

## ESTABILIDADE DO CONCURSO VESTIBULAR DO CEECEM \*

THOMAS F. BREEN III \*\*

### R E S U M O

*Problema:* A questão da estabilidade da classificação dos candidatos ao ensino superior pelo concurso vestibular tem muita importância para os candidatos e para as universidades. Uma vez que os exames vestibulares são mais exames de desempenho que de habilidades, existem muitos fatores que podem mudar os resultados e a classificação dos candidatos de uma aplicação do exame à outra.

*Método:* Esta pesquisa analisou a fidedignidade teste-reteste do concurso vestibular do CEECEM. Usando os resultados dos candidatos que fizeram os exames nos anos de 1972 e 1973, uma análise de correlação momento-produto de Pearson relacionou os resultados de um ano com o ano seguinte.

*Resultados e Conclusões:* Dados preliminares indicam que a média desta estimativa da fidedignidade dos sete subtestes do exame chega acerca de 0,70. Estes dados estabelecem uma estabilidade alta do exame do CEECEM, indicando que fatores consistentes operam nos resultados dos candidatos apesar das influências fortuitas e manipuláveis.

### S U M M A R Y

*Problem:* The stability of rankings of college applicants in the "vestibular" (college entrance exam) is very important both for the candidates and for the universities. Since the exams are achievement tests and not ability tests, there are many factors that may change the scores from one year to the next.

*Method:* This study analyzed test-retest reliability of CEECEM's tests. ("Center for Selection of College Applicants" of Carlos Chagas Foundation). The sample was made up of all applicants that took the exams twice, in 1972 and 1973. Product moment correlations were computed between scores of the two years.

*Results and Conclusion:* The mean reliability coefficient of the seven sub-tests was 0,70. This indicates that the tests have a high stability, which means that test scores are strongly influenced by stable factors in spite of some other influences.

### INTRODUÇÃO

Dentro das áreas de pesquisa sobre a realidade educativa brasileira, apontadas como de interesse pelo INEP, encontra-se a da seleção de candidatos para o ensino superior. Embora recentemente o colégio tenha passado a ser visto como possível ponto terminal de treinamento escolar, o ingresso na universidade continua a ser um objetivo desejável para muitos alunos que o concluem, como se pode ver pelo grande número de candidatos que se apresentam aos vários concursos vestibulares e pelo grande número de estudantes matriculados nos vários "cursinhos".

Esta pesquisa — que faz parte de um programa de pesquisas desenvolvido pela Fundação Carlos Chagas sobre fatores relacionados com seleção para

a Universidade e êxito na Universidade — é etapa preliminar de um projeto que tem por objetivo estudar fatores que produzem mudanças, de um ano para outro, nas médias gerais dos candidatos ao vestibular. Tal projeto pode considerar-se como variante do modelo experimental conhecido como modelo pré-teste/pós-teste.

Qualquer projeto de pesquisa, especialmente com esse modelo, exige um instrumento de avaliação

\* Comunicação apresentada à XXVI Reunião Anual da SBPC, Recife, Julho de 1974.

\*\* Assessor Técnico do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas.

ção que não seja sujeito a grande variação em estilo, formato, e seleção de itens, a qual invalidaria qualquer comparação entre as notas das duas aplicações do teste. Um projeto de pesquisa contínuo, que utilize resultados de mais de um ano, precisa de um critério — neste caso, seleção de candidatos ou resultado no exame vestibular — que possa comparar-se de uma aplicação para outra.

## FUNDAMENTOS

O problema pertence à área da fidedignidade do instrumento. Conseqüentemente, tentar-se-á em primeiro lugar definir e explicar o conceito de fidedignidade, do modo como se aplica a este projeto.

Fidedignidade do instrumento pode considerar-se de várias maneiras. No sentido geral, significa que o instrumento mede o que pretende medir de maneira exata, precisa e consistente. No contexto desta pesquisa, significa que os concursos vestibulares estão medindo de maneira exata a capacidade de ser bem sucedido no curso universitário, isto é: que a medida obtida do desempenho necessário ou da inteligência necessária para estudos superiores (do objetivo medido pelas provas) é precisa e exata, e não muda muito de aplicação para aplicação das provas. Aqui estamos usando a definição de Ebel (1972, p. 409) de que fidedignidade é a "consistência com que um conjunto de escores medem aquilo que medem".

Existem várias maneiras de estimar essa característica numa medida de acordo com o tipo de fidedignidade estimado. O método mais usado nos concursos vestibulares, quando se calcula a sua fidedignidade, é o método da consistência interna. Essa é uma estimativa do grau em que todos os itens do teste medem a mesma coisa (Para o vestibular do CESCEM tais estimativas, utilizando a fórmula 20 de Kuder-Richardson, geralmente estão entre 0,80 e 0,90, indicando uma consistência geralmente alta).

## MÉTODOS

Estimou-se a fidedignidade selecionando uma amostra de candidatos que fizeram o vestibular em dois anos seguidos (1972 e 1973) e comparando os resultados.

O *concurso vestibular* utilizado para a análise foi o exame preparado anualmente pelo CESCEM para os candidatos às áreas de biologia e medicina em São Paulo. O exame, aplicado em janeiro, compõe-se de sete sub-testes: Matemática; Língua Estrangeira (Inglês ou Francês); Conhecimentos Gerais; Química; Português; Física; e Biologia.

O problema desta pesquisa é se os resultados do concurso vestibular de um ano podem comparar-se com os resultados do ano seguinte. Esse problema surge porque, como se sabe, não se aplicam as mesmas provas mais do que uma vez. Como novas provas são elaboradas cada ano, qual a estabilidade que existe entre a avaliação de candidatos de um ano e a do ano seguinte?

Dois outros métodos são o de múltiplas aplicações do mesmo teste e o de formas alternativas do teste. Estes métodos pretendem averiguar respectivamente, se a medida de uma determinada característica é consistente de um momento para outro, ou é consistente com outras amostras de itens destinados a medir a mesma característica. É uma combinação desses dois métodos que interessa ao presente trabalho.

Um teste é geralmente considerado como observação de uma amostra de comportamento. Um teste qualquer, contém apenas uma amostra do universo de comportamento, do qual muitas outras amostras são possíveis. Os últimos dois métodos citados para estimar fidedignidade referem-se à consistência do desempenho de sujeitos em outras amostras de comportamento, ou na mesma amostra, observada em ocasiões distintas.

Nesta pesquisa, considera-se cada vestibular uma amostra de comportamento, isto é, uma seleção de itens a partir de um universo de itens. A cada ano, os elaboradores das provas selecionam mais uma amostra de itens do universo, para compor o exame daquele ano. Nestes termos, o problema desta pesquisa poderia ser, então, reapresentado do seguinte modo: Quão estável é a medida de características pelo concurso vestibular, quando os candidatos são submetidos a duas amostras diferentes de itens, separadas pelo intervalo de um ano?

No processo de seleção dos candidatos, os resultados dos sub-testes são padronizados e combinados numa média ponderada. Essa média ponderada forma a escala usada para a seleção. Nesta pesquisa, contudo, calcularam-se as fidedignidades de cada um dos sub-testes.

A *amostra* para a pesquisa constou de todos os candidatos que fizeram o vestibular do CESCEM em 1972 e 1973. Esses candidatos fizeram exame dois anos seguidos, porque não conseguiram nota suficientemente alta no primeiro ano para serem sele-

cionados, ou porque quiseram entrar em outra carreira.

A amostra foi escolhida pelo seguinte processo. Os nomes dos candidatos que prestaram o exame-vestibular em 1972 e 1973 foram colocados em listas por ordem alfabética. As duas listas foram comparadas selecionando-se os nomes que constavam em ambas. Para evitar maior probabilidade de erro, o número do documento de identidade dos vestibulandos foi considerado. Em todos os casos em que os nomes e números eram idênticos, os resultados do candidato foram utilizados na análise.

Um arquivo foi construído com os resultados de cada candidato nos dois exames. Usando esse arquivo, calcularam-se médias e desvios-padrão dos testes, e uma matriz de correlação. Nesse ponto, para simplificar a análise, foram eliminados os resultados dos candidatos que não fizeram todos os

sub-testes. As análises se baseiam numa amostra final de 4.547 candidatos.

Essa amostra não pode ser considerada representativa da população que prestou o exame em qualquer dos dois anos. Por não incluir os candidatos selecionados em 1972, provavelmente é mais homogênea que a população total, com uma média mais baixa. Este fato leva-nos a predizer que as estimativas da fidedignidade resultantes dessa análise tenderiam a ser um pouco mais baixas do que as fidedignidades verdadeiras, devido à diminuição da variância dos resultados dos testes. Também é razoável esperar que o grupo de candidatos que pretende fazer o exame mais uma vez faça um preparo especial, com mais estudo e talvez um "cursinho", para melhorar seus resultados. Este fato pode contribuir, também, para uma estimativa baixa da estabilidade do teste.

### ANÁLISE E RESULTADOS

Antes de calcular as fidedignidades dos testes concluídos no vestibular do CESCEM, as médias e desvios-padrão das duas amostras, foram comparados com os parâmetros das populações correspondentes dos dois anos. Também foram calculadas matrizes de correlação entre os testes, para os dois anos.

O Quadro I compara os resultados dos testes da amostra de 1972 com os da população. Como se esperava, a amostra escolhida para a pesquisa não é representativa da população que fez o vestibular do CESCEM. A média da amostra em todos os testes é pouco menor que a da população, mas essa diferença é altamente significativa. Um valor de  $z = 3,3$  pode ser significativa com probabilidade menor que 0,001 num teste bicaudal.

Os desvios-padrão da amostra em todos os testes são um pouco menores que os valores da população. A amostra parece, portanto, ser mais homogênea do que o grupo total que fez o exame, e tem, positivamente, um desempenho mais baixo. Essa homogeneidade se deve, possivelmente, à seleção dos melhores candidatos para a universidade e ao desânimo de muitos que ficaram no extremo mais baixo da distribuição.

O Quadro II mostra a matriz de correlações entre os sete testes de 1972. Essas correlações levam a hipótese de que existem provavelmente três ou quatro fatores nos exames, um fator geral de inteligência, um fator de matemática-ciência, um fator de linguagem e, possivelmente, um fator de conhecimento geral. Os testes estão ordenados na matriz para mostrar essas possíveis relações mais claramente.

QUADRO I — MÉDIAS, DESVIOS-PADRÃO E DIFERENÇAS ENTRE POPULAÇÃO E AMOSTRA, NOS TESTES DO VESTIBULAR DO CESCEM, 1972.

| Testes       | População <sup>a</sup> |        | Amostra |        | Diferença | z <sup>b</sup> |
|--------------|------------------------|--------|---------|--------|-----------|----------------|
|              | Média                  | Desvio | Média   | Desvio |           |                |
| Matemática   | 22,8                   | 8,90   | 21,7    | 7,53   | — 1,1     | — 8,06         |
| Conh. Gerais | 31,0                   | 8,00   | 29,9    | 7,36   | — 1,1     | — 9,27         |
| Química      | 25,5                   | 10,10  | 23,7    | 7,95   | — 1,8     | — 12,02        |
| Português    | 54,8                   | 11,70  | 53,3    | 11,00  | — 1,5     | — 8,65         |
| Física       | 25,8                   | 9,20   | 24,5    | 7,50   | — 1,3     | — 9,53         |
| Biologia     | 25,6                   | 8,50   | 24,3    | 6,98   | — 2,3     | — 18,25        |
| Inglês       | 30,6                   | 11,70  | 29,3    | 10,09  | — 1,3     | — 7,49         |

<sup>a</sup> Baseada numa amostra aleatória de 10%

<sup>b</sup>  $z = (\bar{x} - \mu_0) / (\sigma_x / \sqrt{n})$ .

QUADRO II — CORRELAÇÕES ENTRE OS TESTES, PARA A AMOSTRA. VESTIBULAR CESCEM, 1972.

|       | Qui. | Bio. | Mat. | C. G. | Port. | Ling. |
|-------|------|------|------|-------|-------|-------|
| Fis.  | 69   | 65   | 67   | 46    | 42    | 38    |
| Qui.  |      | 67   | 69   | 44    | 39    | 37    |
| Bio.  |      |      | 61   | 48    | 43    | 42    |
| Mat.  |      |      |      | 40    | 44    | 43    |
| C. G. |      |      |      |       | 45    | 40    |
| Port. |      |      |      |       |       | 60    |

As relações entre os testes que se mostram nesses dados são muito semelhantes às aquelas calculadas para o grupo total de 1972 e para outras amostras de outros anos. Barroso (1972) apresenta algumas dessas matrizes. Esses dados indicariam que o padrão de desempenho da amostra foi semelhante ao do grupo total, apesar de ter alcançado um desempenho não representativo em termos de qualidade.

No Quadro III comparam-se as notas da amostra do vestibular 1973, com as da população do mesmo ano. Inesperadamente, a amostra teve um resultado significativamente maior do que a média da população. O fato de que os desvios-padrão da amostra são aproximadamente iguais aos da população indicaria que o grupo cai numa distribuição semelhante, mas com um desempenho significativamente mais alto do que o desempenho da população geral.

Diferenças tão grandes poderiam pôr em dúvida a comparabilidade e consistência dos resultados do teste de um ano para outro. Pareceria que muitos candidatos não selecionados num ano, poderiam ter sido selecionados no ano seguinte.

A matriz de correlação apresentada no Quadro IV para a amostra de 1973 é muito semelhante à do Quadro II, relativa à amostra de 1972. A única diferença é que os valores são ligeiramente mais altos para o segundo ano. Isso provavelmente se deve apenas às variâncias maiores dos resultados dos sub-testes. As correlações entre medidas tendem a ser maiores quando essas medidas têm variâncias maiores.

Apesar de grande melhora apresentada pela amostra do vestibular de 1972 em 1973, o Quadro V mostra que a fidedignidade teste-reteste, para cada um dos sub-testes, é geralmente alta. Embora haja alguma diminuição em relação aos coeficientes de fidedignidade de consistência interna apresentados pelo CESCEM, as fidedignidades para esses exames, calculadas sobre uma amostra homogênea, não representativa, ainda apresenta um valor médio de aproximadamente 0,68.

Isto significa que, apesar das variações na quantidade de treinamento e de estudo ocorrida entre as duas aplicações do exame, apesar do intervalo de um ano entre as duas aplicações, e apesar

QUADRO III — MÉDIAS, DESVIOS-PADRÃO E DIFERENÇAS ENTRE POPULAÇÃO E AMOSTRA, NOS TESTES DO VESTIBULAR DO CESCEM, 1973.

| Testes       | População <sup>a</sup> |        | Amostra |        | Diferença | z <sup>b</sup> |
|--------------|------------------------|--------|---------|--------|-----------|----------------|
|              | Média                  | Desvio | Média   | Desvio |           |                |
| Matemática   | 18,3                   | 7,29   | 21,0    | 7,62   | 2,7       | 24,98          |
| Conh. Gerais | 30,2                   | 7,69   | 31,4    | 7,24   | 1,2       | 10,53          |
| Química      | 28,0                   | 12,13  | 33,4    | 12,86  | 5,4       | 30,03          |
| Português    | 54,7                   | 12,25  | 57,2    | 11,40  | 2,5       | 13,77          |
| Física       | 25,2                   | 9,40   | 28,9    | 9,94   | 3,7       | 26,56          |
| Biologia     | 28,1                   | 8,72   | 31,1    | 8,63   | 3,0       | 23,20          |
| Inglês       | 30,4                   | 11,93  | 32,1    | 11,44  | 1,7       | 9,61           |

<sup>a</sup> Baseada numa amostra aleatória de 10%

<sup>b</sup>  $z = (\bar{x} - \mu_0) / (\sigma_x / \sqrt{n})$ .

QUADRO IV — CORRELAÇÕES ENTRE OS TESTES PARA A AMOSTRA. VESTIBULAR DO CESCEM, 1973.

|       | Qui. | Bio. | Mat. | C. G. | Port. | Ling. |
|-------|------|------|------|-------|-------|-------|
| Fis.  | 84   | 75   | 74   | 47    | 49    | 48    |
| Qui.  |      | 76   | 76   | 51    | 50    | 53    |
| Bio.  |      |      | 63   | 57    | 56    | 53    |
| Mat.  |      |      |      | 39    | 48    | 50    |
| C. G. |      |      |      |       | 50    | 47    |
| Port. |      |      |      |       |       | 65    |

de que o segundo exame contém uma amostra de itens inteiramente distinta, o desempenho dos candidatos, de um para outro ano, é razoavelmente consistente.

Como mostram as diferenças entre os Quadros I e III, essa consistência de desempenho não quer dizer que não haja possibilidade de melhora de um ano para o ano seguinte. O que parece indicar é: a) que o grupo que retorna ao exame uma segunda (ou terceira) vez está mais bem preparado para a prova e para a situação de prova do que estava na sua primeira experiência com o vestibular; e b) que a ordem dentro do grupo é, de modo geral, estável.

Este estudo mostrou que a distribuição de uma amostra de resultados de um ano, no vestibular do CESCEM, é muito comparável aos da mesma amostra observada num exame seguinte. O vestibular do CESCEM demonstra, assim, possuir alta fidedignidade teste-reteste, provavelmente devida à elaboração de provas bem controlada e planejada. Esses resultados são, provavelmente, generalizáveis a todo vestibular que seja preparado com os mesmos ou com melhores padrões que os do CESCEM.

Contudo, surgiram também, da análise, implicações de que haja alguns fatores intervenientes, ainda desconhecido entre as duas aplicações do exame relacionados com uma realização melhor na segunda aplicação. Pesquisa subsequente deve tentar identificar estes fatores, que podem estar relacionados com motivação ou estudo.

Outra área de pesquisa seria a relação entre as mudanças dos resultados no vestibular, de um ano

QUADRO V — FIDEDIGNIDADE TESTE-RETESTE VESTIBULAR DO CESCEM, 1972-1973.

| Testes       | Fidedignidade |
|--------------|---------------|
| Matemática   | 0,64          |
| Línguas      | 0,79          |
| Conh. Gerais | 0,65          |
| Química      | 0,67          |
| Português    | 0,78          |
| Física       | 0,68          |
| Biologia     | 0,60          |

#### CONCLUSÕES

para o seguinte e desempenho acadêmico. Resta ver como essas mudanças com mais um ano de preparação se relacionam com o desempenho na universidade. Barroso (1972) apresenta dados indicando que alunos com êxito no vestibular em sua primeira tentativa, estão mais bem preparados para o trabalho universitário do que aqueles que tem êxito em tentativas posteriores. Esses dados podem significar que habilidade é mais importante para uma boa realização universitária do que motivação ou estudo, caso se demonstre que estas são as razões das mudanças nos resultados das provas.

Esta pesquisa mostrou que o vestibular do CESCEM é um instrumento fidedigno apropriado para seleção para a universidade e pesquisa educacional. Mostrou, também, que muitas perguntas ficam ainda sem resposta, sobre a seleção para a universidade e a relação entre habilidade, desempenho anterior, motivação e desempenho acadêmico.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- BARROSO, C. L. de M. 1972. Estudos de predição do comportamento acadêmico; II — Faculdades de Medicina. *Cadernos de Pesquisa*, (5): 55-76, nov.
- BARROSO, C. L. de M.; RIBEIRO NETTO, A. e COELHO, M. H. M. 1972. Estudos de predição do comportamento acadêmico; I — Faculdade de Medicina Veterinária da USP. *Cadernos de Pesquisa*, (5): 37-53, nov.
- EBEL, R. L. 1972. *Essentials of educational measurement*. Prentice-Hall, Englewood Cliffs, New Jersey.
- GOMES, P. A. e KLARE, G. R. 1967. *Elementary statistics: data analysis for the behavioral sciences*. MacGraw-Hill, Book Company, New York.

[Artigo recebido para publicação em julho de 1974]