

Correlação entre Avaliações por Testes de Múltipla Escolha e por Provas Analítico-Expositivas

Alceu G. de Pinho

Instituto de Física, Universidade de São Paulo

Caixa Postal 20516, CEP 05508-900, São Paulo, SP, Brasil

e

Fundação Universitária para o Vestibular, FUVEST

Antigo Edifício da Reitoria, Cidade Universitária

CEP 05508-900, São Paulo, SP

Trabalho recebido em 6 de dezembro de 1994

Resumo

Neste trabalho são estudadas correlações existentes entre avaliações feitas por meio de testes objetivos e por meio de provas analítico-expositivas. Os concursos vestibulares da FUVEST, realizados em duas fases, prestam-se a esse tipo de comparação uma vez que os candidatos, selecionados na primeira fase por uma prova composta por 72 testes, são, em seguida, submetidos a um conjunto de provas discursivas versando sobre matérias idênticas às examinadas na primeira fase. Uma vez removido, estatisticamente, o componente casual presente nos índices de acerto dos testes, analisa-se a correlação entre os resultados obtidos nas duas formas de avaliação pelo grupo de 8.371 candidatos que, tendo sido submetidos às duas avaliações, alcançaram uma das vagas em disputa nas 69 "carreiras" do Concurso Vestibular de 1994.

I. Introdução

Apesar de testes de múltipla escolha serem utilizados há muitas décadas como um método de avaliação de conhecimentos, persiste, ainda hoje, alguma desconfiança quanto à sua eficácia, quando comparados a avaliações realizadas através de questões analítico-expositivas. Entretanto, desde o aparecimento de leitoras ópticas rápidas e confiáveis, os testes tornaram-se a única maneira capaz de dar, em tempo curto, uma resposta com padrão de correção homogêneo quando se examina o desempenho de dezenas de milhares de respondentes.

Por tal motivo a FUVEST, desde sua fundação em 1977, vem adotando testes de múltipla escolha na primeira fase de seus concursos vestibulares para ingresso na Universidade de São Paulo. Nessa etapa, um número que varia de 25 a 35% dos candidatos, conforme o ano, são selecionados para uma segunda etapa constituída

por provas dissertativas. O número de inscritos nos concursos vestibulares da FUVEST vem se mantendo no patamar da centena de milhar sendo que 134.786 indivíduos realizaram a prova da primeira fase, em 1994.

Essa última prova se compunha de 72 questões, das quais 12 sobre Português e 10 sobre cada uma das seguintes matérias: Biologia, Matemática, Física, Química, História e Geografia. Foram convocados para a segunda fase 38.452 candidatos que realizaram, então, provas discursivas sobre as mesmas matérias arroladas acima, mais uma prova de Língua Estrangeira e uma Redação. A segunda seleção levou à escolha dos candidatos que ocuparam as vagas em jogo. É importante frisar que a classificação final foi feita exclusivamente com as notas das provas dissertativas.

O objetivo do presente trabalho é comparar o resultado obtido pelos candidatos que lograram matrícula nas vagas disponíveis nos dois tipos de avaliação: nos exames da primeira fase (testes) e naqueles da segunda

fase (provas analítico-expositivas). Os dois ocorreram com cerca de 30 dias de intervalo um do outro.

As provas de segunda fase, em cada matéria, eram constituídas por 6 questões, se a prova era de nível 1, ou por 12 questões (incluindo as 6 de nível 1), se a prova era de nível 2. Dependia da carreira escolhida pelo candidato qual o nível, 1 ou 2, de sua prova.

II. Resultados

Foram comparados os seguintes resultados obtidos pelos candidatos que ocuparam as 8.371 vagas:

1) Em cada uma das carreiras, o índice efetivo de acerto na totalidade dos testes da primeira fase com a média ponderada do grupo no conjunto das provas da segunda fase (numa escala de 0 a 10). Tal média leva em conta se uma dada prova foi de nível 1 (peso 1) ou de nível 2 (peso 2), inclui a prova de Língua Estrangeira e a Redação (esta última sempre com peso 2) mas exclui eventuais provas de Habilidades Específicas que ocorrem em carreiras nas áreas de Artes, Arquitetura e Esporte.

2) Num certo número de cursos, o índice efetivo de acerto nos subconjuntos de testes em cada matéria com a média do grupo na prova dissertativa correspondente (numa escala de 0 a 10).

O índice (bruto) de acerto é definido, em todo o teste ou em cada subconjunto de questões, como a média dos índices de acerto em cada uma das questões. Este último é o número de acertos dividido pelo número de respondentes no grupo considerado. É evidente que o índice (bruto) de acerto contém um componente de acertos casuais que, uma vez eliminado, leva ao índice efetivo de acerto. Em trabalho anterior¹ mostramos como o índice efetivo de acerto (IE) pode ser obtido, uma vez conhecido o índice (bruto) de acerto (IA). No caso de questões com cinco opções de resposta foi mostrado, com base nos resultados da prova da primeira fase de 1994, que, com boa aproximação, pode-se escrever

$$IE = IA \left(1 - \frac{1}{\alpha} \ln \frac{B}{IA - 0,2} \right) \quad (1)$$

onde tanto IA como IE são expressos em porcentagem e α e B são coeficientes obtidos empiricamente a partir do ajuste de curvas aos valores do componente casual determinado, em cada questão, pelos métodos descritos na Referência 1. No caso do grupo de alunos matricu-

lados, coeficientes $\alpha = 3,1$ e $B = 0,78$ foram encontrados, o que permite obter diretamente IE pela expressão acima, para $IA > 24\%$, o que ocorreu em todos os casos.

Como, no entanto, trabalha-se com médias sobre várias questões, preferiu-se usar a expressão linear

$$\langle IE \rangle = 1,3 \langle IA \rangle - 30 \quad (2)$$

que dá resultados que diferem, em média, menos de 10% em relação aos valores da expressão (1). Nessa última expressão, $\langle IA \rangle$ e $\langle IE \rangle$ são valores médios sobre várias questões e estão multiplicados por 100.

A figura 1 representa a relação entre $\langle IE \rangle$ para a totalidade dos testes e a média ponderada das provas da segunda fase. Cada símbolo na figura corresponde a uma das 69 carreiras oferecidas pela FUVEST no vestibular de 1994. Os dois símbolos situados mais à direita referem-se aos cursos de Medicina (USP - Capital) e de Mecatrônica (Escola Politécnica).

As figuras 2 a 14 representam a relação entre $\langle IE \rangle$ para o subconjunto de testes em cada matéria indicada e a média na prova dissertativa correspondente. Cada símbolo corresponde a um curso dentro das carreiras da FUVEST.

As figuras 15 e 16 representam a relação entre $\langle IE \rangle$ para o conjunto dos testes (como na figura 1) e as médias nas provas de Língua Estrangeira e de Redação, respectivamente, para vários cursos.

III. Conclusões

Em cada uma das distribuições consideradas, buscou-se verificar a existência de uma correlação linear entre o par de variáveis envolvidas. Para tanto, foi encontrada a reta que melhor se ajusta ao conjunto de dados, pelo método dos mínimos quadrados. A qualidade do ajuste é medida pelo coeficiente de correlação, r , cujo valor absoluto, podendo variar de 0 a 1, mede quão bem a reta calculada ajusta-se aos dados.

No caso da figura 1, o valor $r = 0,95$ indica uma magnífica correlação para o teste como um todo. Foram duas avaliações absolutamente independentes uma da outra, tendo em comum apenas os programas básicos de cada uma das disciplinas. O desempenho na prova de múltipla escolha é medido por $\langle IE \rangle$ (que varia de 0 a 100) e nas provas analítico-expositivas é medido por $\langle M \rangle$ (que varia de 0 a 10). Mesmo sem a subtração dos acertos casuais, isto é, trabalhando-se com $\langle IA \rangle$, é observada a correlação, mas a qualidade do ajuste

Figura	n	r	m	b
1	71	0,95	0,083	+ 0,62
2	22	0,98	0,092	+ 0,87
3	21	0,96	0,091	+ 0,77
4	20	0,99	0,104	+ 0,27
5	20	0,97	0,100	- 1,34
6	22	0,92	0,150	+ 0,77
7	24	0,97	0,106	+ 1,90
8	22	0,97	0,074	+ 0,23
9	30	0,98	0,070	-0,75
10	50	0,79	0,084	- 0,85
11	21	0,97	0,092	+ 1,83
12	39	0,93	0,102	+ 2,30
13	33	0,93	0,115	- 2,22
14	45	0,87	0,105	- 2,39
15	38	0,92	0,092	+ 0,42
16	38	0,82	0,028	+ 3,14

Tabela 1. Para cada uma das figuras 1 a 16, são apresentados: n, total de pares de números utilizados para definir a reta que melhor se ajusta aos dados; r, coeficiente de correlação linear; m e b, coeficientes da reta $y = mx + b$.

linear é muito inferior uma vez que o valor de $\langle IA \rangle$ não cai a zero, em virtude da presença dos acertos ao acaso, tendendo a um valor mínimo de ordem de 20, em testes com cinco opções. Observa-se também que, para o grupo considerado, isto é, os alunos matriculados, os testes foram, no seu conjunto, um pouco mais fáceis que as provas dissertativas. Isso é esperado pois, em princípio, as provas dissertativas tinham como objetivo classificar candidatos já previamente selecionados pelos testes.

Representando por $mx + b$ a equação das retas obtidas nas regressões lineares, os valores do número de pontos utilizados em cada figura, n, do coeficiente de correlação, r, e dos coeficientes m e b de cada reta são dados na Tabela I.

Exceto no que diz respeito às figuras 10, 14 e 16, os valores obtidos para r indicam uma excelente correlação entre os valores de $\langle IE \rangle$ e os valores de $\langle M \rangle$ ou $\langle N \rangle$, no conjunto ou em cada matéria, respectivamente, para os grupos considerados.

No que diz respeito às provas de Geografia e Português o que ocorre é que ambas foram muito fáceis e, conseqüentemente, muito pouco discriminativas na primeira fase. Na prova de Geografia $\langle IE \rangle$ ficou entre 42 e 82 e na de Português, pior ainda, entre 61 e 88. Tal concentração de valores de $\langle IE \rangle$ numa estreita faixa impede a presença de correlação significativa.

Quanto à prova de Redação, a comparação deve ser feita com a prova de Língua Estrangeira. Ambas não tiveram uma prova equivalente na primeira fase e, assim, estão sendo comparadas com o valor médio $\langle IE \rangle$ da totalidade dos testes. A idéia por trás dessa comparação é a seguinte: espera-se que grupos de candidatos com um bom desempenho no conjunto das sete matérias exigidas na primeira fase estejam, também, bem preparados em Língua Estrangeira e, ao lado disso, possuam maior facilidade para dissertar corretamente sobre um tema dado. A primeira hipótese é amplamente confirmada como pode ser visto na Figura 15. Já no que diz respeito à Redação, tal não foi verificado. O autor julga que a não existência de uma correlação nítida deve-se, de fato, a deficiências na avaliação e, portanto, na atribuição de notas às Redações. O baixo valor de m e o elevado valor de b mostram claramente que as notas atribuídas às Redações pouco se desviaram de um valor de ordem de 5, nada tendo a ver com o desempenho dos grupos em qualquer outra prova ou no conjunto delas.

A análise dos dados da Tabela I mostra, também, certas particularidades das provas. Assim, um elevado valor absoluto do coeficiente b indica um desequilíbrio entre os níveis de dificuldade nas provas das duas fases que estão sendo comparadas. Se o valor elevado é positivo, os testes foram muito mais difíceis que as questões da prova dissertativa (exemplos: Física no nível 2 e

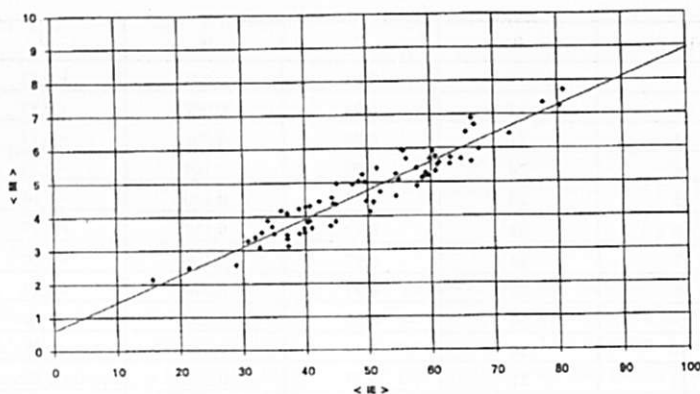


Figura 1. O eixo horizontal dá o valor de $\langle IE \rangle$ para a totalidade dos testes da primeira fase e o eixo vertical dá a média ponderada $\langle M \rangle$ das provas da segunda fase, para o mesmo grupo de candidatos matriculados, conforme explicado no texto. Cada ponto representa uma das 69 carreiras do concurso vestibular FUVEST 94 sendo que os dois pontos de maior valor de $\langle IE \rangle$ representam os cursos de Medicina (USP - Capital) e Mecatrônica (Escola Politécnica, USP).

História nos níveis 1 e 2); se for negativo, é o inverso que ocorreu (exemplos: Química e Matemática no nível 2, Geografia e, principalmente, Português).

Deve ser mais uma vez enfatizado que o uso do índice efetivo de acerto no lugar do índice (bruto) de acerto torna muito mais nítidos e fáceis de interpretar os diferentes coeficientes que caracterizam a regressão linear.

No que foi exposto até agora, a correlação foi examinada comparando-se o desempenho de grupos de candidatos que concorrem em diferentes carreiras ou cursos.

Uma outra maneira de verificar a correlação entre os resultados obtidos nos dois tipos de avaliação será apresentada a seguir. Ela envolve o exame do desempenho de candidatos que concorrem, todos eles, às vagas de uma mesma carreira.

Em cada carreira é necessária uma nota mínima na primeira fase para o candidato ser convocado para a segunda fase. Consideramos candidatos convocados para a segunda fase, numa dada carreira, distribuídos em intervalos de três pontos a partir da nota mínima. Em cada intervalo há C candidatos. Destes, V acabam conseguindo uma das vagas oferecidas na carreira como consequência de seu desempenho nas provas dissertativas da segunda fase. Assim, em cada intervalo de três

pontos, a probabilidade de um candidato aí colocado alcançar uma vaga é V/C . A figura 17 mostra como varia V/C em cada intervalo considerado, para quatro diferentes carreiras: Administração (2008, 400), Odontologia (772, 133), Farmácia-Bioquímica (613, 135) e Engenharia (USP-São Carlos) (453, 120). Os números entre parênteses após cada uma delas indicam, na ordem, o total de convocados para a segunda fase e o total de vagas. Em todas elas é evidente a notável correlação entre os resultados nas provas analítico-expositivas, que determinam a conquista da vaga pelo candidato, e o resultado nos testes de múltipla escolha realizados cerca de um mês antes. Candidatos que passaram para a segunda fase tendo obtido na primeira apenas 6 ou menos pontos acima da nota mínima exigida para a convocação têm escassa probabilidade de lograr uma vaga no final do processo, apesar do resultado da primeira fase não influenciar diretamente na classificação final. Os resultados mostrados para as quatro carreiras acima repetem-se, com pequenas variações, para todas as demais.

Referências

1. de Pinho, A.G., Estimativa de acertos casuais em testes de múltipla escolha, *Rev. Bras. de Ensino de Física*, **17**, 62 (1995).

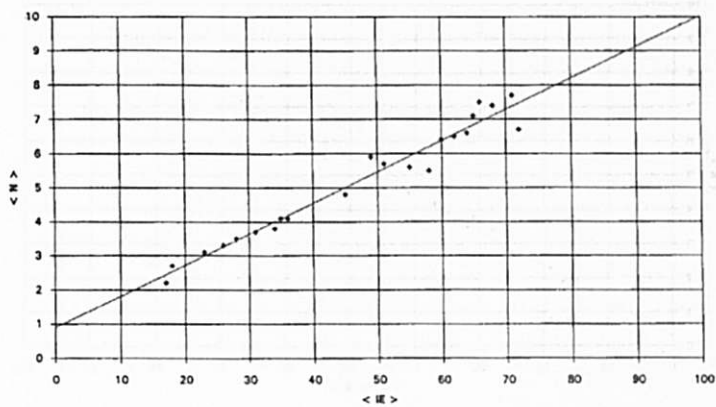


Figura 2. O eixo horizontal dá o valor de < IE > para os alunos matriculados num dado curso e referente aos testes de Biologia e o eixo vertical dá o valor da nota média < N >, para o mesmo grupo, na prova dissertativa de Biologia - Nível I.

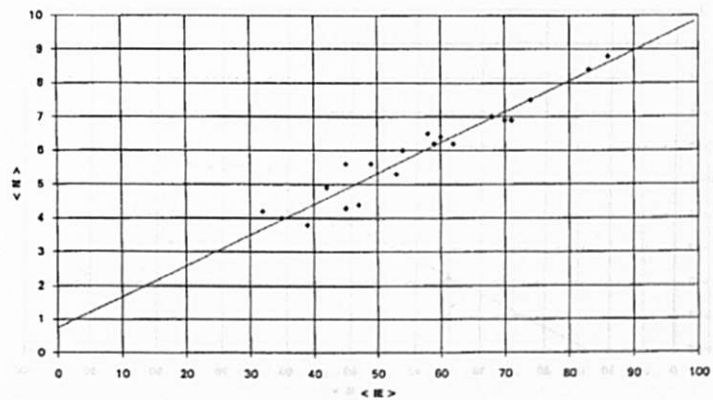


Figura 3. O mesmo que na Fig. 2, para os testes de Biologia e a prova dissertativa de Biologia - Nível 2.

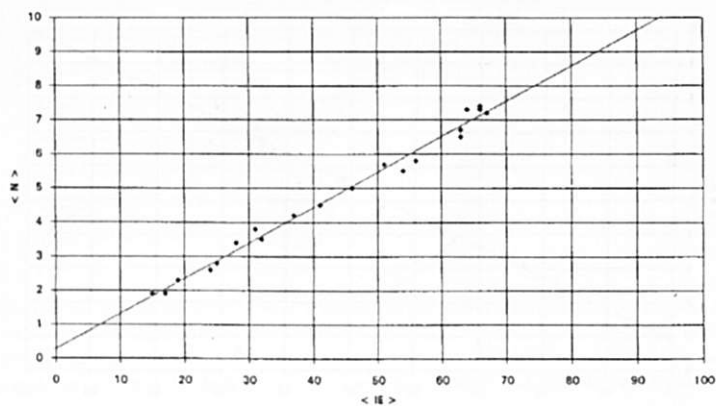


Figura 4. O mesmo que na Fig. 2, para os testes de Química e a prova dissertativa de Química - Nível 1.

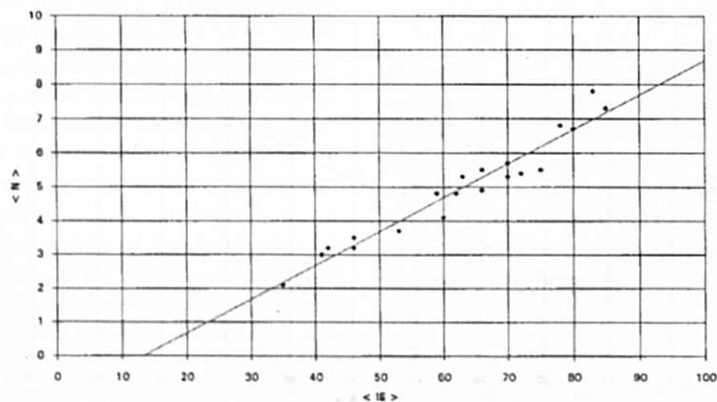


Figura 5. O mesmo que na Fig. 2, para os testes de Química e a prova dissertativa de Química - Nível 2.

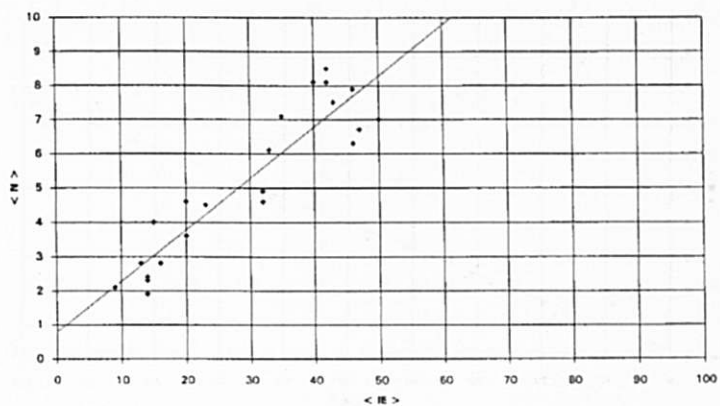


Figura 6. O mesmo que na Fig. 2, para os testes de Física e a prova dissertativa de Física - Nível 1.

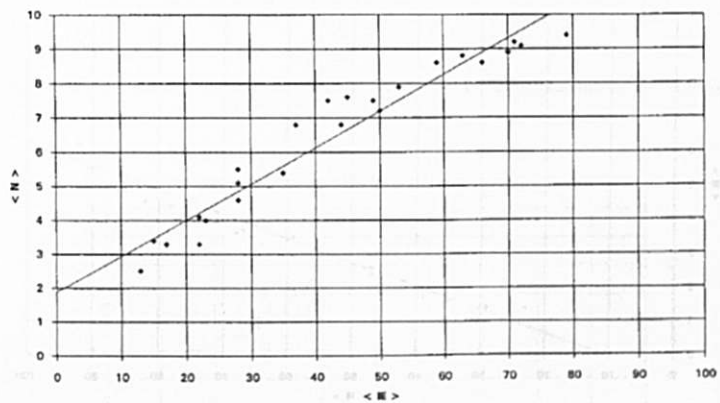


Figura 7. O mesmo que na Fig. 2, para os testes de Física e a prova dissertativa de Física - Nível 2.

Figura 8. O mesmo que na Fig. 2, para os testes de Física e a prova dissertativa de Física - Nível 3.

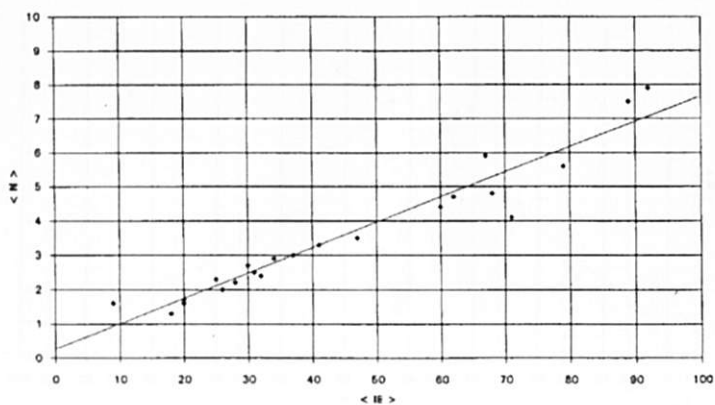


Figura 8. O mesmo que na Fig. 2, para os testes de Matemática e a prova dissertativa de Matemática - Nível 1.

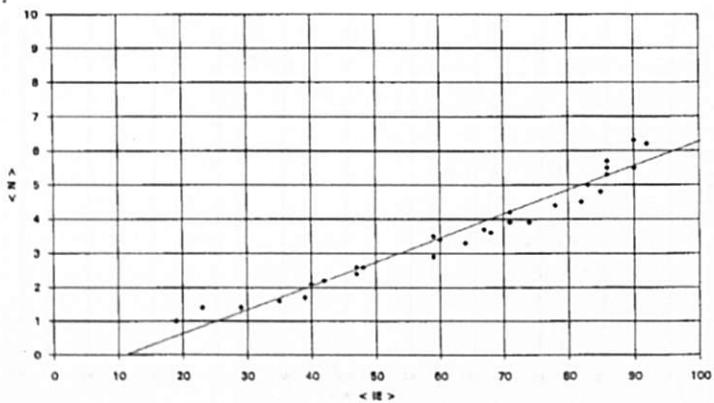


Figura 9. O mesmo que na Fig. 2, para os testes de Matemática e a prova dissertativa de Matemática - Nível 2.

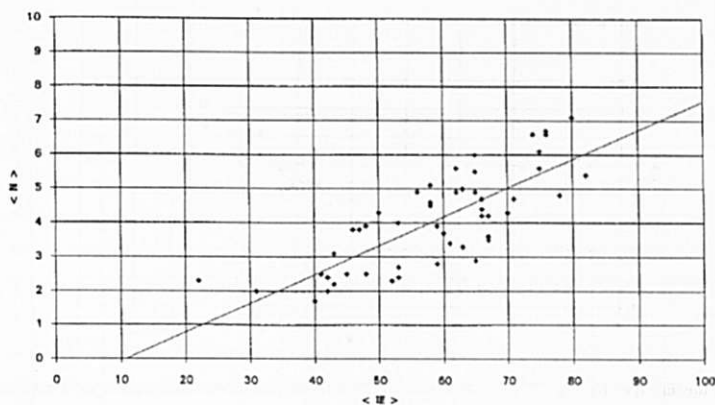


Figura 10. O mesmo que na Fig. 2, para os testes de Geografia e a prova dissertativa de Geografia - Nível 1.

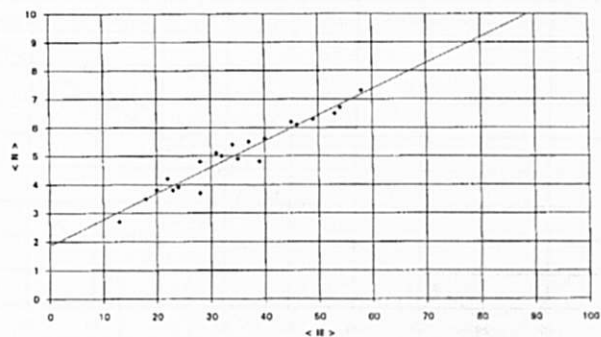


Figura 11. O mesmo que na Fig. 2, para os testes de História e a prova dissertativa de História - Nível 1.

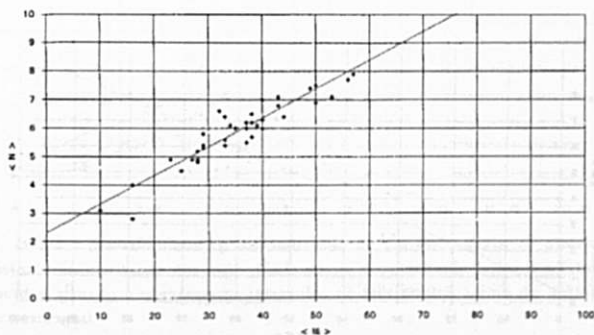


Figura 12. O mesmo que na Fig. 2, para os testes de História e a prova dissertativa de História - Nível 2.

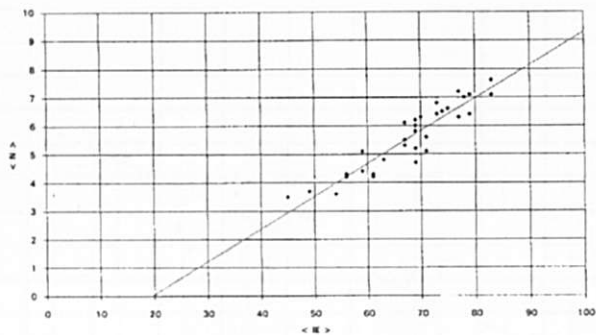


Figura 13. O mesmo que na Fig. 2, para os testes de Português e a prova dissertativa de Português - Nível 1.

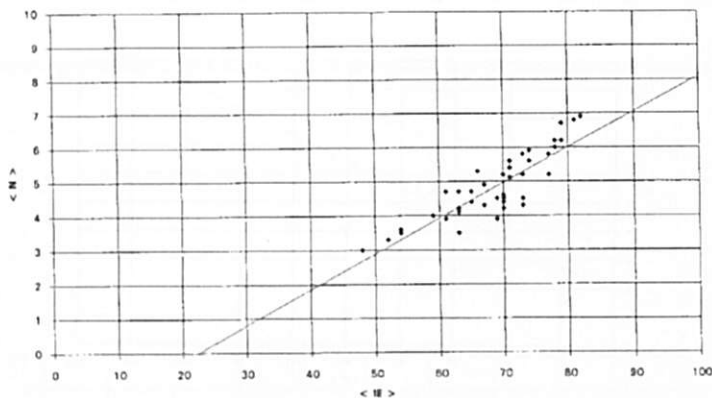


Figura 14. O mesmo que na Fig. 2, para os testes de Português e a prova dissertativa de Português - Nível 2.

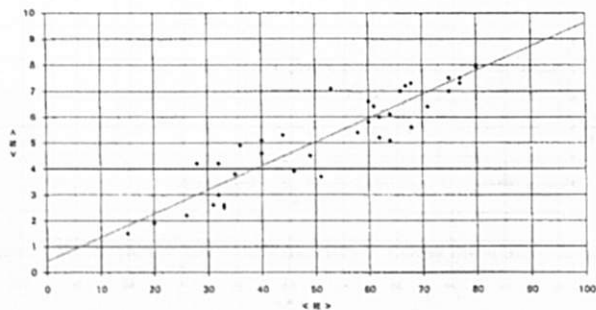


Figura 15. O eixo horizontal dá o valor de < IE > para a totalidade dos testes da primeira fase para os candidatos matriculados num dado curso e o eixo vertical dá o valor da nota média < N > para a prova de Língua Estrangeira, para o mesmo grupo.

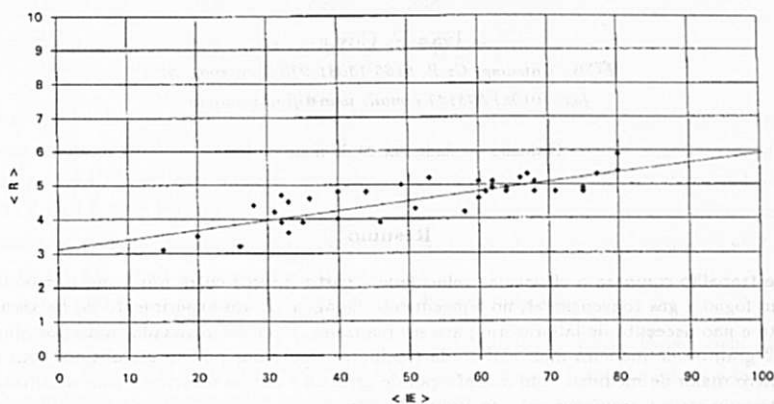


Figura 16. O eixo horizontal dá o valor de $\langle IE \rangle$ para a totalidade dos testes da primeira fase para os candidatos matriculados num dado curso e o eixo vertical dá o valor de $\langle R \rangle$, nota média da Redação para o mesmo grupo.

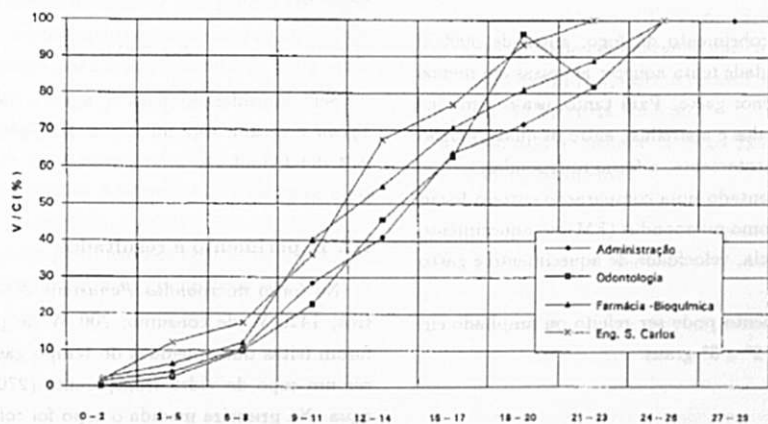


Figura 17: O eixo vertical indica a probabilidade de um indivíduo de um dado grupo de candidatos convocados para a 2ª fase numa carreira conquistar uma das vagas em jogo. Cada grupo é constituído pelos candidatos que passaram para a 2ª fase com um número de pontos situado nos sucessivos intervalos de três pontos acima da nota mínima exigida na carreira, conforme indicado no eixo horizontal.