

APLICAÇÃO DE CRITÉRIOS DE CORREÇÃO EM PROVAS DE REDAÇÃO¹

INTRODUÇÃO

A partir de 1965, a seleção de candidatos para algumas universidades e escolas superiores brasileiras começou a ser feita por meio de provas objetivas, com o emprego de itens de múltipla escolha. Aos poucos, o uso desse tipo de instrumento de medida educacional se difundiu e, na década de 70, o que representava uma experiência limitada a determinados centros educacionais passou a traduzir o comportamento geral das instituições de ensino superior, com apoio em normas oriundas do Ministério da Educação e Cultura. Isso significou, conseqüentemente, o abandono da redação, instrumento de seleção tradicionalmente empregado nos vestibulares, desde a criação desses concursos em 1911.

A reação ao uso exclusivo de provas objetivas não se fez esperar. A Universidade de São Paulo, em 1976, reintroduziu, em seus exames de ingresso, provas dissertativas, que, a partir de janeiro de 1978, também passaram a ser adotadas pelas demais instituições de ensino superior do país, como decorrência da nova orientação do Ministério da Educação e Cultura, que tornou obrigatória a prova de redação em todos os vestibulares.

¹ Artigo publicado em *Cadernos de Pesquisa*, n. 26, p. 29-34, set. 1978.

A redação, na atual sistemática dos vestibulares, que envolvem, muitas vezes, mais de 100.000 estudantes, como nos casos de São Paulo e do Rio de Janeiro, criou, naturalmente, numerosos problemas, entre os quais sobressai o da aplicação de um critério de correção, objetivo do presente estudo.

ELABORAÇÃO DO CRITÉRIO DIRETRIZES GERAIS

A Fundação Carlos Chagas, no primeiro semestre de 1975, iniciou estudos e procurou equacionar, com o máximo de realismo possível, o problema da definição de um critério que permitisse uma correção homogênea de milhares de redações. As diretrizes que orientaram o estabelecimento do critério foram, posteriormente, divulgadas, merecendo destaque o seguinte aspecto:

Buscou-se... operacionalizar um critério que atenda, a um tempo, a dois pressupostos básicos: uniformidade do comportamento analítico e interpretativo' dos diferentes avaliadores, por meio de um padrão objetivo de avaliação; exequibilidade, pela discriminação de um mínimo de subitens que permita considerar um máximo de aspectos determinantes da nota global. (1977a)

O interesse maior dos responsáveis pela elaboração do critério centralizou-se na homogeneidade dos julgamentos. O problema é de importância relativa no caso das redações realizadas em sala de aula, no decorrer de um curso, quando as flutuações de julgamento de um professor se compensam e não prejudicam, ao término da sequência instrucional, a avaliação global do estudante. A situação, no contexto do vestibular, é diferente: a população de candidatos é desconhecida, com escolaridade bem diversificada, e o número de redações a corrigir, em curto espaço de tempo, é elevado. Legítima é, portanto, a preocupação com o estabelecimento de parâmetros que possibilitem avaliações uniformes.

CRITÉRIO DE CORREÇÃO - UM RESUMO

Elaborado por uma equipe de três professores universitários e um professor do 2º Ciclo, do Ensino Oficial do Estado de São Paulo, todos com experiência no ensino de nível médio, o critério foi aplicado, em janeiro de 1976, por quarenta professores, que ofereceram, ao término dos trabalhos, sugestões para modificá-lo em alguns de seus aspectos. Examinadas as críticas e propostas de modificação, a equipe elaborou nova versão do critério, divulgada em 1977, na forma seguinte:

A correção da dissertação deverá considerar os aspectos abaixo discriminados:

	Valores máximos
1 - ESTRUTURA - os julgadores verificarão se o trabalho apresentado pelo candidato é realmente uma dissertação e se essa dissertação constitui um conjunto articulado de partes em torno do tema proposto (forma dissertativa, organicidade e unidade do texto);	20
2 - CONTEÚDO - os julgadores verificarão se a dissertação apresenta ideias fundamentadas e coerentes, que demonstrem senso crítico e que possibilitem uma perfeita relação de entendimento entre o examinando e o avaliador (elaboração crítica, coerência e clareza);	30
3 - EXPRESSÃO - os julgadores verificarão se a dissertação apresenta:	
3.1. adequação vocabular (léxico);	10
3.2. correção gramatical (ortografia, morfologia, sintaxe, pontuação).	40
Total	100

Ressalta do texto adiante transcrito, elaborado à guisa de apresentação, o cuidado dos responsáveis pela formulação do critério com a uniformidade e o rigor dos julgamentos.

Embora não se devessem dissociar forma e conteúdo, ou seja, a articulação do pensamento e sua linguagem, é evidente que, na elaboração do critério, foi preciso fazer uma distinção convencional entre os elementos que entranhadamente constituem um texto. Assim é que, atribuindo valores parcelados a estrutura, conteúdo e expressão, o critério buscou fornecer aos avaliadores instrumentos mais

precisos para o julgamento do texto como um todo. Os três itens são faces diversas do mesmo objeto, que é a prova total como se apresenta em sua redação final, resultante de um processo de elaboração em que o candidato deverá ter considerado necessariamente esses três aspectos. (1977a)

APLICAÇÃO DO CRITÉRIO - PRIMEIRA EXPERIÊNCIA

Após a elaboração do critério, a equipe responsável aplicou-o na correção independente de quatro redações, antes de submetê-lo ao julgamento de outros avaliadores, ainda durante o treinamento. Os resultados dessa experiência inicial são apresentados na Tabela 1.

TABELA 1 - Notas atribuídas a quatro redações por quatro examinadores, em correções independentes. Médias e desvios-padrão. Fundação Carlos Chagas - 1975

PROFESSORES	REDAÇÃO			
	A	B	C	D
1	62,50	48,00	42,50	77,00
2	61,00	23,00	17,00	96,00
3	80,50	41,00	53,00	96,00
4	86,00	35,00	53,00	93,00
\bar{X}	72,50	36,75	41,37	90,50
s	10,67	9,17	15,14	7,88

Os elementos da Tabela 1 demonstram que houve divergências na aplicação do critério pela equipe responsável por sua elaboração, conforme indicam as dispersões (s) das várias notas. Analisando-se, entretanto, as posições atribuídas pelos quatro professores, observa-se que, na aplicação do critério, os julgadores foram unânimes em classificar as redações D e A como as melhores, nas posições 1 e 2, respectivamente. Verifica-se, também, quanto às provas B e C, que os julgadores concordaram que deveriam ocupar os postos 3 e 4, havendo dúvidas quanto à redação que se situaria na posição 3 ou 4, já que

dois deles acharam que a B deveria estar na posição 3, enquanto que os dois outros optaram pela redação C, para esse mesmo posto. Levando-se em consideração que a média dos julgadores é mais fidedigna do que um julgamento isolado (VIANNA, 1977), a aplicação do critério permitiu classificar as quatro provas, a partir da melhor, na seguinte ordenação: D, A, C, e B.

APLICAÇÃO DO CRITÉRIO - NOVA EXPERIÊNCIA

A inclusão de uma prova de redação, no concurso vestibular de 1976, exigiu a seleção de quarenta (40) experientes professores para os trabalhos de avaliação. Os responsáveis pela coordenação das atividades de correção promoveram uma reunião de todos os envolvidos no processo, para apresentação e debates dos problemas. A preocupação da equipe coordenadora relacionava-se diretamente com a possível multiplicidade de julgamentos, pois a literatura técnica é rica em pesquisas que demonstram a diversidade dos critérios individuais de avaliação de provas dissertativas (VIANNA, 1976). A reunião objetivou, sobretudo, sensibilizar os avaliadores para os problemas da correção e vantagens da adoção de um único critério, a fim de atenuar as conhecidas diferenças individuais de julgamento.

Um documento explicativo, detalhando o critério (VIANNA, 1977 b) foi divulgado para encaminhar as discussões e detalhes. A parte introdutória deste documento mereceu especial atenção, sendo enfatizada, mais uma vez, a importância da homogeneidade dos julgamentos.

A correção de redações é tarefa complexa, que envolve grande quantidade de variáveis. O critério geral... procurou abranger os aspectos que devem ser necessariamente observados, a fim de que se obtenham os dados fundamentais para uma avaliação um tanto quanto possível objetiva.”

“Sempre que lemos um texto, é inevitável sermos envolvidos por uma impressão geral, favorável ou desfavorável. Todo professor passa por essa experiência, mesmo que não o queira. No caso presente, o fato de os candidatos estarem empenhados numa situação competitiva obriga-nos a um maior controle da subjetividade. O que se pretende é o estabelecimento de critérios que possibilitem uniformidade

de julgamento por parte de um grupo de avaliação marcado pela diversidade de seus elementos.

Após os debates e a aceitação do critério pelo grupo de avaliação, distribuíram-se as redações anteriormente avaliadas pela equipe de coordenação, solicitando-se aos professores que a corrigissem segundo o critério. A redação D, por solicitação dos examinadores, foi avaliada com base na impressão geral. A Tabela 2 apresenta os resultados dessas correções.

TABELA 2 - Notas atribuídas por quarenta professores a quatro redações, em correções independentes. Médias e desvio-padrão. Fundação Carlos Chagas - 1976

NOTAS	REDAÇÕES			
	FREQUÊNCIAS			
	A	B	C	D
90-96	--	--	--	8
83-89	1	--	--	10
76-82	3	--	--	6(*)
69-75	1	--	--	13
62-68	9	--	2	1
55-61	9(*)	--	3	2
48-54	5	1	5	--
41-47	4	3	2	--
34-40	6	8	16(*)	--
27-33	1	11(*)	7	--
20-26	1	12	2	--
13-19	--	5	2	--
6-12	--	--	1	--
Notas	40	40	40	40
\bar{X}	55,52	29	38,57	79,37
s	14,18	9,27	12,45	8,97

(*) Assinala a classe em que se localiza a média

A análise da Tabela 2 possibilita observar, inicialmente, que as médias do grupo de quarenta professores ordenaram as quatro redações nas mesmas posições apresentadas pela equipe de coordenação. Ainda que as posições sejam idênticas, ressaltam, no caso, as diferenças entre os dois conjuntos de médias. A primeira correção, sob responsabilidade da equipe que elaborou os critérios, foi bem mais tolerante que a de quarenta professores, cujos resultados, entretanto, tendo em vista a teoria da fidedignidade, são mais precisos, por traduzirem um número bem maior de apreciações.

Essas diferenças entre as médias sofreram a influência da grande amplitude das notas dos quarenta professores, conforme a Tabela 3.

TABELA 3 - Amplitude das notas atribuídas a quatro redações, em correções independentes, por quarenta professores. Fundação Carlos Chagas - 1976

NOTAS	REDAÇÕES			
	A	B	C	D
Mínima	88	49	62	95
Máxima	25	14	6	60
Amplitude	64	36	57	36

A essa altura, configuraram-se de imediatos dois problemas graves que, em princípio, podem ter implicações no contexto do vestibular:

- 1º) a grande amplitude das notas, ao que tudo indica, refletiria o fato de que os professores não estariam aplicando os mesmos padrões de correção estabelecidos pelo critério. A atribuição das notas, em muitos casos, foi aparentemente pouco cuidadosa, exemplificados pela redação C, cujas notas variaram de 62 a 6, com uma amplitude, portanto, de 57 pontos; e pela redação A, cujas notas tiveram uma amplitude de 64 pontos;
- 2º) outra constatação, com base na Tabela 2, refere-se à sujeição do estudante à equação pessoal do examinador: seu êxito ou fracasso vai depender da sorte de ter a sua prova corrigida por este ou por aquele professor.

A Tabela 4 reflete uma situação hipotética: considerou-se que somente uma nota superior a 40, nas distribuições apresentadas na tabela 2, traduziria aprovação. A seguir, calculou-se, percentualmente, quantos professores aprovariam ou reprovariam cada uma das quatro redações.

TABELA 4 - Porcentagem de professores que aprovaram e dos que reprovaram, com base em notas atribuídas a quatro redações, em correções independentes, admitindo-se para aprovação um desempenho superior a 40. Fundação Carlos Chagas - 1978

REDAÇÃO	PORCENTAGEM DE		Total
	Aprovação	Reprovação	
A	80	20	100
B	10	90	100
C	30	70	100
D	100	-	100

A correção de uma redação não é assunto pacífico, conforme mostram as porcentagens da Tabela 4. Alguns consideram o trabalho digno de merecer aprovação; outros, ao contrário, o julgam destituído de valor. Os dados sintetizados nas Tabelas 2, 3 e 4 põem em destaque a possível influência do professor no êxito ou insucesso do aluno.

APLICAÇÃO DO CRITÉRIO - VERIFICAÇÃO DE UMA HIPÓTESE

As atividades práticas de 1976 proporcionaram elementos para a reformulação do critério e o estabelecimento de novos procedimentos para sua aplicação. Aos avaliadores solicitaram-se relatórios individuais de sua experiência, especialmente das dificuldades encontradas na aplicação do critério, e apresentação de sugestões para a sua modificação, de modo a garantir o êxito de sua própria utilização. A partir do exame desses relatórios e de suas sugestões modificaram-se alguns elementos do critério, que passou a ter a versão anteriormente apresentada no presente estudo.

A dinâmica da reunião para debate do novo critério, por sugestão dos professores, sofreu, igualmente, modificações. Criaram-se grupos menores para permitir melhor interação dos avaliadores com a equipe de coordenação; desse modo, enquanto na experiência anterior o treinamento ocorreu com a presença de todos os avaliadores (40), nesta organizaram-se grupos com doze (12) professores apenas. Prepararam-se, também, dois documentos (1977a, 1977b) para exame mais detalhado: o primeiro estabelece a metodologia que orientou a elaboração do critério; o segundo define seus diferentes componentes e discute aspectos para a sua aplicação prática e quantificação dos trabalhos. A discussão desses documentos e a aplicação do critério a um novo conjunto de redações ocorreu em duas etapas, totalizando seis horas de atividades, nos meses de novembro e dezembro de 1977. Ao término dessas reuniões, havia, aparentemente, um consenso sobre a aplicabilidade do critério numa situação de vestibular.

Os exames de ingresso à universidade, em janeiro de 1978, permitiram testar a hipótese de que não haveria diferenças estatisticamente significantes entre as médias dos subconjuntos de redações corrigidos por professores que passaram pelas sessões de treinamento. Assim, 2.738 redações foram submetidas ao julgamento de 48 professores treinados, que, em média, corrigiram 57 redações num único dia, em dois períodos, num total de oito horas.

Após a correção das 2.738 provas, calculou-se, inicialmente, a média geral ($\bar{X} = 50,65$); a seguir, a fim de atingir o objetivo maior do presente estudo, estabeleceram-se as médias de cada professor para cada um dos 48 subconjuntos. A Tabela 5 apresenta uma distribuição de frequência dessas médias, nos quarenta e oito subconjuntos. A amplitude das médias foi de trinta (30) pontos.

TABELA 5 - Médias das notas atribuídas a 48 subconjuntos de redações corrigidos por diferentes professores submetidos ao mesmo treinamento. Fundação Carlos Chagas - 1978

x	F
60-62	3
57-59	6
54-56	6
51-53	14
48-50	9
45-47	3
42-44	3
39-41	1
36-38	1
33-35	1
30-32	1
N	48

Observa-se, primeiramente, que 14 médias (29%) se acham na classe modal (51- 53), que também é a classe da média geral. A distribuição demonstra ter havido uma predileção pelos valores centrais, em torno da média. Assim, verificando-se as notas de cada prova, constatou-se que 2133 (77,90%) obtiveram notas na faixa de 30 a 70, numa escala de 0 a 100; que 331 provas (12,10%) receberam notas entre 75 e 100; e que, finalmente, 274 redações (10,00%) tiveram notas de 0 a 25. No conjunto de 2738 provas, atribuiu-se zero a 70, ou seja, 2,56% dos trabalhos receberam essa nota mínima. Apenas 3 provas (0,11%) mereceram nota plena (100). A Tabela 6 apresenta a distribuição de frequência das notas das 2738 redações.

TABELA 6 - Notas atribuídas a 2.738 provas de redação por 48 professores treinados. Fundação Carlos Chagas - 1978

X	F
91-100	19
81-90	111
71-80	201
61-70	396
51-60	512
41-50	631
31-40	460
21-30	232
11-20	93
1-10	13
0	70
N	2.738

A questão fundamental, anteriormente formulada na hipótese, consiste, portanto, em verificar se realmente existe diferença significativa entre as médias dos quarenta e oito subconjuntos de provas. A expectativa, em que pesem elementos anteriormente apresentados, é de que não existam diferenças significativas, tendo em vista o cuidado do critério em eliminar todos os elementos capazes de gerar controvérsias. Além disso, a preocupação da equipe de coordenação em debater e analisar os vários problemas que poderiam ocorrer na fase de aplicação; a constituição de grupos homogêneos de avaliadores, no que tange à formação profissional e experiência docente; e, finalmente, a procura de consenso, supostamente obtido nas sessões destinadas à discussão e aplicação preliminar do critério, permitem esperar que não haja uma variação significativa entre os examinadores.

A fim de testar a hipótese formulada, foi feita a análise da variância (ANOVA *one-way*) das médias das notas atribuídas pelos examinadores aos seus respectivos subconjuntos. A Tabela 7 apresenta os dados da ANOVA.

TABELA 7 - Análise da variância das médias das notas atribuídas a 48 subconjuntos de redações corrigidos por diferentes professores submetidos ao mesmo treinamento. Fundação Carlos Chagas - 1978

FONTE DE VARIAÇÃO	GRAUS DE LIBERDADE	SOMA DOS QUADRADOS	QUADRADOS MÉDIOS	F
PROFESSOR	47	108605,69	2310,76	7,17***
RESÍDUO	2.690	866455,61	322,10	
TOTAL	2.737	975061,30		

*** $p \leq 0,001$ - Altamente significativa

A análise da variância das médias dos 48 subconjuntos mostrou que existem diferenças significativas entre essas medidas de tendência central; conseqüentemente, não se pode admitir que essas discrepâncias sejam puramente casuais, devendo-se atribuí-las à falta de homogeneidade na utilização do critério pelos examinadores. Assim, a hipótese de que não haveria diferenças estatisticamente significantes entre as médias dos subconjuntos de redações, corrigidos por professores que receberam treinamento específico para esse fim, deve ser rejeitada; admite-se, desse modo, que os avaliadores, no processo de correção, empregaram critérios próprios, diferentes do proposto.

CONCLUSÕES

Os elementos coletados no presente estudo possibilitam estabelecer as seguintes conclusões:

- 1- a aplicação de um critério de correção de redação pela própria equipe que o definiu não representa garantia de que haverá uniformidade nos resultados quantitativos apresentados;
- 2- ao corrigirem uma única prova, diferentes examinadores tendem a variações consideráveis nas notas, ainda que concordem quanto à posição dessa prova em relação às demais;
- 3- a variabilidade dos professores em relação a uma única prova é tão grande, mesmo supostamente usando um único critério, que a aprovação ou reprovação do estudante fica sujeita aos azares da sorte;

- 4- a utilização de um número considerável de examinadores, para fins de correção de redações, no contexto do vestibular, exige treinamento específico do pessoal docente, o que, todavia, não garante correção isenta de idiosincrasias individuais;
- 5- apesar da elaboração cuidadosa de um critério, da seleção de avaliadores entre profissionais altamente capacitados e da preocupação em estabelecer, por meio de treinamentos, normas de procedimento uniformes, as diferenças entre as médias dos avaliadores são estatisticamente significantes, conforme os dados da ANOVA – Tabela 7, inferindo-se, portanto, que os 48 professores participantes dessa experiência de avaliação utilizaram critérios – possivelmente, 48 critérios – diferentes daquele que foi apresentado, debatido e supostamente aceito por eles, na fase preliminar de treinamento.

REFERÊNCIAS BIBLIOGRÁFICAS

FUNDAÇÃO CARLOS CHAGAS. *Apresentação do critério de correção*. São Paulo: Fundação Carlos Chagas, 1977a. Mimeo.

_____. *Explicação do critério adotado para julgamento da dissertação*. São Paulo: Fundação Carlos Chagas, 1977b. Mimeo.

VIANNA, Heraldo M. Redação e medida da expressão escrita: algumas contribuições da pesquisa educacional. *Cadernos de Pesquisa*, São Paulo, n. 16, p. 41-47, 1976.

_____. Flutuações de julgamento em provas de redação. *Cadernos de Pesquisa*, São Paulo, n. 19, p. 5-9, 1977.