

DUPLA AVALIAÇÃO DE DISSERTAÇÕES EM CONCURSO VESTIBULAR: COMPARAÇÃO DE ESCORES ⁽¹⁾

Bernardo Buchweitz*
Maria do Horto Soares Motta*

1 – INTRODUÇÃO

Desde a introdução da questão de redação como integrante da prova de Comunicação e Expressão do Concurso Vestibular da Universidade Federal do Rio Grande do Sul (UFRGS) e da Fundação Faculdade Federal de Ciências Médicas de Porto Alegre (FFFCMPA), em 1978, as dissertações dos candidatos vêm sendo avaliadas por um grupo de professores.

Como cada dissertação tem sido corrigida apenas uma vez, somente um escore tem sido atribuído a cada prova, cabendo indagar se diferentes avaliadores tendem ou não a atribuir a mesma nota ao mesmo trabalho. Buscou-se, então, realizar uma experiência com uma dupla correção de dissertações realizadas por um grupo de vestibulandos.

O número de trabalhos sobre concursos para acesso ao ensino superior brasileiro ainda é bem reduzido, particularmente sobre o problema crítico de julgamento de dissertações. No seu livro *Comunicação e Expressão: problemas teóricos e práticos de avaliação*, H.M. Vianna (1984) apresenta e analisa as principais publicações relacionadas com o problema de avaliação de redações. Numa das pesquisas apresentadas nesse livro, aborda a dupla correção em provas de redação. Nesse trabalho, merece destaque a conclusão de que a concordância de duas notas atribuídas à mesma redação em momentos sucessivos, tanto por dois professores como por um único, nunca é absoluta, ocorrendo diferenças consideráveis.

2 – PROCEDIMENTO EXPERIMENTAL

A pesquisa envolveu dissertações de candidatos do concurso vestibular da UFRGS e da FFFCMPA, em 1984. Os vestibulandos realizaram, entre outras, uma prova de redação, que constou de uma questão de dissertação e outra de análise e interpretação de texto. A questão de dissertação foi assim proposta:

(1) Pesquisa parcialmente financiada pela COPERSO.

* Professor da Universidade Federal do Rio Grande do Sul (UFRS).

Leia atentamente o texto e analise as idéias nele contidas. Com base nessa análise, faça uma dissertação em que você exponha seus pontos de vista e suas conclusões.

"O panorama atual das profissões no Brasil é muito inquietante, pois o diploma de curso superior não garante mais emprego, dinheiro ou 'status'. Os jovens de 18 a 20 anos têm diante de si um fenômeno inédito e chocante para o modo de vida brasileiro — está empalidecendo o mito do diploma universitário".

A dissertação deve ter a extensão mínima de 20 linhas e máxima de 30, considerando-se letra de tamanho regular.

Os candidatos dispunham de 3 horas para realizar toda a prova.

As provas foram corrigidas por um grupo de 75 professores universitários e/ou de segundo grau, subdivididos em 5 equipes de 15, cada uma das quais supervisionada por um coordenador e por um supervisor geral. A correção foi feita em 7 etapas. Em cada etapa, os avaliadores recebiam um conjunto de 40 redações e dispunham de 4 horas para corrigi-las. Os escores atribuídos às dissertações obedeciam a uma escala de 0 a 75 pontos, com intervalos de 5.

Na correção da dissertação foram considerados os aspectos de estrutura, conteúdo e expressão, aos quais correspondiam, pela ordem, um número máximo de 15, 25 e 35 pontos. Os avaliadores receberam um texto contendo maiores detalhes sobre esses aspectos e os critérios adotados para julgamento das dissertações, semelhantes aos propostos por Vianna (1984, pág. 47 e 48).

Antes do início das correções, os critérios de avaliação, possíveis tratamentos do tema e outros detalhes foram discutidos e analisados em uma reunião entre coordenadores e avaliadores, com a finalidade de uniformizar a atribuição dos escores às dissertações, isto é, controlar a coerência dos julgadores entre si.

No último turno de correção, 20 avaliadores, sorteados aleatoriamente entre os 75, receberam 40 provas cada um, também escolhidos ao acaso entre as que já haviam sido corrigidas. Esses avaliadores não tomaram conhecimento da nota atribuída na primeira correção e foram orientados por meio do seguinte texto:

Considerando as modificações para o próximo vestibular da UFRGS e valendo-se da oportunidade que ora se apresenta, a COPERSON propõe-se a testar uma nova modalidade de correção pela qual uma mesma redação é avaliada por dois diferentes professores.

Para isso gostaríamos de contar com a sua colaboração no sentido de realizar essa correção com os mesmos critérios até aqui adotados, afastando qualquer tipo de influência que venha a comprometer a validade da pesquisa.

Assim como nas 6 etapas anteriores, na última, os 75 avaliadores corrigiram 40 redações cada um, sendo que 20 deles receberam provas para serem corrigidas pela segunda vez. A atuação dos professores nesse último turno não deixou transparecer nenhuma diferença em relação aos anteriores.

3 – ANÁLISE DOS DADOS

A Tabela I apresenta os escores atribuídos às 800 dissertações submetidas a dupla avaliação. A observação dessas notas indica que, na maioria dos casos, o par de escores, relativo à prova de um mesmo candidato, é constituído de notas diferentes. Essa pequena consistência das notas pode ser melhor observada na Tabela II, que mostra o número de escores iguais

Tabela I. Escores atribuídos às dissertações dos vestibulandos em duas correções

GRUPO VESTIBULANDO	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	X ₁ Y ₁	X ₂ Y ₁	X ₃ Y ₁	X ₄ Y ₁	X ₅ Y ₁	X ₆ Y ₁	X ₇ Y ₁	X ₈ Y ₁	X ₉ Y ₁	X ₁₀ Y ₁	X ₁₁ Y ₁	X ₁₂ Y ₁	X ₁₃ Y ₁	X ₁₄ Y ₁	X ₁₅ Y ₁	X ₁₆ Y ₁	X ₁₇ Y ₁	X ₁₈ Y ₁	X ₁₉ Y ₁	X ₂₀ Y ₁
1	45	50	45	15	25	15	15	65	60	15	50	55	40	45	30	60	60	30	45	
2	60	50	10	40	25	70	50	65	10	25	30	45	55	25	55	40	20	45	35	
3	15	45	65	65	25	35	45	55	65	40	50	15	35	55	40	60	50	30	70	
4	60	50	30	35	30	50	20	15	15	20	40	40	45	60	50	40	30	50	45	
5	40	45	20	35	30	15	45	35	25	40	15	60	50	60	55	40	15	10	65	
6	60	50	20	20	15	15	30	00	70	50	25	35	20	40	45	35	55	20	40	
7	50	35	10	55	30	15	10	25	30	40	35	55	50	50	65	25	30	45	40	
8	05	10	45	70	40	20	65	60	55	40	45	55	60	65	45	40	60	40	30	
9	35	50	75	35	15	35	40	45	45	40	45	50	40	70	45	65	55	30	25	
10	35	35	60	30	40	15	35	50	60	55	40	45	45	30	30	50	40	60	75	
11	55	60	35	40	20	30	15	25	10	30	60	30	35	40	55	40	30	40	35	
12	55	55	35	20	60	25	00	30	40	35	15	65	45	30	45	65	40	25	45	
13	40	40	30	40	25	25	20	55	70	40	55	35	40	30	50	25	35	60	25	
14	55	45	65	55	45	30	40	30	30	25	35	30	35	50	25	35	65	40	25	
15	35	45	50	45	15	60	55	15	20	40	30	50	40	65	50	45	50	45	45	
16	30	35	40	15	70	60	25	30	25	05	30	30	50	20	65	35	25	10	65	
17	30	35	30	45	40	50	35	75	45	55	30	45	35	55	40	50	70	35	45	
18	45	35	25	30	45	55	45	35	65	55	60	40	45	60	50	40	25	15	20	
19	60	30	55	40	65	60	20	25	70	60	15	30	40	25	45	65	35	40	45	
20	55	45	65	55	45	25	55	50	35	45	25	20	15	30	30	50	55	40	35	
21	35	25	15	35	30	45	25	60	40	20	35	45	60	40	25	70	45	45	30	
22	50	65	40	25	65	15	15	35	40	40	30	40	30	25	30	15	35	50	40	
23	40	60	65	60	20	25	35	20	45	35	15	50	20	85	20	35	10	05	30	
24	50	45	35	45	25	30	40	35	40	70	40	45	60	65	60	20	30	30	50	
25	15	40	15	40	45	45	35	45	60	65	40	05	35	30	40	50	35	15	75	
26	40	55	55	35	35	45	50	40	60	20	35	40	40	40	20	40	15	10	25	
27	35	55	25	15	40	20	30	10	50	35	40	65	35	40	55	25	20	35	35	
28	35	25	20	45	55	50	45	45	30	35	20	35	30	55	75	10	15	50	65	
29	45	45	10	25	25	50	45	45	55	35	25	30	60	60	10	05	25	30	55	
30	20	35	75	50	40	55	50	30	30	15	30	35	50	40	55	40	20	40	55	
31	20	40	30	50	65	60	45	55	60	40	55	40	50	70	30	55	60	40	25	
32	20	40	50	15	10	45	45	70	40	45	30	35	45	55	50	35	25	30	35	
33	45	40	55	55	60	35	45	50	45	40	35	55	40	60	55	40	45	30	40	
34	55	50	55	60	50	05	35	20	55	30	40	15	35	40	50	15	35	20	50	
35	15	40	55	45	30	10	25	30	45	50	40	45	40	65	25	35	40	15	40	
36	20	45	25	35	00	30	45	25	75	65	55	30	15	35	30	30	25	35	30	
37	30	50	20	30	75	60	55	55	45	45	15	50	40	15	25	40	35	35	40	
38	70	55	45	60	30	35	30	45	30	10	25	50	50	40	60	55	55	15	15	
39	30	30	35	40	35	45	10	15	65	35	30	30	40	50	70	20	40	40	20	
40	40	25	15	50	50	35	55	60	35	50	50	55	60	50	55	45	40	25	40	

TABELA II - FREQUÊNCIA DAS DIFERENÇAS ENTRE OS ESCORES ATRIBUÍDOS ÀS DISSERTAÇÕES DE 800 VESTIBULANDOS EM DUAS CORREÇÕES. VESTIBULAR DA UFRGS E FFFCMPA, EM 1984.

Diferença ($Y_i - X_i$)	GRUPO																				Total	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	n	%
0	6	6	6	5	6	6	4	6	3	4	4	4	6	5	6	9	10	5	3	5	109	13,6
5	5	5	5	7	5	3	7	2	2	2	2	5	6	4	5	7	5	5	5	5	90	11,2
-5	3	3	5	9	2	6	2	7	7	2	3	5	7	4	8	3	2	3	6	2	89	11,1
10	1	4	6	3	3	3	2	3	5	2	4	2	4	3	1	4	3	9	5	9	76	9,5
-10	7	5	1	3	6	1	7	1	5	8	2	3	4	2	8	1	5	4	4	3	77	9,6
15	4	4	1	2	5	4	4	6	4	1	6	2	3	10	1	4	8	3	3	3	78	9,7
-15	4	3	4	1	1	3	2	3	6	5	4	5	1	3	1	6	3	3	3	58	7,2	
20	5	3	1	2	2	1	3	10	2	3	2	1	2	2	3	3	1	5	1	52	6,5	
-20	2	4	5	2	4	2	2	2	2	3	2	7	2	1	4					42	5,2	
25	3	3	1	1	2	2	2	2	1	2	1	2	1	1	1	2	2	1	1	24	3,0	
-25	1	1	2	1	5			1	4	2	2	1	2	3	1	2	1	1	1	28	3,5	
30	1	2	1	1	3	1			1	1			2	3	1	3	3	2	24	3,0		
-30	1			4	3	1		3	4	1	2	1					1	1	22	2,7		
35		1		1	1	1	1	1	1			1	2					4	12	1,5		
-35			1						1	1	1								4	0,5		
40				1								1				1	1	1	4	0,5		
-40					1						2								3	0,3		
45												1	2						3	0,3		
-45																						
50																						
-50				1																1	0,1	
55																						
-55																						
60				1																1	0,1	
-60																						
65																						
-65																						
70																						
-70																						
75																						
-75																						

e diferentes atribuídos aos candidatos nas duas correções. A diferença $Y_i - X_i$ representa o escore do segundo avaliador (Y_i) menos o escore do grupo de avaliadores (X_i). Verifica-se que apenas 109 (13,6%) das 800 dissertações receberam o mesmo escore, enquanto que as 691 restantes (86,4%) obtiveram uma nota acima ou abaixo daquela atribuída pelo grupo de professores na primeira avaliação. Esses números revelam uma pequena estabilidade das notas atribuídas às provas dos vestibulandos em duas correções independentes.

Para melhor avaliar essa variabilidade dos escores, realizou-se uma análise estatística dessas medidas.

A média dos 800 escores foi de 39,41 (52,6% do escore máximo de 75) e 40,38 (53,8%), para a primeira e segunda avaliação, respectivamente. Esses dados podem ser interpretados como traduzindo uma prova de dificuldade mediana.

A Tabela III mostra o resumo da análise de variância (um fator) para os escores da primeira (X_i) e segunda (Y_i) avaliação. Os dados da primeira avaliação indicam que não há diferença significativa ao nível de significância de 5% ($p < 0,05$) entre as médias dos escores

TABELA III – ANÁLISE DE VARIÂNCIA (UM FATOR) DOS ESCORES ATRIBUÍDOS NA PRIMEIRA E SEGUNDA CORREÇÃO A 20 GRUPOS DE 40 DISSERTAÇÕES. VESTIBULAR DA UFRGS E FFFCMPA, 1984

	Fonte de variação	Graus de liberdade	Quadrados médios	F
Primeira Avaliação	Entre grupos Intragrupos	19 780	294,99 258,51	1,14
Segunda Avaliação	Entre grupos Intragrupos	19 780	1415,62 185,59	7,63*

* Diferença estatisticamente significativa.

dos 20 grupos das populações selecionadas. Isso era esperado, uma vez que as 40 dissertações de cada grupo foram escolhidas aleatoriamente, ficando as eventuais diferenças por conta do acaso. De maneira análoga, para os escores da segunda avaliação, os dados da Tabela III revelam que as médias dos escores dos 20 grupos são significativamente diferentes. Isso mostra que o fator professor teve influência significativa na atribuição das notas da prova de dissertação, ou seja, diferentes avaliadores, embora orientados pelos mesmos critérios de correção, atribuem notas diferentes às dissertações.

Resta ainda examinar as médias e os desvios padrão dos escores de cada grupo nas duas correções, bem como a correlação entre esses escores, a partir dos dados apresentados na Tabela IV. Examinando os índices de correlação ($r_{x,y}$), verifica-se que todos são positivos, não muitos elevados e variam bastante (de 0,22 a 0,73) de grupo para grupo. Isso reflete uma pequena uniformidade das correções. Comparando as médias dos escores de cada grupo na primeira avaliação (\bar{X}) com a da segunda (\bar{Y}), usando o teste t para amostras dependentes, verifica-se que, para sete grupos, a diferença entre as médias é estatisticamente significativa ao nível de 5%.

Para comparar a dispersão dos escores de cada grupo nas duas avaliações, aplicou-se o teste estatístico t para interferências sobre variâncias, usando amostras dependentes. Os valores também encontram-se na Tabela IV e mostram que existem evidências para concluir que a hipótese de que as duas populações têm a mesma variância é rejeitada ao nível de 5% em sete casos. Isso revela que, para esses casos, deve ser descartada a possibilidade de admitir que a diferença entre dispersões dos escores das duas avaliações seja puramente casual, devendo-se atribuí-la à falta de homogeneidade na aplicação dos critérios utilizados pelos avaliadores durante as correções.

4 – CONCLUSÕES

O presente estudo compara os escores atribuídos às dissertações de 800 vestibulandos em duas correções sucessivas e independentes. Com base na análise dos dados, pode-se concluir que diferentes avaliadores, orientados para a adoção de um único critério de correção, tendem a atribuir notas diferentes às mesmas dissertações. Essa falta de uniformidade ocorre tanto para a sua média como para a sua dispersão.

É fácil depreender como a diferença entre as notas atribuídas a um trabalho por dife-

TABELA IV – ÍNDICE DE CORRELAÇÃO, MÉDIA, DESVIO PADRÃO E VALORES DE t PARA OS ESCORES ATRIBUÍDOS, EM DUAS CORREÇÕES, A DISSERTAÇÕES DO VESTIBULAR DA UFRGS E FFFCMPA, EM 1984.

Grupo	n	Correlação	Média		t	Desvio Padrão		t
		r_{xy}	\bar{X}	\bar{Y}		S_x	S_y	
A	40	0,45	39,375	42,50	-1,37	15,41	11,32	2,17*
B	40	0,63	40,125	43,125	-1,32	18,24	14,26	1,98
C	40	0,41	36,50	36,00	0,16	18,65	16,69	0,75
D	40	0,63	36,625	34,75	0,92	16,03	13,49	1,38
E	40	0,44	42,625	44,00	-0,50	18,15	14,29	1,65
F	40	0,40	39,625	32,25	3,00*	16,81	9,13	4,37*
G	40	0,44	37,75	42,00	-1,69	17,39	10,61	3,54*
H	40	0,55	44,00	52,125	-4,05*	13,97	12,55	0,79
I	40	0,34	38,375	37,375	0,39	14,61	13,40	0,57
J	40	0,30	38,125	28,375	4,15*	14,53	9,77	2,65*
K	40	0,26	42,00	40,25	0,66	13,39	14,14	-0,35
L	40	0,50	41,00	36,625	1,84	15,03	15,17	-0,06
M	40	0,44	44,625	40,25	1,84	14,82	13,68	0,55
N	40	0,53	33,25	39,875	-2,58*	16,89	16,70	0,08
O	40	0,73	37,875	35,00	1,66	16,01	12,61	2,18*
P	40	0,38	41,00	52,50	-4,73*	16,26	9,27	3,94*
Q	40	0,69	38,875	38,75	0,06	14,16	17,01	-1,57
R	40	0,52	40,375	43,875	-1,32	18,24	15,91	0,99
S	40	0,22	37,00	42,375	-2,34*	13,67	8,70	2,96*
T	40	0,56	39,00	45,625	-2,50*	17,62	18,12	-0,21

* Diferença estatisticamente significativa.

rentes avaliadores assume uma dimensão crítica em concursos de seleção, podendo a correção de uma dissertação por um ou outro avaliador representar a classificação ou desclassificação do candidato. Portanto, embora não se consiga eliminar a subjetividade da correção, principal responsável pelas flutuações dos escores atribuídos a provas dissertativas, é recomendável que essas notas resultem da média de várias avaliações, o que tende a torná-las mais precisas.

Diante disso, atribuir às questões dissertativas um papel decisivo nos concursos de seleção de candidatos a cursos superiores é muito imprudente. Uma prova dissertativa nunca deverá decidir sozinha o destino do vestibulando, não pela prova em si, mas pela falta de uniformidade nas correções, o que provoca um número considerável de candidatos selecionados ou eliminados pela sorte. Embora se reconheça que o uso dessas questões é extremamente válido nas escolas e, por isso, sua inclusão em concursos seja até recomendável, seu papel não deve predominar na classificação.

Por outro lado, pretender que a inclusão das questões dissertativas no vestibular venha melhorar o ensino e a aprendizagem de segundo grau, significa dar uma ênfase exagerada à dependência dessas áreas em relação ao vestibular. Os problemas da educação de primeiro e segundo graus são tantos e tão grandes, que a dimensão que o vestibular assume para a apren-

dizagem da maioria dos alunos é pequena ou insignificante. Essa pretensão também significaria uma tentativa de afastamento do vestibular da sua finalidade primordial: selecionar, dentre os candidatos, aqueles que apresentam um melhor nível de conhecimento sobre os conteúdos das matérias do segundo grau.

REFERÊNCIAS BIBLIOGRÁFICAS

VIANNA, H. M. (1984). *Comunicação e expressão: problemas teóricos e práticos de avaliação*, IBRASA, São Paulo.

