

Propostas para Mensurar o Grau de Não-Casualidade em Avaliações Objetivas*

LORÍ VIALI

Departamento de Estatística da Pontifícia Universidade Católica do Rio Grande do Sul e Universidade Federal do Rio Grande do Sul
viali@pucrs.br

HÉLIO RADKE BITTENCOURT

Departamento de Estatística da Faculdade de Matemática da Pontifícia Universidade Católica do Rio Grande do Sul, Brasil
heliorb@pucrs.br

Resumo

Neste texto são introduzidos três índices para avaliar o grau de não-casualidade em provas compostas unicamente por questões objetivas de múltipla escolha. Esse tipo de prova é largamente utilizado em concursos de todo tipo, sendo que uma das principais utilizações se dá em vestibulares, ou seja, em concursos destinados aos candidatos a vagas nas universidades brasileiras, sejam elas públicas ou privadas. Em provas objetivas, o número de acertos casuais pode ser previsto pelo Modelo Probabilístico Binomial significando que, mesmo para alunos que respondam todas as questões a esmo, existe um número esperado de acertos. A título de ilustração serão utilizados os resultados dos últimos oito anos do concurso vestibular da Universidade Federal do Rio Grande do Sul, Brasil. Os resultados mostram que os índices são indicadores capazes de medir a distância do número de acertos do grupo de vestibulandos em relação ao número esperado de acertos casuais.

Palavras-chave: acerto casual, provas objetivas, avaliação.

Resumen

El artículo presenta tres índices para evaluar el grado de no-casualidad en exámenes objetivos de opción múltiple. Esta categoría de examen es utilizada en muchas situaciones como, por ejemplo, cuando se aplican pruebas de selectividad para entrada en las universidades. Son exámenes objetivos, donde el número de respuestas correctas ocasionales puede ser estimado por el Modelo Probabilística Binomial. Esto significa que, mismo cuando los estudiantes marcan a todas las preguntas al azar hay un número previsto de respuestas correctas. A título de ilustración se presentan los resultados de los exámenes de selectividad aplicados por la Universidad Federal del Río Grande del Sur durante los últimos ocho años. Los resultados demuestran que los índices son indicadores capaces de medir la distancia entre el número de respuestas correctas del grupo de candidatos en relación con el número esperado de aciertos ocasionales.

Palabras-clave: acierto casual, pruebas objetivas, evaluación.

* Os autores agradecem à Comissão de Vestibular da Universidade Federal do Rio Grande do Sul (Coperse) por disponibilizar os resultados dos vestibulares de 1999 a 2006, na Internet, o que possibilitou a realização deste trabalho.

Abstract

This paper presents three indexes to evaluate the degree of non-randomness in objective assessments of the multiple-choice kind. This form of assessment is widely used in an array of competitive examinations, including the university entrance examinations of the main private and public universities in Brazil. In objective tests, the number of casual correct answers can be predicted by the Binomial Probability Model, which means that even for students who answer all of the questions at random, there is an expected number of correct answers. The results for the university entrance examinations for the Federal University of Rio Grande do Sul over the last 8 years will be used in order to illustrate this fact. The results of these tests show that the indexes are indicators capable of measuring the distance between the number of correct answers by the group of candidates and the expected number of random correct answers.

Key words: random guessing, assessment, evaluation.

1 INTRODUÇÃO

O ingresso de alunos às universidades brasileiras se dá por meio de um conjunto de provas, em geral objetivas, denominado de concurso vestibular, ou simplesmente vestibular. Esse concurso é composto, na maioria das instituições, por provas com questões de escolha múltipla. Os vestibulares das universidades públicas brasileiras, geralmente, são concorridos, em especial os cursos de Medicina, Odontologia e Direito. Cada instituição é responsável pela elaboração e execução de seu próprio concurso vestibular.

O número de questões pode variar entre as instituições, mas normalmente as provas seguem padrão semelhante. Excetuando-se a questão de redação, que é essencialmente descritiva e subjetiva, as demais são normalmente objetivas e constituídas de cinco alternativas. Há pelo menos dois motivos para o vestibular não contar com questões abertas: 1) facilidade de correção, pois normalmente existem milhares de concorrentes; 2) imparcialidade na correção, visto que provas descritivas permitem um alto componente subjetivo. Provas objetivas possibilitam acertos casuais, significando que, mesmo que um aluno tenha conhecimento nulo sobre o conteúdo abordado, ainda assim existe a possibilidade do acerto. Por essas razões, Vianna (1976) diz que testes objetivos costumam ser louvados por uns e criticados por outros.

A utilização de conjuntos de provas objetivas não é uma exclusividade das universidades. Muitas empresas públicas, autarquias, fundações, tribunais, forças armadas, estados, municípios e a própria união precisam, por força de lei, recrutar novos funcionários por meio de concursos públicos, fazendo uso de provas objetivas de múltipla escolha para selecionar seus funcionários. Entretanto, muito pouco tem sido feito ou realizado para verificar até que ponto as provas utilizadas nesses concursos estão calibradas, ou seja, adequadas ao nível de conhecimentos dos candidatos. A pergunta a ser feita é se essas provas de fato discriminam os candidatos pelo seu nível de conhecimento, ou permitem que candidatos com pouco ou nenhum conhecimento também sejam aprovados com base apenas no acaso e algumas dicas dos especialistas na preparação para esses concursos.

Nos tópicos subseqüentes será apresentada uma modelagem probabilística dos acertos casuais e serão propostos três índices de não-casualidade. Uma aplicação a ser considerada é a utilização dos índices propostos para avaliar as diversas provas dos últimos oito vestibulares da Universidade Federal do Rio Grande do Sul (UFRGS).

2 MODELAGEM DE ACERTOS CASUAIS

A Teoria da Probabilidade modela fenômenos não-determinísticos ou probabilísticos. O número de acertos casuais em uma prova objetiva é um desses fenômenos. Assim, em uma prova objetiva de, digamos, 30 questões, e cada questão tendo cinco alternativas, o número esperado de acertos, se um candidato responder a prova inteiramente ao acaso, ou seja, “chutar” toda a prova, será uma variável aleatória. Se essa variável for representada por X , então ela terá uma distribuição de Probabilidade Binomial de parâmetros $n = 30$ e $p = 1/5 = 0,20$. A Figura 1 mostra o gráfico da função de probabilidade da variável aleatória $X =$ número de acertos casuais, numa prova de 30 questões, e cada questão tendo cinco alternativas.

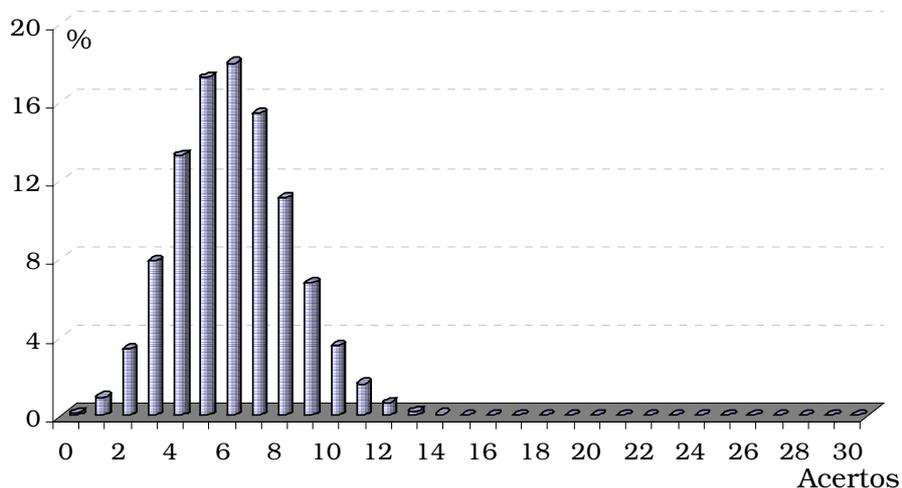
A distribuição binomial revela que, no caso de provas respondidas totalmente ao acaso, aproximadamente 90% dos vestibulandos apresentaria entre três e nove acertos. No concurso vestibular de 2006 da UFRGS, por exemplo, a prova de matemática teve a menor média dos últimos oito anos, com 68,7% dos candidatos apresentando entre três e nove acertos. A Tabela 1 mostra a proporção de candidatos com acertos dentro desta faixa.

Genericamente, se uma prova apresenta “ n ” questões objetivas e cada questão é composta de “ m ” alternativas, então a probabilidade de um número “ x ” de acertos casuais é dado por:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

onde $p=1/m$ e $q=1 - (1/m)$. A variável $X =$ número de acertos casuais varia no conjunto $\{0, 1, 2, \dots, n\}$. Valores altos terão probabilidade de acerto casual próximo de zero, assim como é pouco provável um número muito baixo de acertos casuais, conforme pode ser verificado na Figura 1.

Figura 1 – Modelagem do número de acertos casuais nas provas da UFRGS



O número esperado (médio) de acertos casuais é dado por $\mu = E(X) = np$. Neste caso como $p=1/m$, então o número de acertos casuais é de $\mu = n/m$. Assim, quanto maior o número de alternativas, mais baixa é a expectativa de acerto casual. Por exemplo, a UFRGS utiliza provas com $n = 30$ questões e com $m = 5$ alternativas cada questão. Neste caso, o número esperado (média) de acertos casuais é $\mu = E(X) = 30/5 = 6$ questões. Assim, se uma prova ficar muito acima do nível dos vestibulandos, a média da prova será baixa, e o número de vestibulandos que podem alcançar um escore razoável ao acaso tende a crescer.

Convém lembrar que a probabilidade de acerto casual em uma dada questão é de 20% se um candidato marcar umas das cinco opções inteiramente ao acaso. Como isso em geral não ocorre, pois, normalmente, o candidato lê a questão e, mesmo que ele pouco saiba do assunto, é bem provável que consiga eliminar uma, duas ou, às vezes, até três alternativas. Não é incomum então que ele tenha que decidir, em algumas situações, apenas entre duas alternativas. Quer dizer que é possível que a probabilidade de acerto seja de um em cinco (20%), de um em quatro (25%), de um em três (33%) ou no caso mais otimista de um em dois (50%). Nesse caso, o número esperado mínimo de acertos seria 6 e o máximo 15. Com tal suposição, estamos envolvendo quase todos os valores médios observados em praticamente todas as provas realizadas. Raramente uma prova tem uma média de acertos igual ou superior a 15. A Tabela 1

apresenta os valores médios de acertos de cada uma das provas que compõe o concurso vestibular da UFRGS no período de 1999/2006.

Tabela 1 – Valores médios de acertos nas provas da UFRGS no período 1999/2006

Prova	Período							
	1999	2000	2001	2002	2003	2004	2005	2006
Biologia	11,39	12,28	11,56	11,36	10,57	11,21	9,59	9,18
História	14,47	10,69	11,30	13,15	11,32	11,32	11,36	11,56
Matemática	10,31	9,82	8,77	9,04	8,90	10,15	8,52	8,18
Literatura	12,77	12,53	11,64	13,12	13,70	13,94	12,90	12,16
Língua Portuguesa	18,24	16,84	17,69	*	*	*	15,36	15,80
Química	9,03	10,89	10,34	11,13	12,53	8,98	9,62	8,94
Espanhol	16,85	14,89	15,97	18,18	16,53	16,01	14,24	16,11
Inglês	14,47	13,34	12,40	13,04	13,66	13,30	15,11	13,08
Física	9,03	11,03	9,42	8,96	10,16	10,23	9,72	10,94
Geografia	12,83	10,62	10,95	9,93	13,27	9,62	14,41	11,58

* Dados não disponíveis.

Pode-se observar que as provas que apresentam as médias mais altas são as de Língua Portuguesa (máximo = 18 e mínimo = 15) e Espanhol (máximo = 18 e mínimo = 14). Já as provas de Matemática (máximo = 10 e mínimo = 8) e Física (máximo = 11 e mínimo = 9) são as que apresentam os menores valores.

Para uma análise mais consistente do número de acertos casuais é necessário levar em conta não apenas o valor esperado, mas também a variabilidade em torno desse valor, isto é, o desvio-padrão (σ). Para o modelo que está sendo considerado o desvio-padrão é $\sigma = [p(1-p)]^{1/2}$. Logo, para as provas da UFRGS, o desvio-padrão do número de acertos casuais é de 2,19 questões. Ora, para n grande e p pequeno a distribuição binomial pode ser razoavelmente aproximada por uma curva normal. Dessa forma, o número de acertos casuais em cada prova pode variar de três desvios abaixo da média a três desvios acima da média, ou seja, entre zero e 12 acertos com probabilidade de 99,74% – praticamente a totalidade dos acertos casuais envolvidos. A Tabela 2 mostra a porcentagem de candidatos

com número de acertos situado entre zero e 12. Como era esperado, as provas com médias mais elevadas (Língua Portuguesa e Espanhol) apresentam a menor proporção de candidatos com número de acertos não superior a 12.

Tabela 2 – Porcentagem de candidatos com número de acertos entre zero e doze no concurso vestibular – UFRGS, 2006

<i>Disciplina</i>	<i>% de candidatos entre 0 e 12 acertos</i>
Biologia	80,2
História	62,0
Matemática	87,5
Literatura	55,7
Língua Portuguesa	25,2
Química	81,6
Espanhol	24,0
Inglês	50,2
Física	69,6
Geografia	62,4

Se a média de uma prova for de dez acertos, o que ocorre com certa frequência – veja-se as provas de Matemática e Física –, teríamos 6,11% dos candidatos acertando 10 ou mais questões, inteiramente ao acaso. Como o total de candidatos tem girado em torno de $N = 40000$, teríamos aproximadamente 2443 candidatos nessa situação. Se a média de uma prova cair para nove acertos, como ocorreu com Biologia e Química em 2006, então o número dos que acertariam nove ou mais questões seria de 5146. Se a média for de oito acertos, como ocorreu com Matemática, em 2006, então o número de candidatos que acertariam um valor igual à média ou maior seria de mais de 9000.

É claro que a probabilidade de acerto casual elevado em várias provas é baixa, mas a idéia aqui é avaliar a qualidade da prova, isto é, se ela, de fato, está discriminando um candidato preparado de um não preparado. A idéia é que, se uma prova for complicada o bastante, ou seja, tenha uma média próxima ao número de acertos casuais, então esta prova passa a ter o seu poder discriminativo comprometido. Nesse caso, mesmo

candidatos despreparados podem ter sucesso se responderem a todas as questões inteiramente ao acaso. Mas considerando que isso nem sempre ocorre, ou seja, que o candidato tende a aumentar as suas probabilidades de acerto em cada questão, pela eliminação de alternativas, então o problema passa a ter maior relevância, se for lembrado que a grande maioria dos candidatos fazem cursos pré-vestibular ou pré-concursos onde são treinados a aproveitarem todas as chances do concurso.

Silveira (1996, 1997) e Silveira e Pinent (2001) conduziram estudos sobre o poder discriminativo de questões em provas objetivas, bem como estudaram o efeito da prova de redação em concursos vestibulares; entretanto seus estudos não sugerem índices capazes de retratar a qualidade de uma prova excluindo o efeito dos acertos casuais. Para uma discussão sobre medidas educacionais e teste objetivos, consultar Vianna (1976).

3 ÍNDICES DE NÃO-CASUALIDADE

Neste item serão propostos três índices de não-casualidade, doravante denotados por INC, que buscam isolar o efeito dos acertos casuais de provas objetivas. O primeiro, denominado INC_1 , é baseado na discrepância entre a distribuição de probabilidade teórica para o número de acertos casuais e o número de acertos efetivamente observado. O segundo índice, denominado INC_2 , é inspirado no coeficiente de concordância Kappa, proposto por Cohen¹ (1960). O terceiro índice (INC_3) é baseado no coeficiente de correlação de Pearson entre as frequências esperadas pelo Modelo Binomial e as frequências observadas em cada prova.

Para o cálculo do índice de não-casualidade INC_1 será considerada a notação descrita no Quadro 1.

¹ Jacob Cohen (1923 – 1998). Psicólogo americano.

Quadro 1 – Notação utilizada para formulação e cálculo do índice INC₁

$N = \sum f_o$ = número total de candidatos na prova;

$f_e = N \cdot \binom{n}{x} p^x q^{n-x}$ = frequência esperada de acertos casuais para cada valor de x
no conjunto $\{ 0, 1, 2, \dots, 30 \}$;

$f_o - f_e$ = diferenças entre as frequências observadas e esperadas;

$|f_o - f_e|$ = número de candidatos que, supostamente, não acertaram uma dada questão por acaso;

$0,5\sum |f_o - f_e|$ = total de candidatos que, supostamente, não acertaram as questões por acaso.

O índice INC₁ assume valores no intervalo de 0 até 1, o que é altamente desejável em um indicador, sendo calculado por meio da seguinte expressão:

$$INC_1 = \frac{\sum f_o - 0,5\sum |f_o - f_e|}{\sum f_o} = 1 - \frac{\sum |f_o - f_e|}{2\sum f_o}$$

Dessa forma, quanto mais baixo for o valor do índice, maior é a concordância entre as frequências observadas e as esperadas de forma casual. Por outro lado, provas mais fáceis apresentam um índice INC₁ mais próximo de um, indicando que o número de candidatos que acertaram a prova por acaso é pequeno.

O índice de não-casualidade INC₂ é inspirado no coeficiente de concordância de Kappa que foi inicialmente utilizado para quantificar a concordância entre juízes no campo da Psicometria. Neste coeficiente, o autor separa a concordância em duas partes: uma casual e outra denominada “concordância confiável”. O coeficiente Kappa, portanto, procura mensurar a intensidade ou a qualidade da concordância, assim como este trabalho pretende medir a qualidade de uma prova objetiva, isolando o efeito de acertos casuais. Descrições do coeficiente Kappa e de sua utilização podem ser encontradas em Everitt (1992) e Ludbrook (2002). O Quadro 2 mostra a expressão de cálculo do coeficiente INC₂ e a sua analogia com o coeficiente Kappa.

O numerador do coeficiente INC_2 representa a comparação entre a média observada de acertos da prova e o número esperado de acertos casuais, revelado pela distribuição Binomial. O denominador representa a máxima concordância possível, descontando-se os acertos casuais.

Quadro 2 – Coeficiente INC_2 e sua analogia com o coeficiente Kappa de Cohen

<i>Coeficiente Kappa (k)</i>	<i>Coeficiente de não-casualidade 2 (INC_2)</i>
$k = \frac{p_o - p_e}{1 - p_e}, \text{ onde}$ <p>p_o = concordância observada p_e = concordância esperada (casual)</p>	$INC_2 = \frac{m_o - m_e}{n - m_e}, \text{ onde}$ <p>m_o = média observada na prova m_e = média esperada pela Binomial (n/m) n = número de questões (máximo de acertos)</p>

Com o objetivo de tornar a interpretação do coeficiente INC_2 mais fácil e direta, optou-se pela sugestão de parâmetros de referência, tendo como base a tabela de Landis e Koch (1977). O Quadro 3 apresenta uma escala de interpretação sugerida.

Quadro 3 – Escala de interpretação para o coeficiente INC_2 inspirada em Landis e Koch (1977)

<i>Coeficiente INC_2</i>	<i>Poder discriminatório</i>	<i>Acertos casuais</i>
< 0	Muito pequeno	Média de acertos inferior ao número esperado de acertos casuais
Entre 0,00 e 0,20	Pequeno	Média de acertos praticamente igual ao número esperado de acertos casuais
Entre 0,21 e 0,40	Pequeno/Moderado	Média de acertos pouco superior ao número esperado de acertos casuais
Entre 0,41 e 0,60	Moderado/Grande	Média de acertos superior ao número esperado de acertos casuais
Entre 0,61 e 0,80	Grande	Média de acertos muito superior ao número esperado de acertos casuais
Entre 0,81 e 1,00	Muito Grande	Média de acertos muitíssimo superior ao número esperado de acertos casuais

Coeficientes INC_2 inferiores a zero são raros, assim como coeficientes próximos do valor máximo. A faixa desejável de valores do coeficiente se situa entre 0,21 e 0,60, com valores supostamente ideais na casa do 0,40. À medida que a distribuição do número de acertos se aproxima da curva normal, o coeficiente INC_2 tende a se posicionar próximo de 0,50.

O coeficiente INC_3 é obtido por meio do coeficiente de correlação de Pearson (R) entre as frequências esperadas pelo modelo binomial e o número observado de candidatos para cada acerto possível. Com a finalidade de deixar o indicador INC_3 com variação fixa no intervalo de zero até um, com o número zero indicando baixo poder discriminativo, foi realizada a seguinte transformação sobre o coeficiente R de Pearson:

$$INC_3 = \frac{1 - R}{2}$$

4 A INTERPRETAÇÃO DOS COEFICIENTES

Após a exposição de três diferentes formas de mensurar a não-casualidade em provas objetivas, passamos à interpretação dos coeficientes. Ressalta-se que todos os três coeficientes propostos variam na mesma direção com o número zero indicando baixo poder discriminativo da prova, ou seja, grande possibilidade do número de acertos casuais se assemelhar ao número de acertos por convicção do candidato.

A título de sugestão, propomos o Quadro 4 como referência para interpretação dos três índices, ainda seguindo o modelo de Landis e Koch (1977) utilizado no Quadro 3. Apesar de haver grande associação entre os índices, cabe ressaltar que pode ocorrer uma pequena discordância na classificação. Caso isso ocorra, sugere-se optar pelo índice INC_1 que, além de ser um índice robusto, apresentou resultados práticos satisfatórios.

Comparando os índices INC_1 e INC_3 com INC_2 , percebe-se que este último necessita apenas do conhecimento da média, número de questões e número de alternativas por questão para o seu cálculo. Para o cálculo dos índices INC_1 e INC_3 o pesquisador necessariamente deve ter acesso à distribuição de frequências do número de acertos.

Quadro 4 – Escala de interpretação sugerida para os coeficiente INC_1 , INC_2 e INC_3

<i>Poder discriminatório</i>	INC_1	INC_2	INC_3
Muito pequeno	0	< 0	0
Pequeno	Entre 0,01 e 0,50	Entre 0,00 e 0,20	Entre 0,01 e 0,25
Pequeno/Moderado	Entre 0,51 e 0,80	Entre 0,21 e 0,40	Entre 0,26 e 0,45
Moderado/Grande	Entre 0,80 e 0,90	Entre 0,41 e 0,60	Entre 0,46 e 0,75
Grande	Entre 0,90 e 0,99	Entre 0,61 e 0,80	Entre 0,76 e 0,95
Muito Grande	1	Entre 0,81 e 1,00	> 0,95

5 RESULTADOS

Tendo por objetivo testar os três coeficientes propostos foram analisados os resultados dos últimos oito anos dos vestibulares da UFRGS. No último concurso vestibular, em janeiro de 2006, realizaram a prova de Língua Portuguesa 32993 candidatos, número que não se mantém constante nas demais provas devido às desistências. Nas provas de idiomas há uma divisão da população, em razão da possibilidade de opção entre espanhol ou inglês. A Tabela 3 resume os resultados encontrados para os índices no ano de 2006.

De acordo com os índices calculados, as provas do vestibular da UFRGS apresentam graus de dificuldade de moderado a grande; em algumas provas (Matemática e Física) a média de acertos está próxima do número esperado de acertos casuais.

As Tabelas 4, 5 e 6 apresentam os resultados dos índices INC_1 , INC_2 e INC_3 nos últimos oito anos (1999/2006). As provas de Matemática, Física e Química são tradicionalmente as mais difíceis da UFRGS, seguidas de perto pela prova de Biologia. As provas relacionadas à área de literatura e letras apresentam os menores graus de dificuldade e, conseqüentemente, os maiores valores do indicador. As provas de História e Geografia apresentam índices intermediários.

O índice INC_1 varia no intervalo [0; 1]. Quanto mais próximo de zero ele for maior será o grau de dificuldade da prova que ele está avaliando. O gráfico de acertos, nessa prova, será semelhante ao gráfico dos acertos casuais, Figura 1, cuja média é de seis acertos com um desvio-padrão de 2,19 acertos. A prova de Matemática de 2006, por exemplo,

apresentou média de 8,18 acertos, ficando apenas cerca de um desvio-padrão acima da média esperada dos acertos casuais. Por esse motivo, o seu INC_1 foi baixo (0,266).

Tabela 3 – Índices de não-casualidade INC_1 , INC_2 e INC_3 das provas da UFRGS – 2006

<i>Prova</i>	INC_1	INC_2	INC_3	<i>Poder discriminatório</i>
Biologia	0,354	0,133	0,088	Pequeno
História	0,589	0,232	0,321	Pequeno/Moderado
Matemática	0,266	0,091	0,044	Pequeno
Literatura	0,688	0,257	0,458	Pequeno/Moderado
Língua Portuguesa	0,835	0,408	0,675	Moderado/Grande
Química	0,304	0,122	0,053	Pequeno
Espanhol	0,841	0,421	0,687	Moderado/Grande
Inglês	0,631	0,295	0,365	Pequeno/Moderado
Física	0,523	0,206	0,245	Pequeno/Moderado
Geografia	0,627	0,232	0,378	Pequeno/Moderado

Observando a Tabela 3, pode-se verificar que as provas que apresentam os mais altos índices de não-casualidade são Língua Portuguesa e Espanhol, sendo, portanto, as provas mais discriminativas já que é improvável que um candidato faça bons escores respondendo apenas ao acaso. Já as provas de Matemática, Química e Biologia são as que menos discriminam um candidato preparado de um candidato que realiza escolhas ao acaso, possuindo índices baixos de não-casualidade.

Tabela 4 – Índices de não-casualidade INC_1 das provas da UFRGS – 1999/2006

<i>Prova</i>	<i>Período</i>							
	1999	2000	2001	2002	2003	2004	2005	2006
Biologia	0,594	0,657	0,573	0,583	0,548	0,609	0,427	0,354
História	0,793	0,521	0,606	0,694	0,617	0,584	0,591	0,589
Matemática	0,498	0,375	0,3109	0,382	0,318	0,471	0,278	0,256
Literatura	0,672	0,720	0,6419	0,7072	0,726	0,784	0,727	0,688
Língua Portuguesa	0,905	0,861	0,8795	*	*	*	0,841	0,835
Química	0,374	0,519	0,450	0,477	0,660	0,314	0,421	0,304
Espanhol	0,859	0,812	0,859	0,923	0,857	0,858	0,802	0,841
Inglês	0,673	0,707	0,614	0,595	0,599	0,665	0,707	0,631
Física	0,296	0,510	0,353	0,337	0,456	0,434	0,374	0,523
Geografia	0,716	0,533	0,575	0,544	0,734	0,451	0,799	0,627

* Dados não disponíveis.

Os dados também revelam que, nos últimos oito anos, nenhuma prova apresentou indicador INC_1 maior que 0,90 ou INC_2 maior que 0,51, ou seja, de maneira geral as provas têm baixo ou moderado poder discriminatório. De acordo com as Tabelas 4 e 5, nenhuma prova apresentou poder discriminatório muito grande.

Em relação ao INC_3 , isto é, o indicador baseado no coeficiente de correlação entre as frequências observadas e as esperadas, pode-se observar que a amplitude de variação também se dá no intervalo [0; 1]. Em comparação ao INC_2 este índice apresenta uma maior amplitude de variação, o que pode significar maior sensibilidade do índice, o que é uma característica desejável. Com base neste indicador, uma prova (Língua Portuguesa – 1999) apresentou poder discriminatório grande.

Tabela 5 – Índices de não-casualidade INC₂ das provas da UFRGS – 1999/2006

<i>Prova</i>	<i>Período</i>							
	1999	2000	2001	2002	2003	2004	2005	2006
Biologia	0,225	0,262	0,232	0,223	0,190	0,217	0,150	0,133
História	0,354	0,195	0,221	0,298	0,222	0,222	0,224	0,232
Matemática	0,180	0,159	0,116	0,127	0,121	0,173	0,105	0,091
Literatura	0,282	0,272	0,235	0,297	0,321	0,331	0,288	0,257
Língua Portuguesa	0,510	0,452	0,487	*	*	*	0,390	0,408
Química	0,126	0,204	0,181	0,214	0,272	0,124	0,151	0,122
Espanhol	0,452	0,371	0,416	0,508	0,439	0,417	0,343	0,421
Inglês	0,353	0,306	0,267	0,293	0,319	0,304	0,380	0,295
Física	0,126	0,210	0,143	0,123	0,173	0,176	0,155	0,206
Geografia	0,284	0,193	0,206	0,164	0,303	0,151	0,351	0,232

* Dados não disponíveis.

Tabela 6 – Índices de não-casualidade INC₃ das provas da UFRGS – 1999/2006

<i>Prova</i>	<i>Período</i>							
	1999	2000	2001	2002	2003	2004	2005	2006
Biologia	0,333	0,411	0,298	0,316	0,291	0,357	0,160	0,088
História	0,611	0,243	0,349	0,472	0,373	0,320	0,328	0,321
Matemática	0,229	0,081	0,067	0,129	0,061	0,201	0,040	0,044
Literatura	0,438	0,499	0,400	0,495	0,523	0,590	0,513	0,458
Língua Portuguesa	0,764	0,713	0,749	*	*	*	0,668	0,675
Química	0,120	0,242	0,167	0,172	0,414	0,059	0,151	0,053
Espanhol	0,720	0,637	0,689	0,758	0,700	0,693	0,611	0,687
Inglês	0,429	0,496	0,343	0,297	0,276	0,428	0,492	0,365
Física	0,045	0,212	0,080	0,088	0,181	0,148	0,088	0,245
Geografia	0,499	0,264	0,316	0,300	0,531	0,192	0,614	0,378

* Dados não disponíveis.

Para verificar a coerência entre as três medidas propostas, foi calculada a matriz de correlação entre os três indicadores. A Tabela 7 apresenta os coeficientes R de Pearson entre os três índices nos oito anos. Pode-se perceber que eles cumprem o seu papel uma vez que essas correlações são todas bastante altas com todos os valores oscilando em torno de 0,90.

Tabela 7 – Coeficientes de correlação R de Pearson entre os três índices

<i>Pares</i>	<i>Período</i>							
	1999	2000	2001	2002	2003	2004	2005	2006
INC ₁ e INC ₂	0,951	0,950	0,952	0,951	0,939	0,968	0,962	0,966
INC ₂ e INC ₃	0,965	0,963	0,974	0,933	0,895	0,976	0,965	0,978
INC ₁ e INC ₃	0,994	0,998	0,990	0,989	0,983	0,995	0,996	0,987

A Tabela 8 apresenta as correlações entre os indicadores e as médias das provas. As altas magnitudes observadas nestas correlações mostram que, quanto mais baixa é a média observada em uma prova, menores serão os valores dos três indicadores. Em virtude da formulação do coeficiente INC₂ (uma função linear da média), o mesmo apresenta correlação perfeita com a média.

Tabela 8 – Coeficientes de correlação entre os índices e as médias das provas

<i>Índices</i>	<i>Período</i>							
	1999	2000	2001	2002	2003	2004	2005	2006
INC ₁	0,951	0,950	0,949	0,951	0,939	0,968	0,962	0,966
INC ₂	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
INC ₃	0,965	0,963	0,969	0,933	0,895	0,976	0,965	0,978

6 CONSIDERAÇÕES FINAIS

De acordo com Lindeman (1974), uma prova bem elaborada deve conter questões fáceis, médias e difíceis. As provas da UFRGS que apresentaram os valores mais baixos dos indicadores – Matemática, Química e Física – revelaram que o concurso vestibular pode não estar respeitando esta premissa. Neste caso, torna-se difícil diferenciar um aluno mediano de um aluno que escolhe as alternativas no palpite. Em cursos onde uma vaga é bastante disputada, como Medicina (40 candidatos por vaga em 2006), este problema não é grave, visto que o aluno necessita de um número muito elevado de acertos para ser aprovado. Já em cursos pouco disputados, como Estatística ou Física (aproximadamente três candidatos por vaga em 2006), o problema torna-se mais grave, pois o baixo poder discriminativo das provas dá alta probabilidade de aprovação a um candidato que responde “a esmo” grande parte da prova.

É notório que os vestibulares de universidades públicas federais brasileiras não são considerados concursos “fáceis”; na UFRGS, por exemplo, no ano de 2006, 40816 candidatos disputaram 4212 vagas. Portanto, os baixos valores dos índices não devem surpreender o leitor.

Para evitar que candidatos despreparados possam levar vantagem sobre candidatos que acertam por conhecimento, os autores sugerem que seja empregado algum tipo de penalidade pela escolha aleatória. Alguns concursos brasileiros com questões do tipo “verdadeiro ou falso” estão subtraindo os erros dos acertos, dando a possibilidade do candidato deixar questões em branco.

REFERÊNCIAS BIBLIOGRÁFICAS

COHEN, Jacob. A coefficient of agreement for nominal scales. *Educational Psychology*, n.20, p.27-46, 1960.

EVERITT, Brian. *Analysis of Contingency Tables*. London: Chapman and Hall, 1992, 168p.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, n. 33, p.159-174, 1977.

LINDEMAN, Richard. *Medidas Educacionais*. Porto Alegre: Globo, 1974. 175p.

LUDBROOK, John. Statistical techniques comparing measurers and methods of measurement: a critical review. *Clinical and Experimental Pharmacology and Physiology*, n.29, p.527-536, 2002.

SILVEIRA, Fernando. Relação de desempenho no concurso vestibular da UFRGS com diversas variáveis. *Fascículos Prograd 8*, Porto Alegre, p.5-28, 1996.

_____. Comparação entre três argumentos de concorrência para o vestibular da UFRGS. *Estudos em Avaliação Educacional*, n.10, p.43-57, 1997.

SILVEIRA, Fernando; PINENT, Carlos. A Questão da Redação no concurso vestibular à universidade: validade e poder decisório. *Estudos em Avaliação Educacional*, n.24, p.147-164, 2001.

VIANNA, Heraldo. *Testes em Educação*. 2.ed. Rio de Janeiro: Fename, 1976. 220 p.

Recebido em: junho 2006

Aprovado para publicação em: novembro 2006