



Educação & Sociedade

ISSN: 0101-7330

revista@cedes.unicamp.br

Centro de Estudos Educação e Sociedade

Brasil

Cordeiro, Leonardo

SOBRE A INADEQUAÇÃO DA METODOLOGIA DE CÁLCULO DAS NOTAS DO SISU

Educação & Sociedade, vol. 35, núm. 126, enero-marzo, 2014, pp. 293-320

Centro de Estudos Educação e Sociedade

Campinas, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=87330638017>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais artigos
- ▶ Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe , Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# SOBRE A INADEQUAÇÃO DA METODOLOGIA DE CÁLCULO DAS NOTAS DO SISU

LEONARDO CORDEIRO\*

**RESUMO:** A adoção da Teoria de Resposta ao Item (TRI) no cálculo das notas do Exame Nacional do Ensino Médio (Enem) tem sido apontada como o fator principal de dúvidas, questionamentos e até protestos pela parte da população afetada por esse processo. Embora a introdução, no contexto brasileiro, de um método mais moderno de avaliação educacional marque um necessário avanço metodológico, a utilização corrente da TRI na seleção para o ensino superior é fundamental e conceitualmente equivocada. Isso ocorre porque as habilidades estimadas pela TRI, definidas em escalas específicas, são diretamente combinadas à nota da redação, que é obtida em escala diferente. Dessa forma, o método utilizado pelo Sistema de Seleção Unificada (Sisu/MEC) gera distorções determinantes no processo de seleção. Esses desvios fazem com que até 76% dos candidatos que seriam aprovados, caso não houvesse essa mistura equivocada de notas, não mais o fossem.

**Palavras-chave:** Enem. Ensino médio. TRI. Avaliação educacional.

## ON THE INADEQUACIES OF SISU'S GRADES COMPUTATIONAL METHODOLOGY

**ABSTRACT:** The adoption of the IRT (item response theory) in the National High School Exam (Enem) has generated doubt and complaints from the student population. Even though this type of measurement marks a necessary and salutary step towards a more modern educational assessment culture, the use of the IRT in university entrance examinations is fundamentally flawed. This is a direct consequence of the fact that the IRT grades are combined with essay marks which are obtained through a completely different measurement scale. As such, the current method of combining these incompatible numerical grades generates a significant distortion in the final ranking of approved students. Up to 76% of students that would have been approved if the standard methodology were used fail to do so and have to try again in the following year.

**Key words:** Enem. High school. IRT. Educational assessment.

\* Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Rio de Janeiro (RJ) – Brasil.  
Contato com o autor: <lcordeiro@iag.puc-rio.br>

**SUR L'INADÉQUATION DE LA MÉTHODOLOGIE  
DE CALCUL DES NOTES DU SISU**

**RÉSUMÉ:** L'adoption de la Théorie de Réponse à l'Item (TRI) dans le calcul des notes de l'Examen National de l'Enseignement secondaire (Enem) a été indiquée comme le facteur principal de doutes, questionnements et jusqu'à des protestations de la partie de la population touchée par cette procédure. Bien que l'introduction, dans le contexte brésilien, d'une méthode plus moderne d'évaluation scolaire marque une nécessité de l'avance méthodologique, l'utilisation courante de TRI pour l'accès à l'enseignement supérieur est fondamentale et conceptuellement confondue. Cela se produit parce que les habilités demandées par la TRI sont définies dans des échelles spécifiques, directement attachées à la note de la rédaction, celle qui est obtenue dans une échelle différente. De cette forme, la méthode utilisée par le Système de Sélection Unifiée (Sisu/MEC) produit des déformations déterminantes dans le processus. Ces détours font que jusqu'à 76% des candidats qui seraient approuvé, au cas où il n'y aurait pas ce mélange confondu de notes, ne seront plus.

*Mots-clés:* Enem. Enseignement secondaire. TRI. Évaluation scolaire.

## Introdução

**E**m 1998, o Ministério de Educação e Cultura (MEC) iniciou um programa de avaliação do ensino médio em nível nacional, por meio da aplicação do Exame Nacional do Ensino Médio (Enem) a alunos concludentes do ensino médio (EM). O Enem é estruturado em duas partes: uma objetiva e multidisciplinar, composta por questões de múltipla escolha, que versam sobre o conteúdo típico do EM, e uma subjetiva, composta por uma redação em Língua Portuguesa. Embora o objetivo inicial do exame não contemplasse a seleção de alunos para o ensino superior, e sim a avaliação do rendimento escolar no EM, entre outros,<sup>1</sup> as notas obtidas no exame começaram a, paulatinamente, ser utilizadas para esse fim por algumas instituições de ensino superior (IES) (CASTRO; TIEZZI, 2005). Em 2009, o MEC propôs que o Enem fosse utilizado como instrumento unificado de seleção para as IES federais.<sup>2</sup>

Atualmente, o Enem é largamente empregado como instrumento de seleção para diversas IES públicas e privadas. Em algumas IES públicas, o Enem é a única forma de acesso ao ensino superior.<sup>3</sup> O Sistema de Seleção Unificada (Sisu) é o processo de seleção às vagas oferecidas pelas IES públicas e é administrado pelo MEC.

Observa-se, por conseguinte, a enorme importância que o Enem passou a ter no destino de milhões de jovens e candidatos a uma vaga no ensino superior e, consequentemente, na competitividade nacional futura. Ademais, a indicação de que a

nota do Exame substituirá a da Prova Brasil no cálculo do Índice de Desenvolvimento da Educação Básica (Ideb)<sup>4</sup> – tornando-se, portanto, instrumento fundamental das políticas educacionais do MEC – amplia drasticamente o impacto dessa medida na educação brasileira.

Nessa perspectiva, torna-se fundamental que se faça uma análise crítica da adequação da metodologia de cálculo das médias finais que irão decidir quem será agraciado, ou não, com uma oportunidade de ingressar em uma IES e, futuramente, qual escola será considerada “boa” ou “ruim”.

O objetivo principal desse trabalho é preencher essa lacuna. De especial interesse é a análise das consequências da decisão do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC) de combinar o nível de proficiência produzido pela TRI – uma “nota” relativa, resultado de procedimentos matemáticos complexos – com a nota da redação – uma medida absoluta e que não recebe qualquer tratamento estatístico, em contraste com o procedimento tradicional, aquele em que o percentual de acertos é combinado à nota da redação, de forma consistente e coerente.

Este trabalho está organizado da seguinte forma: após a Introdução, são apresentados aspectos importantes da TRI e da sua utilização pelo MEC/Inep no Enem. A seguir, o procedimento de obtenção da amostra é descrito, assim como a metodologia de análise. A quarta seção discorre sobre os resultados obtidos e reflete sobre suas possíveis implicações para a seleção de candidatos ao ensino superior. Finalmente, a quinta seção conclui o trabalho.

## TRI, Inep, Enem e Sisu

Ao longo do período de aplicação do Enem, o cálculo das notas dos candidatos sofreu importante alteração metodológica, capitaneada pelo Inep, autarquia federal responsável pelo exame. De 1998 a 2008, a teoria de medição educacional empregada foi a Teoria Clássica dos Testes (TCT) e o desempenho dos candidatos na prova objetiva foi medido pelo percentual de acertos, normalizado na escala [0, 100]. Tais notas eram combinadas às notas da redação para a obtenção de uma nota média final.

A partir de 2009, a Teoria de Resposta ao Item (TRI) começou a ser utilizada no cálculo das notas da parte objetiva do exame. A nota da redação continuou a ser obtida da mesma forma como no período “pré-TRI”, sendo combinada com as notas geradas pela TRI para a obtenção da nota final do candidato.

Outra característica marcante da TRI é que, caso as premissas do modelo sejam satisfeitas, os parâmetros referentes aos itens de uma determinada prova são

universais e não dependem dos candidatos. Da mesma forma, as estimativas de proficiência não dependem do conjunto de itens. Em outras palavras:

Uma propriedade importante da TRI é a de invariância dos parâmetros, isto é, os parâmetros dos itens obtidos de grupos diferentes de alunos testados e os parâmetros de proficiência baseados em grupos diferentes de itens são invariantes, exceto pela escolha de origem e escala. (KLEIN; FONTANIVE, 1995, p. 31)

Assim, uma vez conhecidas as características dos itens – geralmente estabelecidas pelos infames “pré-testes”<sup>5</sup> – estes se tornam o centro da análise, em vez da prova como um todo, antigo foco da TCT. As vantagens dessa abordagem são assim resumidas na literatura:

Ao escolher o item [...] como unidade de análise e peça de reposição intercambiável na elaboração das provas e na elaboração de seus resultados, a Teoria de Resposta ao Item incentivou inúmeras inovações na prática da aplicação das provas, especialmente no desenvolvimento no banco de itens e testes adaptativos. (FLETCHER, 1994, p. 23.)

Evidencia-se, portanto, que a principal vantagem da TRI é a possibilidade de comparação entre resultados obtidos em provas distintas, aplicadas em grupos distintos de alunos, fato que é possibilitado pela *escala comum entre os itens e pela invariância populacional dos parâmetros dos mesmos*.

Coerentemente com o defendido por Hambleton e Swaminathan (1985), o Inep assim justifica a utilização da TRI no cálculo das notas das provas objetivas do Enem a partir de 2009 (KARINO; ANDRADE, 2011, p. 2):<sup>6</sup> “A decisão de implementar no Exame Nacional do Ensino Médio (Enem) a Teoria de Resposta ao Item (TRI) teve duas finalidades principais: (1) permitir a comparabilidade dos resultados entre os anos e (2) permitir a aplicação do Exame várias vezes ao ano”.

Sem dúvida, a TRI permite que tais objetivos sejam alcançados, se os seus pressupostos teóricos forem verdadeiros, o que não seria possível caso o percentual de acertos continuasse a ser empregado, tendo em vista as limitações da TCT.

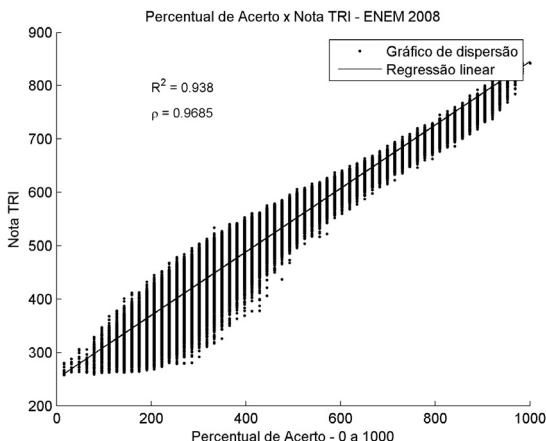
Uma revisão detalhada sobre a TRI foge ao escopo deste trabalho<sup>7</sup> e detalhes técnicos sobre essa metodologia podem ser obtidos em referências clássicas sobre o assunto, como os livros escritos por Lord (1980), Hambleton e Swaminathan (1985), Hambleton, Swaminathan e Rogers (1991) e Embretson e Reise (2000). Como representante nacional, o trabalho de Andrade, Tavares e Valle (2000) é um bom exemplo de uma apresentação técnica do tema.

No âmbito nacional, a mudança para a TRI, em si, não implica, necessariamente, uma alteração tão radical no resultado final de seleção, considerando-se o

percentual de acertos como medida referencial. A Figura 1 é o resultado da superposição do gráfico de dispersão entre as variáveis “percentual de acertos” e “nota TRI”<sup>8</sup> com a reta obtida pela regressão linear da segunda variável pela primeira. Percebe-se, visual e nitidamente, que as variáveis são fortemente relacionadas, o que é também comprovado pelo valor elevado do coeficiente de determinação ( $R^2$ ) entre elas. Isso indica que um alto percentual de acertos implica uma alta nota TRI, relativamente aos demais candidatos. A forte relação entre as variáveis ajudaria a dirimir algumas dúvidas dos candidatos em relação à nota TRI.<sup>9</sup>

**Figura 1**

Gráfico de dispersão entre o percentual de acertos e a nota TRI, sobreposta à reta obtida pela regressão linear da nota TRI (variável dependente) pelo percentual de acertos (variável independente). O coeficiente de determinação ( $R^2$ ) e a correlação linear entre as variáveis ( $r$ ) também se encontram na figura.



Por exemplo, um candidato que tenha acertado 80% do caderno amarelo do Enem de 2008 teria recebido uma nota TRI que variaria entre 680 e 730, aproximadamente, dependendo de quais questões ele acertou e errou. O fato de a nota TRI ser numericamente menor não altera significativamente a posição relativa do candidato em relação aos demais. É importante salientar que todas as notas dos mais de 700 mil candidatos<sup>10</sup> estão marcadas na figura. Tomando-se um aluno com um percentual de acertos menor, igual a 30%, por exemplo, percebe-se que o intervalo de variação possível da nota TRI é maior, cerca de 290 a 500. Contudo, a nota TRI ainda preserva o seu valor relativo, ou seja, em média, esse estudante receberá uma nota TRI menor que a de outro que acertou 40% da prova. Ainda, o candidato de alto desempenho, aquele com um percentual de 90% de acertos, muito provavelmente manterá a sua posição relativa, pois a sua nota TRI – que pode variar de 720 a 820 – continuará

sendo uma das mais altas da amostra. Ou seja, ater-se ao valor nominal da nota TRI é contraproducente, uma vez que, para fins de seleção educacional, a classificação relativa é o fator crucial, não a nota numérica em si. Percebe-se, adicionalmente, que a dispersão ao redor da reta de regressão é maior na região de percentuais de acerto menores, onde a relação linear entre as variáveis é mais fraca. Entretanto, de forma geral, a aproximação linear aparenta ser razoavelmente satisfatória.

As notas obtidas pela metodologia TRI são utilizadas pelo MEC/Inep no Sisu, compondo a média final que definirá a classificação final em cada curso. Dependendo do curso e da IES, o cálculo da média final varia. Em cursos em que não há predominância da nota de uma disciplina sobre a de outra, a média final é computada pela média aritmética simples.

$$\text{Média Aritmética} = \frac{\text{NAT} + \text{HUM} + \text{LIN} + \text{MAT} + \text{REDAÇÃO}^{11}}{5}$$

Alguns cursos, no entanto, elegem pesos diferentes para cada disciplina, como é o caso da Universidade Federal do Rio de Janeiro (UFRJ) no concurso de admissão para os cursos de graduação em 2012. Para o curso de Medicina, a média final é calculada da seguinte forma:

$$\text{Média Medicina} = \frac{\text{NAT} \times 2 + \text{HUM} \times 1 + \text{LIN} \times 2 + \text{MAT} \times 1 + \text{REDAÇÃO} \times 3}{9}$$

Similarmente, a média para os diversos cursos de Engenharia é assim calculada:

$$\text{Média Engenharia} = \frac{\text{NAT} \times 4 + \text{HUM} \times 1 + \text{LIN} \times 1 + \text{MAT} \times 4 + \text{REDAÇÃO} \times 3}{13}$$

Os pesos atribuídos à nota da redação variam da seguinte maneira: 20.0%, 23.1% e 33.3%, na média aritmética, de Engenharia e de Medicina, respectivamente.

Assim, apesar de o Inep empregar uma metodologia moderna no cálculo da nota objetiva do Enem, a média final do candidato inclui a nota da redação sem qualquer ajuste para colocá-la na mesma escala da prova objetiva. *Essa prática vai de encontro ao objetivo fundamental de se usar a TRI que é, em primeiro lugar, o de produzir notas comparáveis ao longo do tempo e entre diferentes populações.* Empregando os exemplos típicos encontrados na literatura de TRI, o Inep, dessa forma, soma centímetros com polegadas ou graus Celsius com graus Kelvin. Se, ainda, essas notas forem utilizadas para avaliar a “qualidade” das escolas brasileiras, corre-se o risco

de se exacerbar ainda mais os potenciais efeitos nocivos que o uso inapropriado de indicadores quantitativos pode exercer na educação (FERRÃO, 2012).

## Metodologia

Em linhas gerais, desenvolveu-se uma metodologia para avaliar se as notas geradas pela TRI na parte objetiva, relativas, podem ser combinadas com as notas da redação, de caráter estritamente absoluto, sem causar mudanças drásticas no resultado final da seleção. Essa avaliação é feita comparando-se a classificação dos alunos em cenários distintos de cálculo das médias finais, e aferindo-se a diferença do número de alunos aprovados, em relação a uma classificação padrão, que corresponde ao grupo de controle.

O problema possui duas dimensões: o instrumento de medida utilizado para se aferir as notas das provas objetivas e a forma de obtenção das notas finais dos candidatos. A primeira dimensão corresponde à escolha entre o percentual de acertos e a TRI, e a segunda, entre a utilização, ou não, da redação na composição da nota final. A Figura 2 ilustra, graficamente, essa problemática.

**Figura 2**

Diagrama com os quatro cenários possíveis da “Problemática Sisu”.

		Instrumento de medida	
		% de Acertos	TRI
Obtenção da nota final	Sem redação	% de Acertos	TRI
	Com redação	% de Acertos + Redação	TRI + Redação

Se o método adotado pelo Inep for adequado, isto é, se for válida a combinação da nota TRI com a nota da redação, espera-se que a classificação final, nesse caso, não difira muito da classificação obtida se somente a nota TRI fosse utilizada, pois o

único fator que se alterou foi a inclusão da nota da redação. Similarmente, a classificação decorrente da média calculada a partir do percentual de acertos com a redação não pode apresentar muitos desvios, quando comparada à obtida com o uso apenas dos percentuais de acerto. As variações relevantes estão representadas pelas setas na Figura 2. Adicionalmente, foram analisadas também as variações “percentual de acertos x TRI” e “(percentual de acertos + redação) x (TRI + redação)”.

A variável de análise será construída observando-se o percentual de alunos aprovados na metodologia de controle que também seriam aprovados pela metodologia analisada.<sup>12</sup> Quanto maior for esse percentual, mais as classificações se parecem e menos problemática é a alteração de uma metodologia para a outra. Por outro lado, quanto menor for esse percentual, ou quanto maior for a diferença percentual dos alunos aprovados na metodologia de análise, em relação à metodologia de controle, definido como “erro de aprovação”, mais questionável é a mudança.

Para ilustrar esse procedimento, a Tabela 1 apresenta um exemplo numérico hipotético dessa metodologia. A coluna “Classificação controle” representa o resultado de uma seleção fictícia que oferece dez vagas, com uma relação candidato-vaga igual a dois. De acordo com a tabela, os aprovados seriam aqueles classificados nas dez primeiras posições: os candidatos {A, E, C, D, B, J, T, I, L, F}.

Assumindo que a variável de interesse seja a inclusão da nota de redação na nota final, o cenário de controle seria “prova objetiva SEM redação” e o cenário de análise “prova objetiva COM redação”, estes denominados “Cenário I” e “Cenário II”, ilustrando dois possíveis resultados. Na classificação final obtida no “Cenário I”, nove dos dez estudantes aprovados no cenário de controle continuariam sendo aprovados, indicando que a alteração na variável de interesse não produziu efeitos qualitativamente tão graves.<sup>13</sup> No entanto, se a classificação final tivesse sido a do “Cenário II”, nota-se que o erro de aprovação, ou seja, o percentual de alunos aprovados no cenário de controle que não mais seriam aprovados após a alteração, seria de 40%. Portanto, se o total de vagas ofertadas fosse 100, 40 alunos perderiam a vaga.

Como os dados específicos de candidatos – a quais cursos e a quais IES se candidataram – não são divulgados, este estudo utiliza uma metodologia de simulação de Monte Carlo, que cria uma grande quantidade de cenários semelhantes ao da Tabela 1, permitindo que se construa uma distribuição para o erro de classificação, possibilitando uma análise estatística do mesmo. Os dados utilizados para a análise foram os relativos ao Enem 2008.<sup>14</sup> Uma análise mais recente não pôde ser feita, pois o Inep não divulgou os microdados das últimas três edições do Enem, apesar de reclamações de diversos setores da sociedade.<sup>15</sup>

Em 2008, segundo os microdados fornecidos pelo Inep, 4.018.070 candidatos se inscreveram no Enem. Destes, 2.920.588 compareceram aos dois dias de prova.

Subtraindo-se desse número as provas com respostas em branco e outros erros na base de dados,<sup>16</sup> chegou-se ao total de 2.911.699 observações. Para evitar problemas estatísticos decorrentes de possíveis diferenças entre os quatro cadernos do exame,<sup>17</sup> considerou-se apenas os candidatos que responderam ao caderno amarelo, resultando em uma amostra com 737.967 observações.

**Tabela 1**

Exemplificação da metodologia de comparação para um curso hipotético, com vinte candidatos e dez vagas. A área cinza representa a zona de aprovação. A “classificação controle” simboliza a classificação obtida de acordo com o método padrão, i.e., o que não sofreu variação na variável de interesse. As colunas “cenário I” e “cenário II” representam classificações hipotéticas geradas pela variação da variável de interesse.

Candidato	Classificação controle	Candidato	Cenário I	Candidato	Cenário II
A	1	C	3	E	2
E	2	I	8	N	15
C	3	A	1	A	1
D	4	D	4	B	5
B	5	M	11	H	20
J	6	T	7	I	8
T	7	J	6	T	7
I	8	B	5	P	12
L	9	L	9	L	9
F	10	F	10	M	11
M	11	E	2	C	3
P	12	Q	13	F	10
Q	13	G	14	G	14
G	14	H	20	R	17
N	15	S	16	J	6
S	16	R	17	S	16
R	17	P	12	Q	13
K	18	O	19	D	4
O	19	K	18	O	19
H	20	N	15	K	18
Erro		10%		40%	

Essa amostra foi utilizada na estimativa dos parâmetros dos itens e no cálculo da habilidade, de acordo com a mesma metodologia TRI utilizada pelo Inep.<sup>18</sup> O modelo utilizado é o logístico de três parâmetros de Birnbaum (1968), estimado

por máxima verossimilhança marginal, assumindo-se a distribuição normal padrão para o traço latente, seguindo a metodologia de Bock e Aitkin (1981). Por sua vez, a medida de habilidade é estimada pelo método bayesiano EAP – “*expected a priori*” – e transformada para uma medida com média igual a 500 e desvio-padrão igual a 100. Os procedimentos de estimação foram realizados no programa Bilog-MG (ZIMOWSKI et al., 1996). Os parâmetros estimados dos itens encontram-se no Anexo I.

O próximo passo na metodologia de comparação é a obtenção das classificações em um curso hipotético. O cálculo das médias finais a serem utilizadas na seleção é feito de três formas: média aritmética simples, média “medicina” e média “engenharia”. A diferença fundamental entre elas são os pesos atribuídos às notas da prova objetiva e da redação. Caso os dados do “novo Enem” estivessem disponíveis, a média aritmética seria calculada dessa forma:

$$\text{Média Aritmética} = \frac{\text{NAT} + \text{HUM} + \text{LIN} + \text{MAT} + \text{REDAÇÃO}}{5}$$

No entanto, como o Enem 2008 é composto por uma única prova, contendo questões de todas as disciplinas, em vez de quatro provas, uma para cada grande área do conhecimento, a forma de cálculo se altera para:

$$\text{Média Aritmética} = \frac{\text{OBJETIVA} \times 4 + \text{REDAÇÃO}}{5}$$

Essa aproximação é razoável, uma vez que existe uma compensação em termos de desempenho em uma cada uma das grandes áreas.<sup>19</sup> Analogamente, as médias “medicina” e “engenharia” seriam assim calculadas, caso os dados do “novo Enem” estivessem disponíveis:

$$\text{Média Medicina} = \frac{\text{NAT} \times 2 + \text{HUM} \times 1 + \text{LIN} \times 2 + \text{MAT} \times 1 + \text{REDAÇÃO} \times 3}{9}$$

$$\text{Média Engenharia} = \frac{\text{NAT} \times 4 + \text{HUM} \times 1 + \text{LIN} \times 1 + \text{MAT} \times 4 + \text{REDAÇÃO} \times 3}{13}$$

Esses pesos são idênticos aos pesos atribuídos às respectivas disciplinas no processo de seleção da Universidade Federal do Rio de Janeiro no concurso para o ingresso em 2012, e refletem escolhas alternativas para o peso de cada nota no cálculo da média final. Adaptando os cálculos anteriores para a prova do Enem 2008, obtém-se:

$$\text{Média Medicina} = \frac{\text{OBJETIVA} \times 6 + \text{REDAÇÃO} \times 3}{9}$$

$$\text{Média Engenharia} = \frac{\text{OBJETIVA} \times 10 + \text{REDAÇÃO} \times 3}{13}$$

O último passo da metodologia é a obtenção de conjuntos hipotéticos de candidatos por meio de uma rotina de Simulação de Monte Carlo. Essencialmente, o procedimento de simulação gera diversas tabelas semelhantes à Tabela 1, possibilitando que o número de erros de aprovação seja aferido em cada uma delas. Repetindo-se o processo milhares de vezes,<sup>20</sup> simula-se o que poderia ter acontecido realmente, caso fosse possível repetir o experimento “Enem” múltiplas ocasiões.<sup>21</sup>

Considerando a relação candidato-vaga desse curso hipotético igual a 50 – em alguns cursos essa relação é até maior, como em Medicina – e considerando uma oferta de 100 vagas, chegou-se ao tamanho da amostra de simulação: 5.000 observações. Esse processo é repetido 10.000 vezes, gerando uma distribuição para os erros de aprovação, definidos como a diferença percentual entre o número de alunos aprovados na metodologia de controle que continuariam sendo aprovados pela metodologia teste. Esse cálculo foi feito para os três tipos de média definidos anteriormente.

Como os alunos que geralmente pleiteiam os cursos mais concorridos estão entre os mais bem preparados, toda a análise foi repetida, restringindo-se o universo de amostragem aos candidatos que obtiveram nota TRI superior a um desvio-padrão em relação à média. Isto equivale a utilizar como base de simulação apenas aqueles com “habilidade” TRI acima de 1.0, sob a premissa de que o traço latente é distribuído conforme a distribuição normal padrão. A rigor, a distribuição *a posteriori* do traço latente não é, necessariamente, igual à distribuição *a priori*. No entanto, assume-se que os resultados qualitativos não se alterem tanto, uma vez que as distribuições tendem a ser semelhantes, como pode ser observado no Anexo II. Além disso, como a Curva de Informação do Teste alcança o máximo quando a habilidade é igual a 1.88, de acordo com a curva do Anexo IV, o corte proposto aumenta a precisão da análise, pois limita a amostra à região da curva com maior poder estatístico.

Concluindo, poder-se-ia argumentar que a validade dos resultados obtidos neste trabalho é questionável devido à limitada verossimilhança entre os dados das edições do “antigo” e “novo” Enem. Contudo, o que se está analisando nesse estudo é a validade da metodologia empregada pelo Inep/Sisu de uma forma geral, aplicada aos dados disponíveis. Antecipa-se que, diante do considerável tamanho da

amostra e da esperada ergodicidade das habilidades dos vestibulandos, os resultados da aplicação dessa metodologia sejam satisfatoriamente estáveis em amostras distintas. Além disso, a presidência do Inep afirmou que a utilização dos microdados de edições mais antigas é perfeitamente aceitável, pois “podem não ser os mais atuais, mas não são tão diferentes de um ano para o outro”.<sup>22</sup>

## Análise dos resultados

A Tabela 2 apresenta um sumário das estatísticas descritivas da amostra utilizada na simulação. Percebe-se que a distribuição da nota TRI é inteiramente concentrada no intervalo [258.18, 842.17], ao passo que a da nota da prova objetiva abrange um intervalo mais amplo: [15.90, 1000.00].

**Tabela 2**

Estatísticas descritivas da amostra do caderno amarelo do Enem 2008 após exclusão dos candidatos que obtiveram nota zero na parte objetiva ou na redação. A variável média “aritmética” corresponde ao cálculo ( $4 \times$  Prova Objetiva + Redação)/5; a variável média “medicina” corresponde ao cálculo ( $6 \times$  Prova Objetiva + 3 x Redação)/9 e a variável média “engenharia” corresponde ao cálculo ( $10 \times$  Prova Objetiva + 3 x Redação)/13, onde “Prova Objetiva” corresponde ao percentual de acertos multiplicado por 1000. A “nota TRI” corresponde à nota calculada pela TRI e as médias “aritmética TRI”, “medicina TRI” e “engenharia TRI” foram calculadas conforme as fórmulas anteriores, substituindo-se a “Prova Objetiva” pela nota TRI. Q1 equivale ao primeiro quartil da distribuição e Q3 ao terceiro quartil.

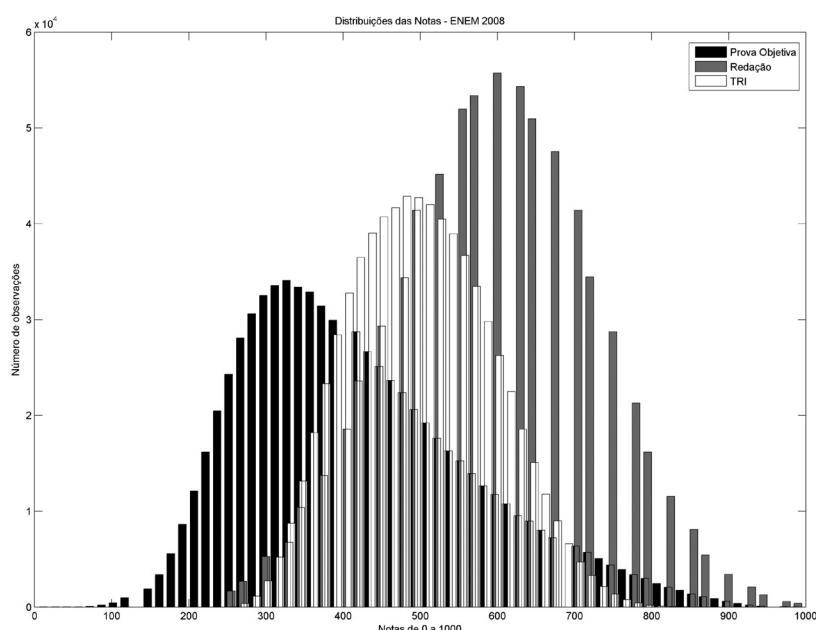
Variáveis	N	Média	Desv. Pad.	Min	Q1	Mediana	Q3	Max
Prova objetiva	721651	419.53	149.83	15.90	301.60	396.80	507.90	1000.00
Redação	721651	596.22	128.71	250.00	500.00	600.00	675.00	1000.00
Média aritmética	721651	454.87	131.77	62.72	356.64	434.80	536.32	974.64
Média medicina	721651	478.43	122.71	93.93	388.87	462.93	555.27	978.87
Média engenharia	721651	460.31	129.44	69.92	364.69	442.38	540.69	975.62
Nota TRI <sup>23</sup>	721651	500.36	92.00	258.18	431.39	496.55	564.85	842.17
Média aritmética TRI	721651	519.53	86.94	256.54	455.63	516.45	580.16	851.79
Média medicina TRI	721651	532.31	87.49	255.45	469.67	529.96	592.62	876.49
Média engenharia TRI	721651	522.48	86.78	256.29	459.03	519.52	582.84	857.49

Nota-se, também, que tanto a média, quanto a mediana da distribuição das notas de redação são significativamente superiores às do percentual de acertos e da

nota TRI. Como os candidatos que obtiveram nota zero na redação foram excluídos da amostra – zero na redação é fator eliminatório no Sisu –, fica evidente que a distribuição das notas de redação é truncada em 250, ou seja, não há *notas de redação no intervalo (0, 250)*, o que causa uma descontinuidade relevante na distribuição, de acordo com a Figura 3.

**Figura 3**

Histogramas das distribuições dos percentuais de acerto (“Prova Objetiva”), das notas da redação e das notas calculadas pela metodologia TRI.



A distribuição da nota TRI também é truncada no seu limite inferior, como a da redação. Este efeito introduz uma distorção adicional em qualquer comparação e combinação entre as diferentes notas. Com efeito, tomando-se a nota 800 como medida de comparação, exatos 49.027 candidatos obtiveram nota na redação igual ou superior a 800. Ao mesmo tempo, exatos 10.933 obtiveram nota na prova objetiva nesse mesmo intervalo – um número 4.48 vezes menor – e apenas 155 (cento e cinquenta e cinco) conseguiram obter uma nota TRI igual ou superior a 800, um número 316.3 menor.

Inicialmente averiguou-se se há diferenças marcantes no erros de aprovação entre a utilização do percentual de acertos e a nota TRI, sem incluir a redação na nota final. Os resultados encontram-se na Tabela 3.<sup>24</sup>

**Tabela 3**

Percentuais de erros de aprovação causados pela variação no instrumento de medição da nota objetiva. A primeira coluna corresponde aos resultados utilizando-se toda a amostra, e a segunda aos resultados obtidos quando a amostra foi restrita aos alunos com notas TRI acima de 1 desvio-padrão.

	Toda a amostra	Apenas as notas acima de 1 desvio-padrão
Média	11,6	20,8
Desvio-padrão	2,4	3,0
Mínimo	4,0	11,0
Máximo	21,0	35,0

Percebe-se que, em média, cerca de 12% dos alunos aprovados com base no percentual de erros não seria aprovados com base na nota TRI. Esse percentual se eleva para 21%, aproximadamente, quando se considera apenas os alunos de maior desempenho. Esse resultado pode ser explicado pela Figura 1, que mostra que a dispersão das notas TRI ao redor da reta de regressão diminui marcadamente para níveis elevados de desempenho. Assim, pequenas diferenças nas notas acarretam mudanças relevantes na classificação, o que gera um erro de aprovação maior do que na situação em que toda a amostra é utilizada e a classificação é mais robusta. As duas figuras a seguir ilustram os histogramas relativos às distribuições dos erros de aprovação nesses dois cenários.

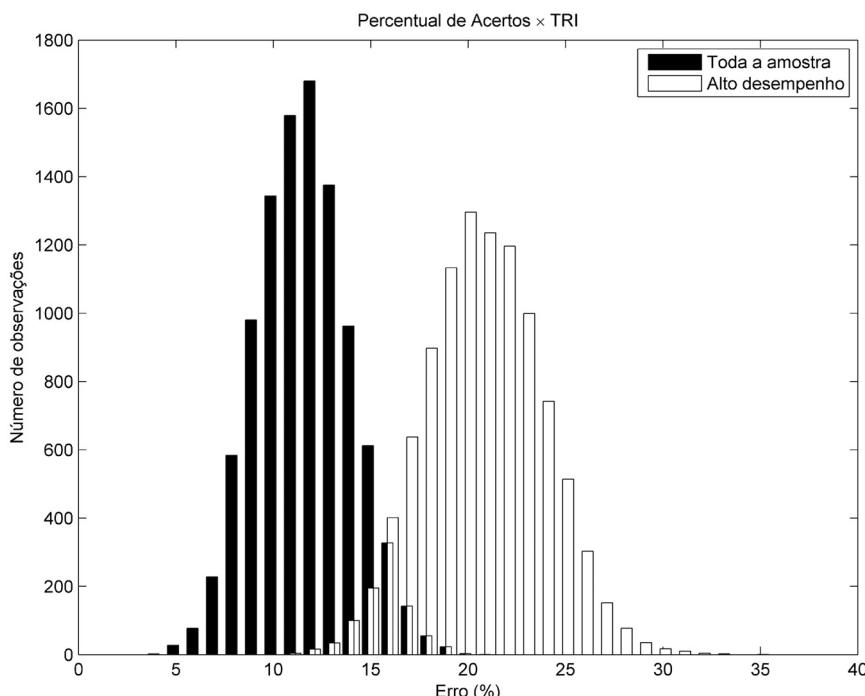
Observa-se, comparando-se as Figuras 3 e 4, que houve uma translação de toda a distribuição de erros para a direita, devido ao aumento de sensibilidade da classificação em relação à amostra considerada.

Na Tabela 4, encontram-se os resultados relativos ao impacto da adoção da redação no cálculo da nota final, quando a nota objetiva é medida pelo percentual de acertos.

Evidencia-se que o erro médio aumenta na seguinte ordem: média aritmética, média “engenharia” e média “medicina”. Tal fato é consistente com os diferentes pesos da redação no cálculo da nota final: 20,0%, 23,1% e 33,3%, para as médias aritmética, “medicina” e “engenharia”, respectivamente. De acordo com a tabela, cerca de 28% dos candidatos para o curso de Medicina perderiam vaga com a inclusão da redação, podendo atingir até 43% dos candidatos. A perda média no caso da média aritmética é de cerca de 16%, ao passo que a perda para os candidatos de Engenharia atinge aproximadamente um em cada cinco candidatos. A comparação entre os histogramas dos erros de aprovação de acordo com cada tipo de média encontra-se na Figura 5.

**Figura 4**

Histogramas da distribuição do erro de aprovação, originados pela mudança na utilização do percentual de acerto para a nota TRI, sem considerar a redação no cálculo da média final. Todos os candidatos foram incluídos na simulação.

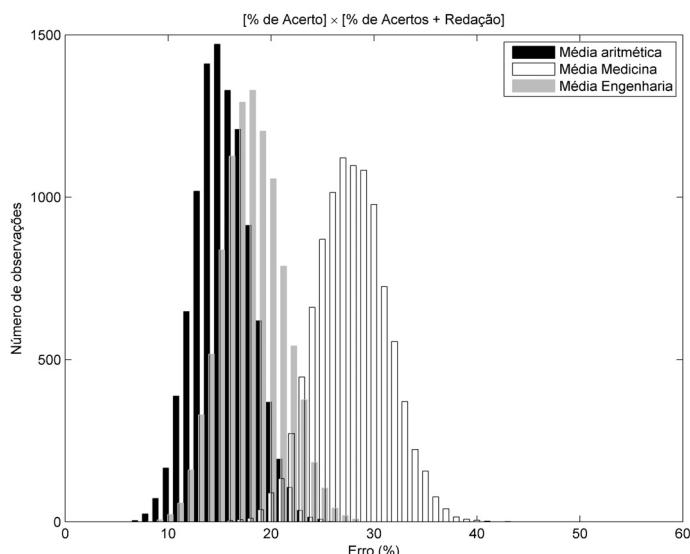
**Tabela 4**

Percentuais de erros de aprovação causados pela variação na forma de cálculo da média final, pela adoção da redação, quando o instrumento de medição da nota da prova objetiva é o percentual de acertos. As variáveis “aritmética”, “medicina” e “engenharia” correspondem às definidas anteriormente. O erro de aprovação é definido como a diferença percentual dos alunos aprovados na metodologia de análise, em relação à metodologia de controle, nesse caso, a utilização do percentual de acertos sem a redação. A amostra completa foi considerada.

	Aritmética	Medicina	Engenharia
Média	15.5	27.9	18.1
Desvio-padrão	2.7	3.5	3.0
Mínimo	7.0	16.0	8.0
Máximo	26.0	43.0	29.0

**Figura 5**

Histogramas das distribuições dos percentuais de erro, com base na média aritmética, média “medicina” e média “engenharia” anteriormente definidas, quando a nota da prova objetiva é dada pelo percentual de acertos. Todos os candidatos foram considerados na amostra utilizada na simulação.



O descolamento da distribuição dos erros de aprovação para as médias do curso de Medicina é evidente. Considerando-se apenas os candidatos de alto desempenho, para o mesmo cenário “percentual de acertos” × “percentual de acertos + redação”, obtém-se os dados mostrados na Tabela 5.

**Tabela 5**

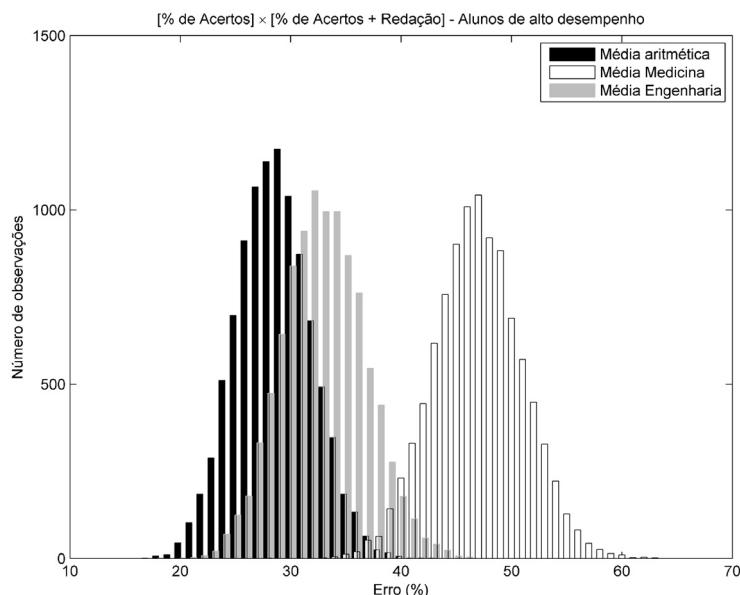
Percentuais de erros de aprovação causados pela variação na forma de cálculo da média final, pela adoção da redação, quando o instrumento de medição da nota da prova objetiva é o percentual de acertos. As variáveis “aritmética”, “medicina” e “engenharia” correspondem às definidas anteriormente. O erro de aprovação é definido como a diferença percentual dos alunos aprovados na metodologia de análise, em relação à metodologia de controle, nesse caso, a utilização do percentual de acertos *sem* a redação. Apenas os candidatos com notas TRI acima ou superior a 1 desvio-padrão foram considerados.

	Aritmética	Medicina	Engenharia
Média	28.6	47.0	32.9
Desvio-padrão	3.4	4.0	3.8
Mínimo	17.0	28.0	18.0
Máximo	43.0	63.0	46.0

Como anteriormente, a distorção gerada pela inclusão da nota da redação no cálculo da média final aumenta consideravelmente quando a amostra de simulação é restringida aos candidatos de alto desempenho. Em Engenharia, um em cada três aprovados perderia a vaga, com limite máximo de distorção igual a 46%. Os problemas são ainda mais graves em Medicina, em que aproximadamente metade dos estudantes originalmente aprovados perderia a vaga, em média. No pior cenário, 63% dos candidatos sofreriam essa perda. A Figura 6 ilustra o histograma referente a essa situação.

**Figura 6**

Histogramas das distribuições dos percentuais de erro, com base na média aritmética, média “medicina” e média “engenharia” anteriormente definidas, quando a nota da prova objetiva é dada pelo percentual de acertos. Apenas os candidatos de alto desempenho, ou seja, com nota TRI acima de 1 desvio-padrão, foram considerados.



A análise a seguir comprova a hipótese de que a combinação da nota TRI com a nota da redação amplifica significativamente os efeitos desse procedimento no erro de aprovação. De acordo com os resultados apresentados na Tabela 6, a distorção é de 32% para Engenharia, ou seja, cerca de um estudante em cada três aprovados; quando apenas a nota TRI é empregada no cálculo da nota final, perderia a vaga. O impacto da combinação da nota da redação com a nota TRI é tão significativo que metade dos aprovados em Medicina perderia a vaga, em média, e até 60% seriam prejudicados por esse equívoco.

**Tabela 6**

Percentuais de erros de aprovação causados pela variação na forma de cálculo da média final, pela adoção da redação, quando o instrumento de medição da nota da prova objetiva é a TRI. As variáveis “aritmética”, “medicina” e “engenharia” correspondem às definidas anteriormente. O erro de aprovação é definido como a diferença percentual dos alunos aprovados na metodologia de análise, em relação à metodologia de controle, nesse caso, a utilização da TRI sem a redação. A amostra completa foi considerada.

	Aritmética	Medicina	Engenharia
Média	27.9	44.7	32.0
Desvio-padrão	3.5	3.5	3.7
Mínimo	16	29.0	18.0
Máximo	41.0	60.0	47.0

Além disso, comparando-se os resultados contidos nas Tabelas 4 e 6, nota-se que a diferença entre o erro médio em Medicina, por exemplo, no cenário em que a prova objetiva é medida pelo percentual de acertos, e o erro no cenário em que a prova objetiva é medida pelo Enem é de 17%. Esse aumento pode ser atribuído única e diretamente à incompatibilidade entre as diferentes escalas de medida desses dois instrumentos.

Finalmente, o erro de aprovação, quando apenas os candidatos de desempenho superior são incluídos na amostra, chega a surpreendentes 76% em Medicina e 60% em Engenharia, com valores médios iguais a 60% e 45%, respectivamente, como o disposto na Tabela 7. Como essa premissa é mais realista, tendo em vista que esses cursos são, de fato, muito concorridos e é razoável assumir que atrairão uma quantidade maior de alunos mais bem preparados, esses resultados são muito preocupantes.

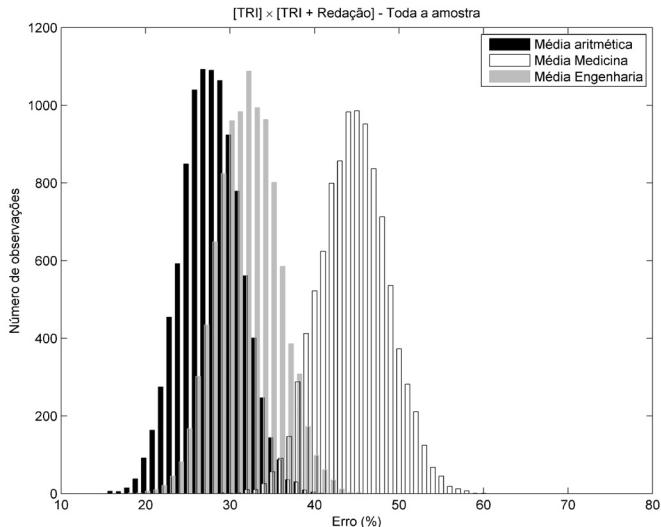
**Tabela 7**

Percentuais de erros de aprovação causados pela variação na forma de cálculo da média final, pela adoção da redação, quando o instrumento de medição da nota da prova objetiva é a TRI. As variáveis “aritmética”, “medicina” e “engenharia” correspondem às definidas anteriormente. O erro de aprovação é definido como a diferença percentual dos alunos aprovados na metodologia de análise, em relação à metodologia de controle, nesse caso, a utilização da TRI sem a redação. Apenas os candidatos com notas TRI acima ou superior a 1 desvio-padrão foram considerados.

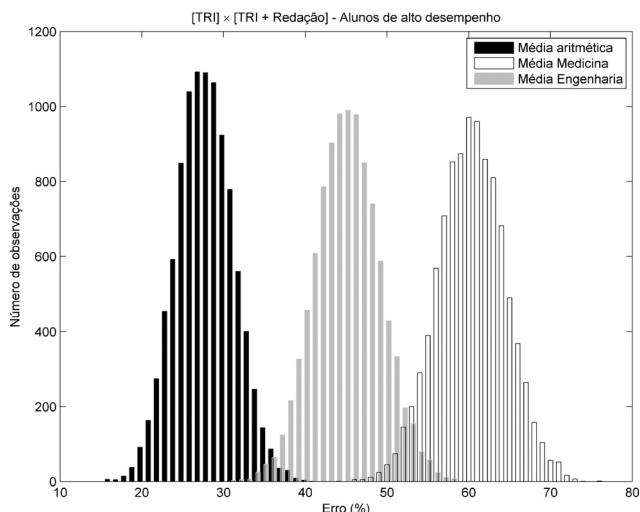
	Aritmética	Medicina	Engenharia
Média	39.5	60.3	45.0
Desvio-padrão	3.8	4.1	4.0
Mínimo	26.0	44.0	31.0
Máximo	54.0	76.0	60.0

**Figura 7**

Histogramas das distribuições dos percentuais de erro, com base na média aritmética, média “medicina” e média “engenharia” anteriormente definidas, consequência da inclusão da nota da redação na média final, quando a nota da prova objetiva é dada pela TRI. Todos os candidatos foram considerados na amostra utilizada na simulação.

**Figura 8**

Histogramas das distribuições dos percentuais de erro, com base na média aritmética, média “medicina” e média “engenharia” anteriormente definidas, consequência da inclusão da nota da redação na média final, quando a nota da prova objetiva é dada pela TRI. Apenas os candidatos de alto desempenho, ou seja, com nota TRI acima de 1 desvio-padrão, foram considerados.



Qualitativamente, os histogramas da Figura 7 e 8 são semelhantes aos das Figuras 5 e 6: a distribuição das médias de Medicina se isola à direita das outras duas, comprovando que as consequências do uso da redação na média final agravam-se nos cursos em que o peso dado à prova da redação é relativamente grande.

## Conclusão

A adoção em massa do Enem como principal, e em muitas vezes única, forma de acesso aos cursos de graduação de IES públicas faz necessária uma análise crítica de toda a metodologia empregada no processo de seleção. A mera sugestão de que o Enem possa se tornar obrigatório<sup>25</sup> para todos os concluintes do ensino médio amplifica essa necessidade, assim como a provável substituição da Prova Brasil pelo Enem no cálculo do Ideb, ampliando o alcance do exame à esfera de formulação de políticas públicas.

Este trabalho verifica empiricamente a invalidade da combinação da nota estimada pela TRI com a nota da redação. Conceitualmente, se um dos motivos principais para a adoção da TRI no Enem, em 2009, foi exatamente a vantagem que esse modelo apresenta no sentido de permitir que diferentes avaliações possam ser comparadas, por estarem calibradas em uma métrica única, é altamente questionável a combinação dessas duas notas de forma livre. Além disso, aberrações na distribuição de notas de redação, documentadas na Figura 3, contribuem para o argumento de que *as notas de redação não deveriam ser combinadas de forma alguma com a prova objetiva*, independentemente do instrumento de medida dessa prova. Considerando, ainda, a quase impossibilidade de garantir a homogeneidade da correção das redações nesse processo, questiona-se a validade da inclusão dessa nota a qualquer tipo de avaliação, individual ou coletiva.

Naturalmente, essas distorções serão também transmitidas a todo tipo de avaliação escolar que possa ser realizada com base nessas notas. Consequentemente, quaisquer esforços futuros de responsabilização derivados dessas avaliações estariam, da mesma forma, comprometidos.

Os resultados sugerem, inequivocamente, que a inclusão da nota da redação distorce de forma expressiva a classificação final em um processo de seleção hipotético, com parâmetros pautados, porém, em premissas realísticas. O impacto é estatisticamente significativo e pode comprometer de 27% a mais de 60% das vagas ofertadas.

Espera-se que este trabalho tenha contribuído para o debate em torno do Enem, que se torna cada vez mais relevante e urgente, em virtude da importância que o exame vem ganhando no cenário atual da educação brasileira e na definição

da competitividade nacional futura. O trabalho destaca também os riscos de se empregar medidas quantitativas frias como possíveis indicadores da “qualidade” da educação, uma vez que esses representam uma agregação dos dados individuais. Claramente, é necessário que medidas de avaliação mais robustas sejam desenvolvidas. Estas devem considerar definições mais amplas do conceito de “qualidade” educacional, assim como imbutir na sua construção aspectos relevantes da realidade educacional brasileira, como, entre outros, o nível socioeconômico da população e a qualidade dos insumos educacionais investidos no sistema de educação brasileiro.

## Notas

1. Artigo 1º, Portaria MEC n. 438, de 28 de maio de 1998.
2. Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_docman&task=doc\\_download&gid=768&Itemid=>](http://portal.mec.gov.br/index.php?option=com_docman&task=doc_download&gid=768&Itemid=>)>. Acesso em: 1º jan. 2012.
3. Por exemplo, a UFRJ e a Unirio, no estado do Rio de Janeiro.
4. Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_content&view=article&id=18031](http://portal.mec.gov.br/index.php?option=com_content&view=article&id=18031)>. Acesso em: 22 de agosto de 2012.
5. Ver, por exemplo: <<http://www1.folha.uol.com.br/saber/997015-escola-do-ceara-diz-que-usou-questoes-de-pre-teste-do-enem-leia-nota.shtml>>.
6. Uma argumentação mais social pela utilização da TRI é encontrada em Fletcher (1994).
7. Não foram abordadas neste trabalho discussões sobre a relação entre variáveis observáveis e traços latentes, assim como sobre métodos de estimação dos parâmetros dos itens e do traço latente, nem sobre a validade das premissas matemáticas e estatísticas da TRI e os testes de ajuste do modelo aos dados observados.
8. Mais detalhes sobre o método de obtenção das notas TRI encontram-se na seção de metodologia. A metodologia de estimação empregada é a mesma do Inep.
9. Como o relatado em: <<http://www1.folha.uol.com.br/saber/1026606-alunos-questionam-pontuacao-do-enem-e-fazem-ate-site-de-queixas.shtml>>, entre outras fontes. Acesso em 1º jan. 2012.
10. Candidatos que responderam ao caderno amarelo do Enem 2008.
11. NAT = Ciências da Natureza, HUM = Ciências Humanas, LIN = Linguagens e MAT = Matemática.
12. Adicionalmente, poderiam ter sido utilizadas algumas estatísticas mais sofisticadas, que avaliam a distância entre dois vetores de classificação, entre elas: “Spearman’s rho”, “Spearman’s footrule” e “Kendall’s tau”.
13. Assumindo-se, nesse caso, que um nível de erro igual a 10% é aceitável. Em cursos que oferecem um número maior de vagas, por exemplo, 200 vagas, essa taxa de erro indicaria que vinte alunos perderiam a vaga que lhes seriam de direito. Cabe à sociedade decidir se esse nível de erro é aceitável, ou não.
14. O Inep disponibiliza os microdados de diversos exames, incluindo o Enem, em: <<http://portal.inep.gov.br/basicas-levantamentos-acessar>>. Acesso em: 2 jan. 2011.
15. Jornal *Valor Econômico*, Caderno A – Brasil, p. A5, 30 jun. 2011.
16. Alguns registros no banco de dados estavam incompletos e não puderam ser incluídos.
17. Apesar dos quatro cadernos conterem exatamente as mesmas questões, apresentadas em ordem distinta, é possível que haja diferenças entre eles (KLEIN; FONTANIVE, 1995).

18. Idealmente, para simular ao máximo a estrutura do “novo Enem”, a prova de 2008 deveria ser subdividida em quatro provas únicas, com cerca de 14 questões cada. Em seguida, a TRI deveria ser aplicada em cada uma dessas provas. No entanto, apesar da literatura não ser precisa quanto ao número mínimo de itens necessários para garantir eficiência na estimativa dos parâmetros, há indícios de que esse limite esteja entre 20 e 30 (ANDRADE; TAVARES; VALE, 2000). Assim, essa alternativa não se mostra viável.
19. Essa compensação nas habilidades é base para modelos mais sofisticados de TRI como, por exemplo, os modelos multidimensionais (RECKASE, 2009).
20. O gerador de números pseudoaleatórios utilizado foi o “*Mersenne Twister*”. Detalhes em: <<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>>.
21. Um tratamento filosófico mais informal, embora igualmente relevante, pode ser encontrado em Taleb (2001).
22. Disponível em: <<http://www.estadao.com.br/noticias/vidae,presidente-do-inep-critica-cobertura-da-imprensa-do-enem-por-escolas,773056,0.htm>>. Acesso em: 2 jan. 2011.
23. Essas estatísticas se referem à distribuição *a posteriori* das habilidades.
24. As estatísticas-t são calculadas da forma convencional, assumindo a hipótese nula de que a média populacional seja igual a zero:  $t = \frac{\bar{x}}{s/\sqrt{n}}$ , onde  $\bar{x}$  é a média amostral,  $s$  é o desvio-padrão amostral e  $n$  é o tamanho da amostra de simulação.
25. Disponível em: <<http://blogs.estadao.com.br/jt-cidades/haddad-defende-enem-obrigatorio/>>. Acesso em: 1º jan. 2012.

## Referências

- ANDRADE, D.F.; TAVARES, H.R.; VALLE, R.C. *Teoria de Resposta ao Item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística, 2000.
- BAKER, F.B. *The basics of item response theory*. Washington, DC: Eric, 2001.
- BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. In: LORD, F.M.; NOVICK, M.R. (Ed.). *Statistical theories of mental test scores*. Reading: Addison-Wesley, 1968.
- BOCK, R.D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: an application of an EM algorithm. *Psychometrika*, Williamsburg, v. 46, n. 4, p. 443-459, 1981.
- CASTRO, M.H.G.; TIEZZI, S. A reforma do ensino médio e a implantação do Enem no Brasil. In: BROCK, C.; SCHWARTZMAN, S. *Os desafios da educação no Brasil*. Rio de Janeiro: Nova Fronteira, 2005. p. 119-154.
- COURVILLE, T.G. *An empirical comparison of Item Response Theory and Classical Test Theory Item/Person Statistics*. 2004. Tese (Doutorado) – Texas A&M University, Kingsville.
- EMBRETSON, S.E.; REISE, S.P. *Item Response Theory for psychologists*. New Jersey: LEA, 2000.

FAN, X. Item Response Theory and Classical Test Theory: an empirical comparison of their item/person statistics. *Education and Psychological Measurement*, v. 58, n. 3, p. 357-381, 1998.

FERRÃO, M.E. Avaliação educacional e modelos de valor acrescentado: tópicos de reflexão. *Educação & Sociedade*, Campinas, v. 33, n. 119, p. 455-469, 2012.

FLETCHER, P. R., A Teoria da Resposta ao Item: medidas invariantes do desempenho escolar. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 1, n. 2, p. 21-27, 1994.

HAMBLETON, R.K.; VAN DER LINDEN, W. Advances in Item Response Theory and applications: an introduction. *Applied Educational Measurement*, v. 6, n. 4, p. 373-378, 1982.

HAMBLETON, R.K.; SWAMINATHAN, H. *Item Response Theory: principles and applications*. Boston: Kluwer, 1985.

HAMBLETON, R.K.; SWAMINATHAN, H.; ROGERS, H.J. *Fundamentals of Item Response Theory*. Newbury Park: Sage, 1991.

KARINO, C.A.; ANDRADE, D.F. *Entenda a Teoria de Respostas ao Item (TRI)*. Nota técnica. Brasília, DF: Inep, 2011.

KLEIN, R.; FONTANIVE, N.S. Avaliação em larga escala: uma proposta inovadora. *Em Aberto*, Brasília, DF, v. 15, n. 66, p. 29-34, 1995.

KOLEN, M.J.; BRENNAN R.L. *Test equating, scaling, and linking: method and practice*. New York: Springer, 2004.

LORD, F.M. *Application of Item Response Theory to practical testing problems*. Hillsdale: Erlbaum, 1980.

LORD, F.M.; NOVICK, M.R. *Statistical theories of mental test scores*. Reading: Addison-Wesley, 1968.

MACDONALD, P.; PAUNONEN, S.V. A Monte Carlo comparison of item and person statistics based on Item Response Theory vs. Classical Test Theory. *Educational and Psychological Measurement*, Durham, v. 62, n. 6, p. 921-943, 2002.

MOREIRA JUNIOR, F. Aplicações da Teoria da Resposta ao Item (TRI) no Brasil. *Revista Brasileira de Biometria*, São Paulo, v. 28, n. 4, p. 137-170, 2010.

RECKASE, M.D. *Multidimensional Item Response Theory*. New York: Springer, 2009.

TALEB, N. *Fooled by randomness: the hidden role of chance in the markets and in life*. New York: Springer, 2001.

VAN DER LINDEN, W.; HAMBLETON, R.K. *Handbook of modern Item Response Theory*. New York: Springer, 1997.

VON DAVIER, A.A. *Statistical models for test equating, scaling, and linking*. New York: Springer, 2010.

ZIMOWSKI, M.F. et al. *Bilog-MG: multiple-group IRT analysis and test maintenance for binary items*. Chicago: Scientific Software International, 1996.

---

*Recebido em 12 de junho de 2012.*

*Aprovado em 18 de dezembro de 2012.*

## Anexos

### Anexo I

Parâmetros TRI e outras estatísticas relativas aos 63 itens do Enem 2008. "% Acertos" corresponde ao percentual da amostra que acertou a questão; "dificuldade", "discriminação" e "acerto casual" correspondem aos parâmetros homônimos da TRI; "corr. biserial" corresponde ao coeficiente de correlação biserial da Teoria Clássica.

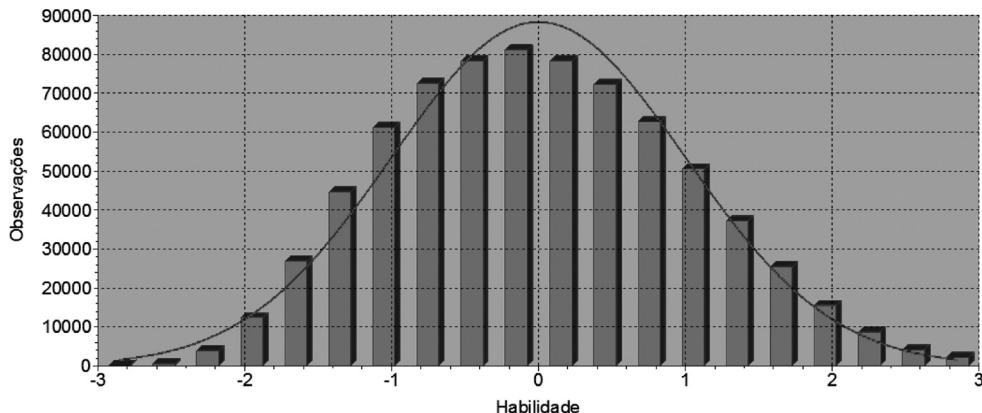
Item	% Acertos	Dificuldade	Discriminação	Acerto casual	Corr. Biserial
1	49,7	0,039	0,294	0,004	0,262
2	22,6	2,118	0,986	0,166	0,256
3	31,8	1,350	1,233	0,196	0,410
4	74,9	-0,419	0,750	0,372	0,410
5	12,9	3,324	0,563	0,076	0,211
6	53,4	0,532	0,661	0,244	0,380
7	30,1	1,253	0,688	0,083	0,438
8	33,9	1,353	0,807	0,174	0,387
9	58,1	-0,459	0,447	0,009	0,370
10	65,6	-0,202	1,145	0,216	0,564
11	50,3	1,804	0,768	0,423	0,187
12	30,8	1,994	0,266	0,015	0,231
13	31,9	2,151	0,603	0,213	0,228
14	48,2	0,429	0,852	0,149	0,502
15	40,0	0,636	1,175	0,122	0,595
16	50,5	1,122	1,469	0,396	0,339
17	60,9	-0,183	0,570	0,155	0,399
18	38,1	1,457	1,213	0,285	0,311
19	54,8	0,233	0,950	0,196	0,518
20	61,9	-0,634	0,524	0,002	0,413
21	27,5	2,064	2,256	0,253	0,130
22	35,7	1,645	0,762	0,233	0,293
23	49,0	0,371	0,887	0,144	0,523
24	38,8	1,796	1,581	0,343	0,194
25	35,0	1,254	0,919	0,188	0,406
26	35,1	1,102	1,348	0,197	0,473
27	43,8	0,433	0,845	0,077	0,542
28	44,2	0,391	0,401	0,003	0,346
29	24,8	1,727	1,144	0,164	0,338

30	39,3	1,026	0,960	0,200	0,439
31	28,9	1,951	1,292	0,239	0,212
32	47,9	0,577	1,026	0,208	0,501
33	34,7	1,301	0,825	0,179	0,391
34	47,7	1,016	0,645	0,265	0,330
35	42,2	0,693	1,005	0,158	0,525
36	49,0	0,891	0,759	0,277	0,363
37	53,4	-0,428	0,186	0,003	0,156
38	10,8	2,590	0,968	0,070	0,268
39	40,0	1,123	0,719	0,194	0,371
40	85,0	-2,036	0,588	0,008	0,391
41	36,8	0,733	0,841	0,072	0,537
42	61,8	0,322	0,007	0,237	-0,033
43	29,3	1,844	0,588	0,143	0,306
44	65,3	-0,156	0,907	0,242	0,495
45	38,6	1,071	0,837	0,185	0,422
46	9,7	2,508	1,201	0,069	0,252
47	67,9	-0,494	0,988	0,118	0,558
48	54,8	0,562	0,661	0,274	0,364
49	89,0	-1,445	0,854	0,375	0,440
50	37,3	1,396	1,721	0,290	0,334
51	12,4	2,397	1,493	0,102	0,198
52	39,9	0,979	0,931	0,194	0,441
53	28,7	1,398	0,937	0,139	0,430
54	59,8	-0,534	0,438	0,026	0,361
55	31,1	1,012	1,013	0,095	0,553
56	21,3	1,931	1,203	0,152	0,302
57	40,4	0,819	1,213	0,187	0,521
58	17,6	2,176	1,120	0,127	0,274
59	49,5	0,476	1,304	0,216	0,555
60	34,3	2,081	1,357	0,309	0,174
61	29,5	1,451	0,951	0,159	0,404
62	55,7	1,314	0,656	0,421	0,239
63	17,4	2,141	1,975	0,150	0,181

## Anexo II

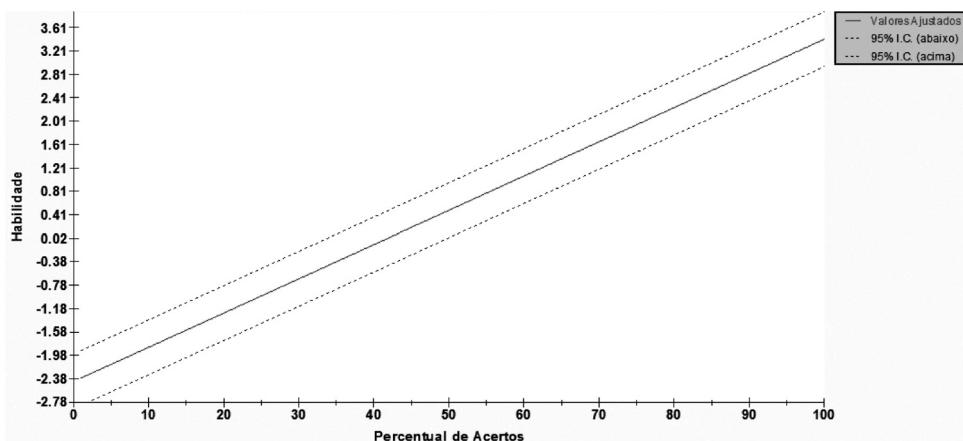
Histograma do traço latente da amostra, Enem 2008. O traço contínuo indica a distribuição normal padrão e as barras a distribuição *a posteriori* das habilidades.

**Distribuição das Habilidades - Enem 2008**



## Anexo III

Regressão linear da habilidade estimada pela TRI com o percentual de acertos da Teoria Clássica, na escala da distribuição normal padrão. Gráfico produzido automaticamente pelo software de cálculo da TRI, que equivale à Figura 1, que foi obtida diretamente dos dados.



#### Anexo IV

Curva de Informação do Teste – Enem 2008

