

QUE CONHECIMENTOS E HABILIDADES ESTÃO EM TESTE NO ENEM? ANÁLISE DE *PERFORMANCE* *INDUZIDA REVERSA*

WHICH KNOWLEDGE AND SKILLS ARE ASSESSED IN ENEM? ANALYSIS OF
REVERSE INDUCED PERFORMANCE

Nelio Bizzo, Ana Maria Santos-Gouw, Paulo Sérgio Garcia, Paulo Henrique Nico Monteiro
Faculty of Education and EDEVO Research Nucleus, São Paulo University – Brazil

Luiz Caldeira Brant Tolentino-Neto
Santa Maria Federal University – Brazil

ABSTRACT: Two instruments based on real questions of the Brazilian National Exam (“ENEM”) were presented to two randomized groups of high school students. One group received real questions (n=1,631), and the other received the same questions (n=1,400) with abridged context information, but with the same command and alternatives. Results were analyzed with statistical tools, showing higher performance in some questions with abridged information (*reverse induced performance*). Qualitative tools were used to explain this awkward phenomenon. Conclusions show that students’ performance may rely more heavily on perception of subtleties in questions’ alternatives, reading skills and time management ability than on previous knowledge and mental skills.

KEYWORDS: assessment of competences; concepts and context; performance assessment; reverse induced performance.

INTRODUÇÃO

O Ministério da Educação (MEC) organiza um exame nacional para os alunos no final do Ensino Médio (“ENEM”) desde o ano de 1998 no bojo das políticas de avaliação da qualidade da educação do Brasil (Vianna, 2003; Zákia e Oliveira, 2003). Participar do exame permite ao aluno candidatar-se a uma vaga em universidades públicas e privadas, além de certificar conclusão de estudos.

No ano de 2012 mais de cinco milhões de estudantes se matricularam no ENEM, concorrendo a 130.000 vagas no sistema unificado de seleção de universidades públicas (“SISU”). Após a realização do exame, cerca de um milhão de estudantes passaram a concorrer com sua nota no ENEM a pouco mais de 170.000 bolsas de estudo em instituições privadas (“PROUNI”), que podem chegar a cobrir 100% do valor das mensalidades, observadas algumas características socioeconômicas do candidato. O exame custa aos cofres públicos diretamente cerca de R\$ 400 milhões (US\$ 200 milhões), disputados por empresas privadas especializadas em um processo que dispensa licitação, além da renúncia fiscal requerida para o fornecimento de bolsas de estudo do PROUNI. No ano de 2013 se inscreveram quase

oito milhões de estudantes, embora os concluintes do ensino médio devam corresponder a menos de 25% desse total.

O ENEM é conhecido por evitar questões tradicionais, que dependem fortemente de conhecimento anterior dos alunos. Em seu lançamento, em 1998, foi apresentado como uma nova estratégia para avaliar diretamente as competências dos alunos, definidas por um documento oficial como “modalidades estruturais de inteligência” (Franco e Bonamino, 1999:29). O novo teste foi calorosamente recebido pela imprensa brasileira e amplamente promovido em impressos avulsos oficiais. Ele foi tomado não só como uma estratégia para uma reforma educativa baseada em avaliação, mas também como instrumento de reforma social, uma vez que supostamente iria ajudar os estudantes pobres a ingressar na educação superior e, além disso, também poderia direcioná-los para o mercado de trabalho. O MEC apresentou o teste como uma oportunidade para que os jovens planejassem seu futuro, a partir de uma “idéia clara do seu potencial pessoal e profissional”, uma vez que esse exame “permitiria avaliar seu potencial para planejar futuras escolhas” (Zakia e Oliveira, 2003: 884).

O exame está baseado em extensa lista de competências e habilidades, ao lado das referências conceituais para o conhecimento disciplinar (“objetos de conhecimento”) (MEC/INEP, 2012). Consta de 180 questões, mais uma redação obrigatória, aplicado em dois dias consecutivos, com controle rigoroso de tempo. Cada questão traz um contexto visual e escrito, seguido por um comando e cinco alternativas. Conhecimento prévio raramente é necessário, pelo menos na forma de definições conceituais.

OBJETIVOS E MÉTODOS

Esta pesquisa colocou em teste a relevância do contexto inicial das questões e buscou nexos causais para os resultados encontrados, apoiando-se na Linguística e na Análise de Discurso (Orlandi, 2012).

Sete questões claramente relacionadas à biologia do ENEM 2009 e 2010 foram selecionadas e apresentadas a dois grupos aleatórios de alunos do Ensino Médio. Um grupo ($n_0 = 233$) foi convidado a responder as perguntas com o contexto inicial original, em um questionário de seis páginas (denominado “grupo 0”); outro grupo similar de estudantes ($n_1 = 200$) foi convidado a responder às mesmas perguntas, com uma diferença, uma vez que o contexto inicial tinha sido removido, deixando, no entanto, o mesmo comando e alternativas, sob a forma de um questionário de três páginas (denominado “grupo 1”). Outras três perguntas foram incluídas nos dois conjuntos de questionários, exatamente com o mesmo comando breve e cinco alternativas, como um padrão para fins de comparação (denominadas “questões padrão”, tabela 2). O formulário do “grupo 1” teve ainda a inserção de uma questão adicional com a finalidade de tornar os dois formulários mais parecidos, do ponto de vista editorial e, ao mesmo tempo, garantir a identidade do grupo em teste no processamento dos resultados.

Cada assistente de pesquisa recebeu um dos dois conjuntos de questionários e foi responsável por apresentar aos estudantes de uma escola pública da cidade de São Paulo (SP). Quatorze escolas foram escolhidas aleatoriamente, no mesmo intervalo de tempo (meados de setembro de 2011). Os assistentes de pesquisa não estavam cientes das diferenças dos dois conjuntos de questionários. A carta-convide com o Termo de Consentimento Livre e Esclarecido (TCLE) fazia parte de cada questionário, e convidava os alunos a colaborarem em uma pesquisa sobre a avaliação, sob a condição de que eles não seriam identificados na folha de respostas, e as várias escolas participantes não seriam identificadas ou classificadas.

A análise estatística descritiva incluiu ensaios paramétricos, com medidas de tendência central, variabilidade, assimetria e curtose e busca de *outliers* (Bisquerra et al, 2004:39). Confirmada a homogeneidade das variâncias do campo amostral, e descartada dos cálculos uma escola *outlier* (escola 9), a aplicação do Teste Exato de Fisher com os resultados das três questões padrão (tabela 2) confirmou a similaridade dos dois grupos de alunos (Q_2 , $p = 0,787$; Q_4 , $p = 0,116$ e Q_6 , $p = 0,140$), com variações

atribuíveis apenas ao acaso. As mesmas análises realizadas com as questões do ENEM (Tabela 1) apontaram p-valor sem diferenças estatisticamente significativas entre os grupos em quatro questões (Q1, $p = 0,906$; Q5, $p = 0,077$; Q9, $p = 0,901$; Q10, $p = 0,152$), e diferenças estatisticamente significativas em três questões, em favor de questões SEM o contexto inicial do ENEM (Q03, $p = 0,006$; Q7, $p < 0,001$ e Q8, $p < 0,001$).

As questões com maior significância estatística ($p < 0,001$) na diferença de desempenho dos estudantes dos dois grupos foram analisadas do ponto de vista qualitativo, utilizando ferramentas linguísticas e a metodologia de análise de discurso (Orlandi, 2012).

RESULTADOS

O número total de perguntas focalizando o ENEM foi 3.031 (TABELA 1), outras 1.299 perguntas padrão foram processadas, a fim de testar a homogeneidade da amostra (Tabela 2), com um número total de 4.330 perguntas respondidas e processadas.

Tabela 1.

Resultados de 3.031 questões do ENEM com e sem contexto (T.E.F.= Teste Exato de Fisher)

Questões ENEM com contexto inicial ("Grupo 0")									
N	Escola	n_0	Q1	Q3	Q5	Q7	Q8	Q9	Q10
1	EEA0HD	15	1	12	13	4	2	3	3
2	EEB0NL	52	18	30	32	11	12	10	12
3	EEC0SB	32	15	22	20	6	16	10	21
4	EED0XS	13	7	9	12	9	6	7	8
5	EME0EA	32	12	8	17	7	4	17	7
6	EEF0PM	34	21	27	28	23	19	3	17
7	EEG0GC	13	7	10	11	11	3	9	11
8	EEH0BT	42	27	32	34	26	9	21	20
	Total	233	108	150	167	97	71	80	99
			46%	64%	72%	42%	30%	34%	42%
Questões ENEM sem contexto inicial ("Grupo 1 - Teste")									
		n_1	Q1	Q3	Q5	Q7	Q8	Q9	Q10
9	EEI1FB	73	44	14	3	49	57	12	7
10	EEK1BM	43	14	35	36	25	29	13	15
11	ETL1HV	25	10	19	20	20	7	11	18
12	EEM1BC	19	14	17	13	9	5	14	6
13	EEN1MS	18	4	10	12	9	10	3	11
14	EEO1HF	22	20	19	21	16	18	4	14
	Total	200	106	114	105	128	126	57	71
			53%	57%	53%	64%	63%	29%	36%
	F.E.T	p-valor	0.906	0.006	0.077	<0.001	<0.001	0.901	0.152

Tabela 2.
Resultado das 1299 questões-padrão (T.E.F.= Teste Exato de Fisher)

Questões Padrão – “Grupo 0”					
N	Escola	n ₀	Q2	Q4	Q6
1	EEA0HD	15	4	10	2
2	EEB0NL	52	8	17	4
3	EEC0SB	32	8	13	7
4	EED0XS	13	4	7	5
5	EME0EA	32	3	11	7
6	EEF0PM	34	4	11	7
7	EEG0GC	13	3	11	4
8	EEH0BT	42	15	6	10
	Total	233	49	86	46
			21%	37%	20%
Questões Padrão – “Grupo 1” - Teste					
		n ₁	Q2	Q4	Q6
9	EEI1FB	73	1	7	17
10	EEK1BM	43	4	9	4
11	ETL1HV	25	10	16	10
12	EEM1BC	19	6	16	6
13	EEN1MS	18	1	1	1
14	EEO1HF	22	4	16	12
	Total	200	26	65	50
			13%	33%	25%
	T.E.F	p-valor	0.787	0.116	0.140

A Tabela 1 apresenta os resultados dos dois grupos de questões experimentais. A hipótese de que o contexto inicial das questões é relevante para o desempenho de estudante não encontrou apoio nas análises estatísticas, o que configura resultado surpreendente. Ao comparar os dois grupos de respostas, as questões Q3, e principalmente Q7 e Q8 permitiram um desempenho melhor, com diferença estatisticamente significativa, quando são apresentadas sem a informação escrita de contexto. Isso nos levou a rotular as questões como indutoras de performance reversa, uma vez que sua leitura integral diminui a probabilidade de acerto.

A análise qualitativa das questões com maior *performance induzida reversa* (PIR), com ferramentas da Linguística e da Análise de Discurso, bem como a análise estatística da dispersão dos resultados, permitiu estabelecer hipóteses para explicar a razão do desempenho tão distante do esperado. Foram localizadas duas características importantes a induzir desempenho reverso; uma delas é a apresentação de termos-chave na contextualização inicial, que se repetem com destaque nas alternativas erradas e não na correta. No caso da questão de maior “PIR” (Q8) nenhum unitermo (ou equivalentes léxicos ou semânticos) presente na alternativa correta aparecia no contexto inicial. A outra característica refere-se à coesão e coerência entre contexto e comando, muito evidente nessa mesma questão.

DISCUSSÃO

Esses resultados trazem uma nova luz para pesquisas futuras, uma vez que eles revelaram a existência de “PIR” nas questões de Biologia do Novo ENEM. A ausência de desempenho desigual entre as questões

com e sem contexto é, em si, indutora de desempenho reverso, vez que leva o aluno a consumir tempo inutilmente em uma prova em que seu gerenciamento é crucial para completar todas as tarefas. Mais surpreendente ainda foi a localização de questões em que o desempenho do estudante que não lê o contexto inicial é muito superior ao daquele que o lê. A análise linguística das questões com alto “PIR” revelou a existência denexo causal nesse desvio, comprovando que o desempenho do estudante nesse tipo de questão do ENEM depende fortemente de certo tipo de habilidade de leitura, que permite o rápido descarte diferencial de informação irrelevante, com ganho de tempo.

Os chamados “cursinhos de preparação” para o ENEM, em atuação no mercado privado pelo menos desde 2003 (Zakia e Oliveira, 2003: 885), ensinam os alunos a gerenciar o tempo de forma inteligente durante o teste. A indicação de alterar a sequência de leitura das questões, iniciando pela linha de comando, faz parte das técnicas ensinadas em tais cursinhos. Essas técnicas parecem ser úteis para evitar que a concentração do estudante seja desviada para argumentos irrelevantes, bem como para gerenciar eficientemente o tempo. As ações envolvidas nessas estratégias parecem ser mais importantes para explicar o desempenho dos estudantes nas questões de Biologia do que o domínio equilibrado de conhecimento biológico e habilidades de pensamento.

Uma reforma em 2009 criou o chamado “Novo ENEM”, tornando-o não apenas um instrumento para selecionar alunos para universidades públicas, mas também para monitorar a qualidade da educação em todo o território nacional. A descoberta do fenômeno das questões com “PIR” coloca em dúvida a pertinência do uso da Teoria da Resposta ao Item (TRI) no tratamento dos resultados, com consequências profundas. Da mesma forma, o caráter certificador de conclusão da educação básica do ENEM merece ser examinado com muito cuidado.

A democratização do acesso ao ensino superior possibilitada pelo ENEM (se alguma) pode ser devida a mudanças súbitas, que demandam novas estratégias de estudo e postura no exame, e tenderia a desaparecer com o tempo, na medida em que as técnicas de gestão de tempo e descarte de informação contextual irrelevante sejam diferentemente distribuídas pelo espectro socioeconômico.

A proposta de transformar o ENEM em um exame obrigatório de Estado está atualmente em discussão. Nossos resultados sugerem que a reforma educacional e o monitoramento da qualidade da educação baseada nesta avaliação devem ser consideradas com cautela. Mais pesquisas são necessárias, abrangendo outras áreas de conteúdo, a fim de ter uma ideia mais clara do significado real do desempenho do aluno no ENEM e da real dimensão numérica das questões “PIR” no contexto de tal exame.

REFERÊNCIAS

- Bisquerra, R., J.C. Sarriera e F. Martínez. (2004). *Introdução à Estatística: enfoque informático com o pacote estatístico SPSS*. Porto Alegre: Artmed Ed.
- Franco, C.; A. Bonamino; (1999). O ENEM no contexto das políticas para o ensino médio. *Química Nova na Escola* 10: 26-31.
- Ministério da Educação. (1999). Exame Nacional do Ensino Médio – ENEM: documento básico 2000. MEC/INEP.
- /INEP. (2012) Matriz de Referência do ENEM.
- Orlandi, E. P. (2012). *Discurso e leitura*. São Paulo: Cortêz.
- Vianna, H.M. (2003) Avaliações nacionais em larga escala: análises e propostas. *Textos Fundação Carlos Chagas*, n.23. 41p.
- Zakia, S.; R. P. Oliveira. (2003). Políticas de avaliação da educação e quase mercado no Brasil. *Educação e Sociedade* 24(84): 873-895.