



# Multi-Science Journal



## Editorial

### ENEM E A TÔNICA AVALIATIVA À QUALIDADE EDUCACIONAL: INCONGRUÊNCIAS PARA ALÉM DA VISÃO ROMÂNTICA

Prof. Dr. Marcos Fernandes Sobrinho<sup>1,1</sup>

<sup>1</sup>Instituto Federal Goiano – Câmpus Urutaí – Área de Física - Urutaí, GO, Brasil

\* E-mail: [marcos.fernandes@ifgoiano.edu.br](mailto:marcos.fernandes@ifgoiano.edu.br)

*O presente editorial decorre de pequeno trecho revisado e sistematizado acerca de discussão/reflexão desenvolvida [sobre o Exame Nacional do Ensino Médio (Enem)] no âmbito de um dos capítulos de nossa tese de doutoramento (Fernandes-Sobrinho, 2016). Discorreremos brevemente sobre a temática da avaliação, dando ênfase ao Enem, tido como exame de alta relevância, e perpassando pelo histórico de mudanças de teorias e procedimentos utilizados para auferir resultados de estudantes que dele participam. Por oportuno, incongruências sinalizadas por pesquisadores da área de educação em [e para as] Ciências se juntam a outras que levantamos para, então, sugerirmos a necessidade do estabelecimento de uma agenda de discussão com propósitos reformuladores que acenem para possíveis soluções.*

O terceiro capítulo da Constituição Federal (CF) do Brasil refere-se a ações no âmbito educacional em que institui princípios de igualdade de acesso e permanência na escola, gratuidade, liberdade, garantia de padrão de qualidade e gestão democrática (Brasil, 1988, Dourado, Oliveira, & Santos, 2007).

Entre os termos elencados nessas ações, pinçamos o que se refere à “qualidade”, cujo conceito figura, recorrentemente, nos documentos e discursos vinculados a avaliações. Além disso, essas ações compõem argumentos que auxiliam em justificativas para a elaboração e implementação de políticas públicas educacionais. A partir da tônica que contorna esse conceito, avaliar processos, produtos e serviços, revela-se como imprescindível à obtenção (e à melhoria) da qualidade, não sendo diferente na educação.

A depender do contexto, o termo “avaliação” pode adquirir diversos significados e propósitos e, por

isso, implicar reflexos sobre a vida das pessoas, seja no planejamento de suas atividades, seja na divisão de tarefas ou mesmo como medida de suas performances, em determinadas situações.

Na pesquisa em educação, muito se tem debruçado sobre esse tema com discussões que envolvem os sentidos e as finalidades dos mais variados processos de avaliação. Encontramos discussões sobre esses processos em pequena escala, como no ambiente da sala de aula, e também em situações com exigência de análises mais complexas. Essas situações vão desde exames até aquelas que se dão em decursos de tempo mais longos, certamente com propósitos que visam, sobretudo, à superação de caráter meramente classificatório e excludente, em geral, pertinente às tradicionais concepções e práticas avaliativas.

Entre várias ações com propósitos avaliativos (sistemas, índices, indicadores, programas, exames, outros) discorreremos, brevemente, sobre um deles: o Exame Nacional do Ensino Médio (Enem), sinalizando algumas incongruências sem, no entanto, aprofundar a discussão e, por óbvio, sem qualquer pretensão de esgotar essas questões diante dos propósitos de um editorial. Comentamos, rapidamente, a respeito de mudanças e teorias utilizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), para auferir o que denominam “desempenho dos estudantes”, no Enem.

Tomemos inicialmente a natureza conceitual das objetivações do exame, quando de sua criação. Concebido, lá em 1998, o exame manteve sua formatação básica até 2008, com o “objetivo fundamental de avaliar o desempenho do aluno ao término da escolaridade básica, para aferir o desenvolvimento de competências fundamentais ao

<sup>1</sup> Realiza estudos de Pós Doutorado no Programa de Pós-Graduação em Educação para a Ciência, da Faculdade de Ciências, na Universidade Estadual Paulista “Júlio de Mesquita Filho” – Câmpus Bauru (PPGEC/FC/Unesp/Bauru).

exercício pleno da cidadania” (Brasil, 2005; Fernandes-Sobrinho, Ramos, & Santos, 2016; Sobrinho, Ramos, & dos Santos, 2013; Fernandes-Sobrinho & Santos, 2014; Almeida, Silva, Fernandes, & Fernandes-Sobrinho, 2016). Nesse período, a nota do candidato era medida pelo total de acertos nos itens da prova objetiva, aos moldes da Teoria Clássica dos Testes (TCT).

Isso posto, como seria possível, com apenas provas objetivas, o exame aferir aspectos do “desenvolvimento de competências” da atividade humana, atrelados aos pressupostos do conceito de cidadania, em geral, vinculados a valores sociais, ao agir responsável?

Essa metodologia – TCT – e suas limitações, inclusive aspectos relativos aos fundamentos matemáticos, podem ser consultados em Gulliksen (1950), Lord & Novick (1968) e Vianna (1987). Com base nessas limitações, a partir de 2009 o Enem passou a adotar a Teoria de Resposta ao Item (TRI) que, segundo Andrade, Tavares e Valle, (2000), trata-se de:

[...] um conjunto de modelos matemáticos que procura representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. Os vários modelos propostos na literatura dependem fundamentalmente de três fatores: (i) da natureza do item - dicotômicos ou não dicotômicos; (ii) do número de populações envolvidas - apenas uma ou mais de uma; (iii) e da quantidade de traços latentes que está sendo medida - apenas um ou mais de um.

Significa dizer, grosso modo, que a TRI estima as dificuldades dos itens e a proficiência dos participantes, mas também possibilita que os itens, mesmo pertencentes a diferentes edições do exame, situem-se em um mesmo estrato de uma determinada escala.

Aqui, consideramos relevante focarmos a discussão em modelos que avaliam apenas um traço latente ou habilidade, os chamados modelos unidimensionais [a razão para isso, oportunamente explicamos adiante]. Essa unidimensionalidade pressupõe a homogeneidade do conjunto de itens. Ainda que haja exigência de outras habilidades para execução de tarefas, a que representa em um teste, o traço latente é aquela habilidade dominante, requerida à resolução de determinado item.

Este modelo pressupõe outra necessidade que é a independência local (ou condicional), a qual admite que, para uma determinada habilidade, as respostas são independentes, para diferentes itens da prova, o que é imprescindível para fazer uma boa estimativa dos parâmetros que são levados em consideração, no modelo. Dado que a unidimensionalidade implica independência condicional, significa que os itens obrigatoriamente precisam ser elaborados de sorte a satisfazer esta suposição de unidimensionalidade.

Para o caso da avaliação educacional, comumente se define uma população (representada por um ou mais grupos, entendidos aqui como uma ou mais

amostras da população) por características que podem variar, a depender dos propósitos do estudo e que, por decorrência, podem ou não ser relevantes para a distinção de populações.

Apesar de ser uma teoria desenvolvida na década de 1960, somente nos últimos anos, em diversas áreas do conhecimento, especialmente em avaliação educacional, temos observado um aumento expressivo de aplicações de técnicas estatísticas consequentes da TRI, que propõe modelos para traços latentes, traduzidas por características do indivíduo não observáveis diretamente. Esse padrão de variável deve ser inferido a partir da observação de outras, secundárias, relacionadas a ela.

O que esta metodologia sugere são formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes, proficiências ou habilidades na área de conhecimento avaliada (Andrade et al., 2000).

Entre esses traços latentes, podemos entendê-los como variáveis descritas e listadas com certa facilidade. A título de exemplo, citamos aspectos ligados à ansiedade, inteligência, habilidade na execução de determinada ação, nível de compreensão de um texto etc., mas não mensuráveis diretamente como o comprimento de um veículo, a altura de uma pessoa, ou a massa de um objeto.

Uma das grandes vantagens da TRI sobre a TCT reside no fato de ela permitir a “comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes” (Andrade et al., 2000, p. 3), dado que uma das características mais relevantes da TRI é ter como elementos centrais os itens, e não o conjunto destes que compõe o todo da prova.

Entre os modelos utilizados na TRI destacamos o modelo logístico unidimensional de três parâmetros (ML3). O ML3 leva em consideração: a) a dificuldade do item; b) a discriminação do item: diferença entre as médias do item, calculadas para os 27% que atingiram maiores escores no teste, dito grupo superior, e para os 27% com menores escores no teste, dito grupo inferior (Silveira, 1982); e c) a probabilidade de uma resposta correta ter sido dada por indivíduos de baixa habilidade. O índice de discriminação do item dá consistência (ou confiabilidade) ao teste.

O MEC/Inep, com olhares voltados à qualidade de suas avaliações, tem recorrido à TRI no processamento dos resultados nas aplicações de suas provas e geração dos escores dos alunos. Este é o caso do processamento das notas do Enem, em que os escores (resultados obtidos) representam e vão muito além da simples contagem de itens assinalados como certos. Utiliza a TRI sob o modelo logístico de três parâmetros (ML3) [por isso focamos nossa discussão em modelos unidimensionais: é o utilizado pelo MEC/Inep], dado que possibilita comparar resultados entre diversas fases da avaliação. Aqui há um rompimento ao que antes de 2009 era colocado em prática no Enem.

Até 2008, um bom desempenho se dava a partir do número de acertos, valendo inclusive, a possibilidade do “chute”, sem qualquer prejuízo ao respondente. A partir da implementação da TRI, no Enem, compreender e desenvolver linhas de raciocínio aumentam as chances de acerto dos itens. Embora o “chute”, em tese, ainda ocorra [ou possa ocorrer], passa a ser substituído por uma alternativa mais coerente de resposta, em consonância à inteligência particular daquele que responde o item.

Feitas as considerações acima sobre a TRI, apresentamos críticas que permeiam a discussão acerca da validade e aceitação do exame, por pesquisadores na área de educação em Ciências (EC), seguida por nossas considerações.

Nessa direção, e para o caso dos itens de Física, entre os presentes na prova de Ciências da Natureza e suas Tecnologias (CNT), da edição de 2009, Gonçalves Júnior (2012) chama a atenção para o fato de que, em geral, esses itens podem apresentar mais de uma habilidade o que, segundo esses autores, contraria a unidimensionalidade da TRI. Além disso, identificaram uma distribuição não homogênea dos itens analisados, para os objetos específicos de conteúdos de Física, privilegiando temas que envolvem conceitos de eletricidade e termodinâmica e, com isso, podendo implicar distorções no currículo de Física, ao nível médio de ensino, no Brasil.

Em outro estudo realizado sobre as questões da área de CNT do Enem, Costa-Beber (2012) alerta para a existência de certos distanciamentos entre o que se preconiza nos documentos oficiais e o que se encontra nos itens desse exame. A autora evidencia a necessidade de se investir na melhora qualitativa e de natureza tipológica das questões do exame para que, assim, “possa efetivamente avaliar o que se propõe” (Costa-Beber, 2012).

Em outra publicação, destacam-se a identificação de diversos itens com sérios problemas de formulação, enunciados no comando dos itens inconsistentes com o conhecimento científico, outros com mais de uma ou nenhuma alternativa de resposta correta (Silveira & Barbosa, 2014). Segundo estes autores, a Sociedade Brasileira de Física (SBF) se manifestou, em 2013, sobre a qualidade dos itens de Física do Enem, apontando entre vários equívocos, o uso exacerbado e forçado de contextualização nos comandos das questões.

Já Lopes & López (2010) afirmam que o exame nasce em meio a um contexto, inclusive internacional, em que os olhares se voltam às avaliações padronizadas e vinculadas a políticas de responsabilização da Escola. No entanto, os documentos oficiais, a ele relacionados, não fazem alusão clara, especificamente, à finalidade da “prestação de contas” de escolas. Preocupa-se, porém, com o atingimento dos estudantes a outros níveis de ensino do sistema educacional. Desse modo, o exame acaba promovendo inversão piramidal acerca das políticas públicas. Significa dizer que, em vez de o exame agir sobre o sistema em busca da promoção de resultados no indivíduo, trabalha sobre o indivíduo, a fim de gerar desdobramentos no sistema.

Diante do todo exposto, incorporamos outra crítica acerca da divulgação, pelo MEC/Inep, do ranking de escolas (Andrade & Soida, 2012; Fernandes-Sobrinho, 2016; Gauche, 2016), a partir dos resultados do Enem, pelo que indagamos: que sentido tem divulgar médias auferidas pelas escolas, junto ao Enem, que utiliza a Teoria de Resposta ao Item (TRI) para estimar a proficiência dos estudantes, a partir de pequenas amostras? Afinal, dado que o exame é optativo, nem todos os estudantes das escolas o fazem.

Notemos que, já não se consegue avaliar 100% da população de estudantes, pelo seu caráter optativo. Além disso, ainda que bastante evoluída, frente à outra utilizada antes (TCT), a TRI está muito distante de nos permitir avaliar 100% dos objetivos educacionais. Estes sólidos argumentos, entre muitos outros (efeito-escola; nível socioeconômico dos participantes), permitem-nos afirmar que não faz sentido falarmos em ranking das escolas, a partir de dados divulgados do Enem!

Nesse sentido e dando seguimento às nossas provocações, perguntamos ao leitor, sugerindo que busquem respondê-las. Afinal, os critérios de validade de testes têm seus limites? Não é dever do nosso Poder Público observar esses critérios e suas limitações? Há um código de ética capaz de regular as ações dos vários sujeitos que participam dessa atividade? O que dizer acerca da indústria instituída no Brasil, voltada para (e que se beneficia indevidamente dos) elementos do Enem (resultados e ranking de escolas), como se fossem critérios válidos para o contorno da avaliação educacional?

O fato é que os domínios do conhecimento, de interesse expressivo e necessário à vida acadêmica, no ensino superior, dificilmente conseguem ser mensurados nesse exame. Não bastasse, o ensino médio (EM) se traduz na etapa final da nossa educação “básica” que, pela sua própria natureza, e cada dia mais, revela-se excludente daquilo que deixa de ser básico [com desculpas à redundância que, diga-se de passagem, não é nossa!].

O que consideramos erro crasso é enfatizar o básico, dado que se corre o risco de ofuscar o que, por vezes é, de fato, bastante relevante. Será que temos presenciado as consequências desse corolário, na vida acadêmica dos estudantes que têm ingressado, ultimamente, nas Instituições de Ensino Superior (IES)?

Lamentavelmente, isso tem sido objeto de mau uso, dado que cada vez mais se tem utilizado como forma de propagar supostos parâmetros, que atestam a qualidade de uma escola em detrimento de outras. Não se leva em conta, sobretudo nas publicidades, aspectos como: efeito-escola (Alves & Soares, 2007), nível socioeconômico (Kleinke, 2017) e desdobramentos indesejáveis quanto à redução curricular e significados por trás das avaliações em larga escala (Schneider, 2013). Avaliações dessa natureza, por si só, têm como objetivo, no sistema escolar, informar o que populações e subpopulações (ou grupos) de estudantes, em diferentes ambientes, sabem e são capazes de fazer em um determinado momento; e acompanhar sua evolução ao longo dos anos. Necessário se faz enfatizar que, nem de longe, têm como objetivo fornecer informações

acerca de alunos ou escolas individuais (Klein & Fontanive, 1995).

Para estes autores, a expressão exame em larga escala, diz respeito a um número bastante elevado de testes aplicados ou à utilização de certa amostragem estatística, geralmente caracterizada por permitir o monitoramento contínuo dos processos educativos, e que possibilitam detectar benefícios e(ou) malefícios consequentes de políticas públicas educacionais.

Como é possível perceber, embora haja adesão crescente e exponencial ao exame, por assim dizer, pela comunidade (estudantes e IES), há ainda muitos ruídos com apontamentos que colocam em cheque a validação do Enem. Isso requer, da comunidade acadêmica e dos órgãos responsáveis por avaliações em larga escala, o estabelecimento de uma agenda de discussão e, certamente, com reformulações pertinentes que acenem para a solução dessas incongruências, entre outras tantas, não trazidas aqui.

Até o fechamento deste editorial (29 de março de 2017), não registramos pronunciamento e/ou publicação oficial por parte das autoridades do MEC/Inep sobre mudanças a serem implementadas a partir da edição do Enem de 2017, o que deverá ocorrer, segundo a assessoria do órgão, no início do mês de abril, do corrente ano. Apesar disso, temos certa expectativa no sentido de não mais convivermos com a divulgação de Ranking de escolas no exame, pelas razões apresentadas e discutidas no presente texto.

Seguimos com propósito de dialogar com a literatura e com as autoridades, de sorte a ampliarmos as discussões e o debate em torno de encaminhamentos sinalizados pela literatura (Alves & Soares, 2007; Fernandes-Sobrinho, 2016; Gauche, 2016; Kleinke, 2017, entre outros) que se debruçam em estudos sobre políticas públicas educacionais e seus desdobramentos no contexto do ambiente escolar, incluem-se as relacionadas ao largo espectro das avaliações.

#### Referências

- Almeida, A. B. B., Silva, L. A. S., Fernandes, P. S. R., & Fernandes-Sobrinho, M. (2016). Potenciais aspectos sociocientíficos em itens de biologia do ENEM. *Indagatio Didactica*, 8(4).
- Alves, M. T. G., & Soares, J. F. (2007). Efeito-escola e estratificação escolar: o impacto da composição de turmas por nível de habilidade dos alunos. *Educação em Revista*, (45), 25-59.
- Andrade, D. F., Tavares, H. R., & Valle, R. da C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. São Paulo: Associação Brasileira de Estatística.
- Andrade, E. de C., & Soida, I. A. I. (2012). O ranking das escolas de ensino médio baseado no ENEM é confiável? Instituto de Ensino e Pesquisa.
- Brasil. (2005). *Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).
- Brasil. Constituição (1988). *Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.

Costa-Beber, L. B. (2012). *Reorganizações curriculares na conquista da educação escolar de melhor qualidade: expectativas acerca do efeito indutor do novo ENEM (Mestrado em Educação nas Ciências)*. Unijuí, Ijuí.

Dourado, L. F., Oliveira, J. F. de, & Santos, C. de A. (Orgs.). (2007). *A Qualidade da Educação: conceitos e definições*. Brasília: Inep/MEC. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Fernandes-Sobrinho, M. (2016). *Temas sociocientíficos no Enem e no livro didático: limitações e potencialidades para o ensino de Física (Doutorado)*. Universidade de Brasília (UnB), Brasília.

Fernandes-Sobrinho, M., Ramos, T. C., & Santos, W. L. P. dos. (2016). *Temas sociocientíficos (des)velados no Enem: potencialidades à ampliação de fontes e de gêneros textuais ao ensino de Física*. *Indagatio Didactica*, 8(1), 514-533.

Fernandes-Sobrinho, M., & Santos, W. L. P. dos. (2014). *Inserções da interdisciplinaridade e contextualização em itens do enem/2013 com potencial ao enfoque CTS*. *Uni-pluri/versidad*, 14(3), 94.

Gauche, R. (2016). *Avaliação: Discursos Infundados, Resultados Tristemente Esperados...* *Revista Thema*, 5-7. <https://doi.org/10.15536/thema.13.2016.5-7.408>

Gonçalves Júnior, W. P. (2012). *Avaliações em larga escala e o professor de Física (Mestrado)*. UFRJ, Rio de Janeiro.

Gulliksen, H. (1950). *Theory of Mental Tests*. New York: John Wiley and Sons.

Klein, R., & Fontanive, N. S. (1995). *Contemporaneidade e educação (Vol. 66)*. Brasília.

Kleinke, M. U. (2017). *Influência do status socioeconômico no desempenho dos estudantes nos itens de física do Enem 2012*. *Revista Brasileira de Ensino de Física*, 39(2). <https://doi.org/10.1590/1806-9126-RBEF-2016-0081>

Lopes, A. C., & López, S. B. (2010). *A performatividade nas políticas de currículo: o caso do ENEM*. *Educação em revista*, 26(1), 89-110.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Schneider, M. P. (2013). *Políticas de avaliação em larga escala e a construção de um currículo nacional para a Educação*. *Eccos*, (30), 17.

Silveira, F. L. da. (1982). *Considerações sobre o índice de discriminação de itens em testes psicrométricos e educacionais*, 31-37.

Silveira, F. L. da, & Barbosa, M. C. (2014). *Análise do ENEM-2014 pelo Conselho do IF UFRGS*. Sociedade Brasileira de Física (SBF).

Sobrinho, M. F., Ramos, T. C., & dos Santos, W. L. P. (2013). *Questões de Física do ENEM/2012 com caráter sociocientífico: um estudo do potencial argumentativo à luz do padrão de Toulmin*. Atas IX Encontro Nacional de Pesquisa em Educação em Ciências. Águas de Lindóia-SP.

Vianna, H. M. (1987). *Testes em Educação*. São Paulo: IBRASA.