

## **Análise da estrutura interna do ENEM com foco em Ciências Naturais**

Rodrigo Travitzki<sup>1</sup>, Ricardo Primi<sup>2</sup>

<sup>1</sup> Universidade São Francisco, r.travitzki@gmail.com

<sup>2</sup> Universidade São Francisco, ricardo.primi@usf.edu.br

### **Resumo**

Neste trabalho, investiga-se um aspecto da validade das provas do ENEM (a estrutura interna), com especial atenção à de Ciências Naturais. Foram analisados os microdados de 2014 e 2015. A análise fatorial sugere que um fator é suficiente para descrever a variância dos 175 itens. O modelo bifator sugere heterogeneidade sistemática, nos dois anos, entre os subfatores, especialmente nas provas de Matemática e de Linguagens e Códigos. Por outro lado, a média das 4 provas se mostrou colinear ao fator g. Em relação à prova de Ciências Naturais, observou-se evidências de considerável multidimensionalidade interna, o que poderia ser uma característica própria do letramento científico. São apontadas possíveis decorrências para as políticas públicas.

**Palavras-chave:** avaliação educacional, validade, modelo bifator, multidimensionalidade, raciocínio científico

### **1. Introdução**

O Exame Nacional do Ensino Médio (ENEM) é uma importante referência para a educação brasileira e carecem estudos de validade das provas a partir de 2009, quando o Exame passou por transformações estruturais. Há evidências de que as primeiras edições do ENEM avaliavam mais a capacidade de raciocínio (inteligência fluida) do que o conhecimento propriamente dito (inteligência cristalizada) (PRIMI et al., 2001). Sobre as edições recentes há poucos estudos de validade – ver por exemplo (TRAVITZKI, 2017) – mas nenhum com a metodologia aqui proposta.

A estrutura interna é uma das fontes de informação para se verificar a validade de um teste (AERA; APA; NCME, 2014). No caso do ENEM, cabe investigar em que medida os itens das quatro provas correspondem a quatro construtos minimamente distintos. Também cabe verificar se há um construto subjacente aos itens da prova de Ciências Naturais e, em caso

afirmativo, se este construto pode ser interpretado como uma forma de raciocínio científico. Tais objetivos norteiam este trabalho.

O que seria o raciocínio científico no contexto de um exame ao final da educação básica? Há evidências de que crianças de 4 a 6 anos possuem elementos básicos do raciocínio científico, como relação entre hipótese e evidência, ou mesmo percepção de covariância (KOERBER et al., 2005). Segundo Stephen Norris e Linda Phillips, o letramento científico<sup>1</sup> tem uma particularidade em relação aos outros tipos de letramento, pois dependente mais de conhecimento prévios específicos. Nesse sentido, os autores consideram que há dois aspectos do letramento científico: 1) o fundamental (capacidade de compreender, interpretar, analisar e criticar qualquer texto) e 2) o derivado (dependente de conteúdos específicos de cada ciência). No entanto, concluem, o sentido fundamental é pouco aplicado na educação científica, justamente em virtude da importância dos conteúdos específicos neste tipo de letramento (NORRIS; PHILLIPS, 2003).

## **2. Metodologia**

Realizou-se inicialmente a análise fatorial dos 175 itens<sup>1</sup> de cada ano, a partir das correlações tetracóricas. A análise fatorial é um dos fundamentos psicométricos para investigação da inteligência (PRIMI et al., 2001). No contexto das avaliações educacionais, é comum observar-se um fator predominante na prova. Uma questão importante, neste caso, é verificar em que medida este fator predominante corresponde ao construto planejado, e em que medida corresponde a habilidades mais genéricas, relacionadas à resolução de provas, tais como a inteligência fluida e a velocidade de processamento – na nomenclatura da Teoria de Cattell-Horn-Carroll (PRIMI, 2003).

Para buscar respostas a esta questão, utilizou-se modelos bifator, que foram inicialmente propostos como alternativa à estrutura oblíqua simples de Thurstone, contemplando diferentes níveis de complexidade no comportamento psicológico (SCHMID; LEIMAN, 1957). Com efeito, tal tipo de modelo permite que se isole o fator predominante do teste a fim de se investigar aspectos mais específicos de sua estrutura interna. É estimado um fator geral (presente em todos os itens) e alguns subfatores (presentes em subconjuntos de itens), que corresponderiam a construtos conceitualmente mais específicos. Normalmente assume-se que o fator geral é ortogonal aos subfatores. Os modelos bifator permitem: a) investigar o particionamento da variância quando se acredita que há um fator mais geral e alguns subfatores; b) o controle da multidimensionalidade em testes essencialmente unidimensionais;

c) avaliar se o fator geral é forte o suficiente para se justificar modelos unidimensionais; d) determinar a adequação do score geral, e se há algum ganho com a inclusão dos sub scores (RODRIGUEZ; REISE; HAVILAND, 2016). Embora não tenham sido utilizados neste trabalho, há também modelos bifator fundamentados na Teoria da Resposta ao Item (TOLAND, 2017).

Foram analisados os microdados do ENEM de 2014 e 2015 com o software R, pacote Psych (REVELLE, 2018). A amostra incluiu apenas os concluintes do Ensino Médio no respectivo ano, provenientes de escolas regulares, presentes nos dois dias. Todos os cadernos foram incluídos, com uma exceção.<sup>2</sup> Para tanto, os itens foram reordenados com as informações fornecidas nos próprios microdados.<sup>3</sup> A referência para a ordenação de itens aqui apresentada é o caderno azul para o sábado – Ciências Naturais (CN) e Ciências Humanas (CH) – e amarelo para o domingo – Matemática (MT) e Linguagens e Códigos (LC). Foi utilizada uma amostra aleatória de 300 mil alunos, em virtude de limitações computacionais.

### 3. Resultados e Discussão

As oito provas apresentaram coeficiente Alfa de Cronbach aceitável (maior do que 0,7), sendo os valores mais baixos nas provas de CN (0,76 em 2014 e 0,79 em 2015). Analisando o conjunto das quatro provas, o coeficiente alfa foi de 0,95, embora tenham sido encontrados alguns itens negativamente correlacionados com o todo (4 em 2014 e 3 em 2015). Cabe notar que um coeficiente alfa maior do que 0,9 pode indicar número excessivo de itens.

A análise paralela dos 175 itens revelou, nos dois anos, um fator predominante e um segundo fator menor também proeminente, além de diversos outros potencialmente significativos. Analisando as quatro provas separadamente, as de MT e CN apresentaram maior número de componentes do que as outras duas (especialmente CN, com 7 componentes em 2014 e 9 em 2015).<sup>1</sup> Tais resultados são compatíveis com a hipótese de Norris e Phillips (2003), relacionada à maior necessidade de conhecimentos específicos prévios para o letramento científico. A análise fatorial revelou que um fator seria o suficiente para contemplar a variância comum. Por outro lado, uma análise mais completa pode ser realizada com três fatores, tendo em vista que a análise com um fator captou 12% da variância total, enquanto a análise com três fatores captou 14%.

Foi estimado também um modelo bifator com três subfatores para cada ano, contemplando os 175 itens. O coeficiente ômega hierárquico estima a proporção da variância no score total que pode ser atribuída ao fator geral, tratando como erro de medida a variância “decorrente” dos

subfatores (RODRIGUEZ; REISE; HAVILAND, 2016). Quando comparado ao ômega total, ele indica a adequação de modelos unidimensionais e possíveis ganhos com modelos multidimensionais. No ENEM 2014, o coeficiente ômega hierárquico foi de 0,77 e em 2015 de 0,72, sendo que nos dois anos o coeficiente ômega total foi de 0,95. Com efeito, cerca de um quinto da variância nos scores únicos pode ser atribuído à multidimensionalidade.

Porém, quando se observa a média no ENEM, o impacto da multidimensionalidade se mostra irrelevante. A correlação entre a média das 4 notas e o score do fator geral no modelo bifator é de 0,97 em 2015. Em relação aos subfatores, não chega a 0,4. A correlação entre o score do fator geral e as 4 notas separadamente é um pouco menor, especialmente em CN e MT (0,8), mas ainda é alta.

A nota de Ciências Naturais, ainda em 2015, apresentou maior correlação com o fator geral (0,80), em seguida com o sub fator F2 (0,33) – que também é fracamente correlacionado com MT (0,28). Tais resultados sugerem pouca especificidade associada à nota de CN, dificultando a identificação de um construto relacionado ao raciocínio (ou letramento) científico na prova de Ciência Naturais do ENEM.

	2014				2015			
	CH	CN	LC	MT	CH	CN	LC	MT
g	0,53	0,33	0,52	0,37	0,54	0,37	0,47	0,36
F1	0,51	0,13	0,64	-0,02	0,52	0,09	0,63	-0,06
F2	0,23	0,40	0,04	0,59	0,27	0,50	0,11	0,47
F3	0,07	0,23	-0,09	0,63	0,22	0,25	-0,02	0,59
h2	0,45	0,27	0,47	0,44	0,47	0,32	0,48	0,35

Tabela 1: coeficientes de congruência entre fatores da estrutura interna observada (no modelo bifator) e a estrutura esperada idealmente (quatro provas). Observa-se o fator geral (g), os três subfatores (F1 a F3) e a comunalidade (h2). Fonte: elaboração própria.

Para analisar a relação entre a estrutura identificada no modelo bifator e a estrutura de quatro provas do ENEM, calculamos o coeficiente de congruência entre os fatores do modelo e as provas (Tabela 1).<sup>1</sup> Tal coeficiente pode ser interpretado como uma correlação. Nota-se uma estrutura interna semelhante nos dois anos, especialmente clara nas provas de MT e LC. Levando em conta que os subfatores contemplam os resíduos do fator geral, isso significa que, para além das habilidades gerais necessárias para a resolução de provas, há diferenças

claras entre as habilidades avaliadas pelas provas de MT e LC. Embora elas sejam pouco impactantes nas notas do ENEM.

Em relação à prova de Ciências Naturais, observa-se novamente certa proximidade com o sub fator F2. No entanto, este fator está igualmente presente nas provas de MT, dificultando a identificação de um fator específico para CN. O mesmo ocorre entre Ciências Humanas e F1. A análise fatorial da prova de Ciências Naturais 2015 sugere que um único fator seria suficiente para representar a variação comum, explicando 10% da variância nos 45 itens. A análise paralela mostra a existência de um segundo fator proeminente e um total de 15 fatores (ou 9 componentes) com auto valor maior do que o esperado aleatoriamente. O coeficiente Alfa foi 0,8 (considerado adequado, supondo-se a unidimensionalidade). No modelo bifator, o ômega hierárquico foi de 0,64 (ômega total = 0,81) e em 2014 foi de 0,32 (0,78), sugerindo que um score único não representa adequadamente a estrutura interna da prova de CN.

Observou-se também a estrutura de 8 competências e 3 áreas de conhecimento.<sup>2</sup> Analisando os coeficientes de congruência, não foi possível identificar relações consistentes entre os fatores e as competências ou áreas. Analisando as correlações inter item, cada uma das competências (e áreas em menor grau) apresentou coerência interna maior do que a prova como um todo, embora muitas vezes de forma não significativa ( $p < 0,05$ ).<sup>3</sup>

Tais resultados mostram que, embora a estrutura de competências pareça produzir algum efeito empírico, há dificuldades na interpretação da estrutura interna da prova de CN, não apenas na interpretação desta prova em relação ao ENEM como um todo.

#### 4. Conclusões

A estrutura interna do conjunto de 175 itens se mostrou unidimensional na prática, embora tenham sido identificados três subfatores, além do fator g. O score do fator g apresentou correlação de 0,97 com a média das 4 provas. Os subfatores apresentaram congruência sistemática, em 2014 e 2015, com a estrutura de quatro provas do ENEM, observando-se diferença clara entre os itens de Matemática e de Linguagens e Códigos.

Na prova de Ciências Naturais, não foi possível identificar relações consistentes entre a estrutura proposta e a observada, tanto em relação às outras provas quanto internamente (nas 3 áreas ou 8 competências). Há apenas evidências moderadas a favor da estrutura de competências. O coeficiente ômega hierárquico foi de 0,32 em 2014, revelando sérias limitações da nota de CN para representar a multidimensionalidade interna desta prova – resultado compatível com a hipótese de Norris e Phillips sobre o letramento científico.

Em relação às políticas públicas, os resultados apontam necessidade de se aprimorar a especificidade dos construtos relativos a cada prova do ENEM, especialmente em Ciências Naturais e Humanas. Nas condições atuais, há muitos itens – além de considerável heterogeneidade interna – contrastando com a pouca variedade de informação disponibilizada sobre cada aluno. A média no exame corresponde ao fator g e a nota na prova de CN não se refere a uma competência científica geral. Nesse sentido, a análise psicométrica aponta para a transformação do ENEM em alguns exames menores, com construtos mais específicos, o que também contribuiria para a diminuição de fraudes. Por outro lado, do ponto de vista educacional e da gestão, outros argumentos podem contrariar esta recomendação, o que não foi investigado neste trabalho.

## Referências

- AERA; APA; NCME. *Standards for educational and psychological testing*. Washington: American Educational Research Association, 2014.
- KOERBER, S.; SODIAN, B.; THOERMER, C.; NETT, U. Scientific Reasoning in Young Children: Preschoolers' Ability to Evaluate Covariation Evidence. *Swiss Journal of Psychology*, v. 64, n. 3, p. 141–152, set. 2005.
- NORRIS, S. P.; PHILLIPS, L. M. How Literacy in Its Fundamental Sense Is Central to Scientific Literacy. *Science Education*, v. 87, n. 2, p. 224–240, 2003.
- PRIMI, R. Inteligência: Avanços nos Modelos Teóricos e nos Instrumentos de Medida. *Avaliação Psicológica*, v. 1, p. 67–77, 2003.
- PRIMI, R.; SANTOS, A. a. A. Dos; VENDRAMINI, C. M.; TAXA, F.; MULLER, F. A.; LUKJANENKO, M. D. F.; SAMPAIO, I. S. Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, v. 17, n. 2, p. 151–159, ago. 2001.
- REVELLE, W. *psych*: Procedures for Psychological, Psychometric, and Personality Research. Disponível em: <<http://cran.r-project.org/package=psych>>. Acesso em: 4 mar. 2018.
- RODRIGUEZ, A.; REISE, S. P.; HAVILAND, M. G. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, v. 21, n. 2, p. 137–150, 2016.
- SCHMID, J.; LEIMAN, J. M. The development of hierarchical factor solutions. *Psychometrika*, v. 22, n. 1, p. 53–61, 1957.
- TRAVITZKI, R. Avaliação da qualidade do Enem 2009 e 2011 com técnicas psicométricas. *Estudos em Avaliação Educacional*, v. 28, n. 67, p. 256, 28 abr. 2017.
- TOLAND, M. et al. Introduction to bifactor polytomous item response theory analysis. *Journal of school psychology*, v. 60, p. 41–63, 2017.