



Evidências Desfavoráveis ao Postulado de Cargas Fatoriais Simples do Exame Nacional do Ensino Médio (ENEM)

Cristiano Mauro Assis Gomes¹

Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brasil

Hudson Fernandes Golino

Universidade da Virgínia, Estados Unidos

Alexandre José de Souza Peres

Universidade Federal de Mato Grosso do Sul, Campo Grande-MS, Brasil

RESUMO

O Exame Nacional do Ensino Médio (ENEM) possui um modelo de medida restrito, caracterizado pelo postulado de que os itens marcadores de cada domínio se vinculam exclusivamente ao seus domínios-alvo. Estudos preegressos sugerem, por meio de evidências indiretas, que esse modelo não seria válido. No entanto, esse postulado ainda não foi diretamente avaliado. Neste estudo, investiga-se esse pressuposto por meio de análises fatoriais dos itens do ENEM de 2011. Dois modelos foram testados. O primeiro, chamado de estrutura simples de Thurstone, representa o modelo de medida do ENEM. O segundo, de cargas cruzadas, refuta esse modelo. O modelo das cargas cruzadas foi o único que apresentou bom ajuste aos dados de acordo com todos os índices empregados. As evidências encontradas são desfavoráveis ao postulado de cargas fatoriais simples do modelo de medida do ENEM, indicando problemas de validade e na qualidade dos escores produzidos.

Palavras-chave: Exame Nacional do Ensino Médio (ENEM); validade fatorial; análise fatorial confirmatória; modelagem por equações estruturais exploratórias.

ABSTRACT – Evidence Contrary to the Assumption of Simple Factor Loadings of the National High School Education Exam (ENEM)

The National High School Examination (ENEM) has a restricted measurement model characterized by the assumption that the marker items in each domain are exclusively linked to their target domain. Previous studies suggest, through indirect evidence, that this model may not be valid. However, this postulate has not yet been directly assessed. In this study, this assumption was investigated through factor analysis of the items of the ENEM 2011 edition. Two models were tested. The first, called Thurstone's simple structure, represents the measurement model of the ENEM. The second, of crossed loadings, refutes this model. The crossed loadings model was the only one that presented a good fit to the data according to all the indices employed. The evidence found is unfavorable for the assumption of simple factor loadings of the measurement model of the ENEM, indicating issues of validity and in the quality of the scores produced.

Keywords: National High School Examination (ENEM); factor validity; confirmatory factor analysis; exploratory structural equation modeling.

RESUMEN – Evidencias Desfavorables al Postulado de Cargas Factoriales Simples del Examen Nacional de Enseñanza Secundaria (ENEM)

El Examen Nacional de Enseñanza Secundaria (ENEM) tiene un modelo de medición restringido, caracterizado por el postulado de que los ítems marcadores en cada dominio están vinculados exclusivamente a su dominio objetivo. Estudios previos sugieren, a través de evidencia indirecta, que este modelo no sería válido. Sin embargo, este postulado aún no ha sido evaluado directamente. En este estudio se investiga este supuesto a través del análisis factorial de los ítems del ENEM 2011. Se probaron dos modelos. El primero, llamado estructura simple de Thurstone, representa el modelo de medición ENEM. El segundo, de cargas cruzadas, refuta este modelo. El modelo de carga cruzada fue el único que presentó un buen ajuste a los datos de acuerdo con todos los índices empleados. Las evidencias encontradas son desfavorables al supuesto de cargas factoriales simples del modelo de medición del ENEM, lo que indica problemas de validez y en la calidad de las puntuaciones producidas.

Palabras clave: Examen Nacional de Enseñanza Secundaria (ENEM); validez factorial; análisis factorial confirmatorio; modelo de ecuaciones estructurales exploratorias.

O Ministério da Educação (MEC) brasileiro instituiu em 1998 um teste educacional padronizado para avaliar um conjunto de competências dos estudantes ao

final do Ensino Médio (Brasil, 1998). O ENEM (Exame Nacional do Ensino Médio), planejado e operacionalizado pelo Instituto Nacional de Estudos e Pesquisas

¹ Endereço para correspondência: Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, Gabinete 4036, Pampulha, 31270-901, Belo Horizonte, MG. E-mail: cristianomaurogomes@gmail.com

Educacionais Anísio Teixeira (Inep), autarquia vinculada ao MEC, foi, então, concebido a partir de uma abordagem construtivista, com foco no raciocínio e na resolução de problemas (Inep, 2005). O ENEM se propunha a avaliar cinco competências que o estudante do ensino médio deveria desenvolver ao longo do Ensino Fundamental e do Ensino Médio (Gomes, 2010; Inep, 2005; Primi et al., 2001). Essa versão do ENEM foi construída a partir de uma proposta interdisciplinar, representada em uma matriz de habilidades e competências, sendo que cada uma poderia envolver conhecimentos de diferentes áreas (Gomes & Borges, 2009). Até 2008, uma prova objetiva com 63 itens e uma prova de redação eram elaboradas por ano e aplicadas aos estudantes que desejassem realizá-la (Inep, 2005).

A partir de 2009, o ENEM sofreu modificações importantes, pois tornou-se um instrumento utilizado também seleção de estudantes para o acesso ao Ensino Superior. O foco nas cinco competências foi deslocado para quatro domínios de conteúdo: ciências da natureza e suas tecnologias (CN), ciências humanas e suas tecnologias (CH), matemática e suas tecnologias (MT), e linguagens, códigos e suas tecnologias (LC). O foco interdisciplinar ficou mais restrito em relação ao que se observava na versão anterior do exame, mas foi preservado em algum nível (Inep, 2014, 2018). Dentro de cada domínio, há competências que envolvem conhecimentos de mais de uma área. Por exemplo, o domínio CN está relacionado a conhecimentos de física, química e biologia. No entanto, não mais se pressupõe que um item possa avaliar conhecimentos de mais de um desses domínios de conteúdo. A prova objetiva passou de 63 para 180 itens, de modo que cada domínio pudesse ser avaliado por 45 itens específicos (Inep, 2014).

Diversas questões são suscitadas quanto aos aspectos pedagógicos, psicométricos e logísticos envolvidos na elaboração, na aplicação, na correção e na divulgação e uso dos resultados de testes padronizados aplicados em larga escala como o ENEM (AERA et al., 2014; Phelps, 2008). Entre essas questões, destaca-se a preocupação com a validade e a confiabilidade do ponto de vista psicométrico e pedagógico desses testes. O estudo da validade de uma prova ou de um conjunto de provas é uma questão complexa e merecedora de ampla e constante análise, pois envolve investigar seu objetivo mais importante: sua capacidade de aferir aquilo a que se propõe.

Devido à sua importância, pode-se dizer que a investigação da validade é um dos aspectos mais caros, centrais e delicados no âmbito da psicometria, pois, quando as investigações não sustentam evidências favoráveis à validade das provas, isso muitas vezes indica que as provas não funcionam ou que seu modelo

teórico não se sustenta. Considerando que a análise de validade tem um relevante potencial de fazer ruir propostas e teorias, não fortuitamente um modelo teórico e seu conjunto de provas tornam-se muito mais robustos se experienciam o constante escrutínio da investigação de sua validade.

No caso particular do ENEM, investigar a validade de suas provas envolve justamente verificar empiricamente se e como, de fato, elas aferem os domínios de matemática, ciências da natureza, ciências humanas e linguagens. Temos um conjunto de evidências sobre a confiabilidade e alguns aspectos da validade do ENEM. Sobre a validade externa do ENEM, em termos da capacidade preditiva dos microdados do ENEM, localizamos os estudos de Gomes e Jelihovschi (2020), Gomes, Amantes e Jelihovschi (2020) e Gomes, Lemos e Jelihovschi (2020). Sobre a confiabilidade dos escores do ENEM e sua estimativa com a inserção do fator geral de desempenho, temos os estudos de Gomes et al. (2018, 2020).

Sobre a validade estrutural do ENEM, encontramos estudos que investigaram se os escores das competências estariam vinculados aos seus respectivos domínios (e.g., Gomes et al., 2016). Outros estudos (Costa, 2015; Nojosa, 2002; Vieira, 2016) dedicaram-se a investigação da dimensionalidade do ENEM por meio da Teoria de Resposta ao Item Multidimensional (TRIM). Nojosa (2002) utilizou a TRIM para investigar a estrutura fatorial do ENEM de 1999, comparando modelos com uma a cinco dimensões a partir de três critérios: a razão entre a mudança no qui-quadrado e seus graus de liberdade, inspeção das cargas fatoriais e percentual de variância explicada pelos itens. Nojosa (2002) concluiu favoravelmente pelo modelo com cinco dimensões. Seguindo procedimentos semelhantes ao de Nojosa (2002), Costa (2015), por sua vez, avaliou as edições de 1998 a 2008 e também chegou a um modelo com cinco dimensões. Os resultados indicaram que a maioria dos itens se correlacionava a mais de um fator. Costa concluiu que não eram claras as relações entre os fatores teóricos da prova e as cinco dimensões encontradas em seu estudo. Por fim, Vieira (2016) investigou o modelo atual do ENEM, com os dados de 2012, testando-o como uma prova única de 180 itens por meio da análise fatorial de informação plena baseada na TRIM. Vieira concluiu que um modelo com duas dimensões, raciocínio lógico e interpretação de textos, era o mais bem ajustado aos dados e o mais interpretável. Travitzki (2013) também investigou a dimensionalidade do ENEM, mas por meio de análise dos componentes principais e de análise fatorial exploratória. Seus resultados indicaram que um único fator explicaria cerca de 90% da variância do desempenho discente no ENEM anterior a 2009, e

cerca de 65% no ENEM atual.

No entanto, até o presente momento, desconhecemos qualquer estudo publicado em revistas científicas que tenha tido como objetivo investigar se o modelo de medida do ENEM se sustenta empiricamente, considerando seu restritivo postulado do item-relacionado-a-um-único-fator. Cabe ressaltar que os estudos encontrados sobre a validade fatorial do ENEM dedicaram-se a questão da identificação do número de variáveis latentes subjacentes aos dados por meio de modelos multidimensionais (Costa, 2015; Nojosa, 2002; Vieira, 2016), em uma perspectiva semelhante à análise fatorial exploratória ou a exploração de um modelo bifactor para o Exame (Gomes et al., 2016). Apesar disso, podem ser consideradas evidências desfavoráveis ao modelo de medida do ENEM quanto a sua validade fatorial.

A ausência de estudos que tenham diretamente investigado o postulado do item-relacionado-a-um-único-fator é uma lacuna substancial, com perigosas implicações para a qualidade dos escores gerados pelo INEP. No plano da teoria, o ENEM configura-se como um modelo com quatro dimensões, sendo cada uma delas considerada predominantemente unidimensional (Inep, 2011, 2012). Esse pressuposto teórico sustenta a utilização do modelo logístico de três parâmetros da Teoria de Resposta ao Item (TRI) para o cálculo dos escores nos quatro domínios (Inep, 2011). No entanto, até o presente momento, o Inep não apresentou publicamente à sociedade estudos com evidências acerca das propriedades psicométricas do exame, inclusive que sustentem a validade do postulado previamente apontado.

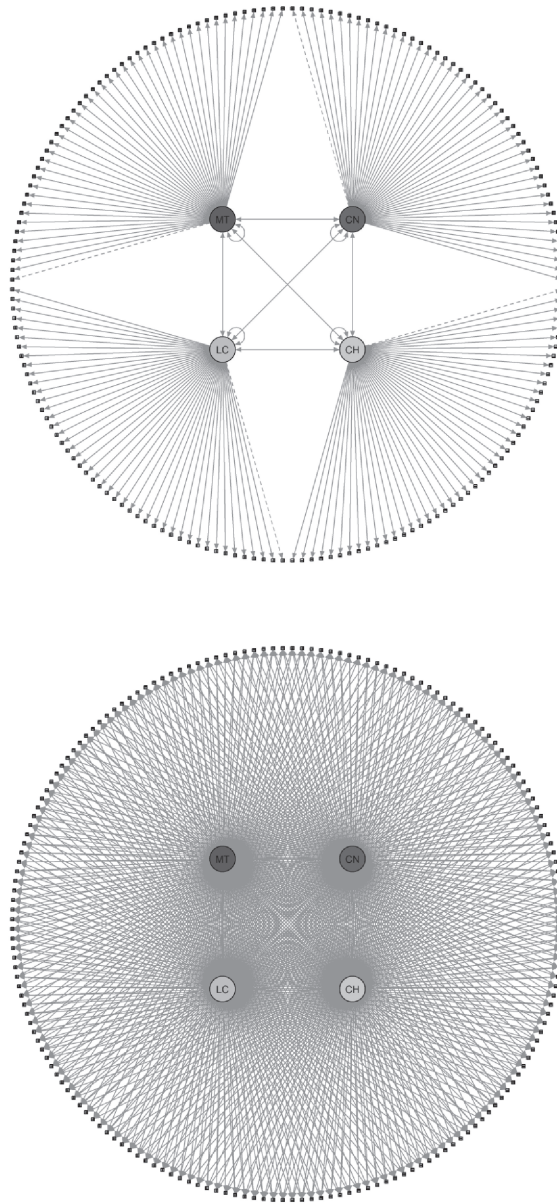
Dois modelos fatoriais permitem testar o postulado do item-relacionado-a-um-único-fator. Um deles é o modelo da estrutura simples, também conhecido como modelo de Thurstone, no qual cada conjunto de 45 itens é explicado direta e exclusivamente pelo domínio-alvo. O modelo da estrutura simples determina que o domínio de matemática, por exemplo, explicaria o desempenho dos estudantes exclusivamente nos 45 itens de matemática da prova do ENEM. Já o domínio de ciências humanas, por sua vez, explicaria apenas o desempenho dos estudantes nos 45 itens de ciências humanas e assim por

diante, ou seja, as cargas fatoriais dos itens em domínios diferentes daqueles que pretendem mensurar têm o valor zero, o que representa ausência de relação. Em outras palavras, esse modelo define a priori que os itens de um determinado domínio não são explicados pelos outros domínios, a não ser pelo próprio domínio em questão. Cabe reforçar que esse é o modelo de medida utilizado para o cálculo dos escores do ENEM (Inep, 2011). O círculo superior na Figura 1 é uma representação gráfica desse modelo.

O segundo modelo é o das cargas fatoriais cruzadas. Da mesma maneira, ele sustenta a presença de quatro variáveis latentes, ou seja, os quatro domínios postulados em teoria, mas determina que o desempenho dos estudantes em todos os itens da prova do ENEM poderia ser explicado parcialmente por cada um dos domínios. Assim, o desempenho nos 45 itens de matemática, por exemplo, poderia ser explicado tanto pela variável latente do domínio de matemática, quanto pela variável latente dos domínios de ciências humanas, de linguagens e de ciências da natureza, ou seja, nenhuma carga fatorial é determinada, a priori, como zero. Esse modelo representaria mais fortemente o caráter interdisciplinar teoricamente concebido para o ENEM desde sua versão inicial (Inep, 2005, 2014, 2018), mas que não é operacionalizado integralmente no modelo de medida utilizado para o cálculo dos escores. O círculo inferior na Figura 1 é uma representação gráfica desse modelo.

Com o objetivo de avaliar a validade de construto do ENEM, esses dois modelos serão testados neste estudo para investigar o postulado do item-relacionado-a-um-único-fator. Essa condição é obrigatória, pois a pertinência das relações teóricas especificadas por um modelo somente é verificada a partir do ajustamento entre as relações determinadas pelo modelo teórico e as relações existentes nos dados analisados. O postulado do item-relacionado-a-um-único-fator do modelo de medida do ENEM encontrará respaldo caso o modelo da estrutura simples apresente ajuste superior ao modelo das cargas fatoriais cruzadas, indicando assim evidências favoráveis de validade fatorial do modelo de medida adotado para o cálculo dos escores do ENEM.

Figura 1
Estruturas dos Modelos de Cargas Fatoriais Simples e de Cargas Fatoriais Cruzadas



Nota. Estrutura do modelo de cargas simples de Thurstone (acima) e estrutura do modelo de cargas cruzadas verificada por meio da modelagem de equações estruturais exploratórias (abaixo).

Fonte. elaboração dos autores

Método

Participantes

Foram analisados os escores de 66.880 participantes do ENEM de 2011, que responderam os cadernos 120, 124, 125 e 129. Desse contingente, a média de idade foi de 21,48 anos (desvio padrão de 7,12 anos); sendo 53,3% do sexo feminino; 50,5% autodeclarados brancos; 10,4% pretos; 33,4% pardos; 2,5% amarelos; 0,5% indígenas e 2,8 que não declararam cor de pele ou etnia.

Instrumento

A prova de 2011 do ENEM é composta por 180 itens, separados em quatro grupos de 45 itens que sustentam a medida para cada um dos quatro domínios: linguagens, matemática, ciências da natureza e ciências humanas. Os itens da prova são retirados de um banco de itens. A prova objetiva de 180 itens do ENEM é realizada em dois encontros de quatro horas. No primeiro encontro, o estudante responde a 90 itens e, no segundo encontro, responde os outros 90 itens. Todos são de múltipla-escolha.

Procedimentos

A análise da estrutura fatorial dos itens da prova de 2011 do ENEM foi realizada por meio de uma integração entre os *softwares* R (R Core Team, 2013) e Mplus (Muthén & Muthén, 2014). A análise dos dados seguiu este procedimento para cada modelo testado no presente estudo: primeiramente, realizou-se o *download*, a extração, a importação e a conversão dos vetores de resposta em escores dicotômicos por meio do pacote ENEM² (Golino, 2014) desenvolvido para o R. Foram selecionados apenas os participantes que estiveram presentes em todas as provas. Os dados faltantes foram transformados em zero para as análises deste estudo. Em seguida, os modelos que seriam testados eram construídos no R, por meio do pacote *MplusAutomation* (Hallquist & Wiley, 2014). Os gráficos foram elaborados por meio do pacote *semPLOT* do R (Epskamp, 2014).

Para a análise do modelo simples de Thurstone, foi feita uma análise fatorial confirmatória, na qual os itens de cada um dos quatro domínios do ENEM eram explicados exclusivamente pela variável latente do domínio-alvo. Nesses termos, a variável latente de matemática explicava exclusivamente o desempenho dos estudantes nos 45 itens de matemática e assim por diante (ver círculo superior da Figura 1). Por sua vez, para a análise do modelo de cargas fatoriais cruzadas (ver círculo inferior da Figura 1), foi feita uma análise por modelagem por equação estrutural exploratória (*exploratory structural equation modeling* – ESEM), com rotação oblíqua Geomin. Detalhes técnicos da ESEM são encontrados em Asparouhov e Muthén (2009) e em Gomes et al. (2017), especificamente quanto ao seu uso nas áreas de psicologia e de educação. Ambas as análises fatoriais foram realizadas utilizando o método de estimação *Weighted Least Squares Means and Variance* (WLSMV), pertinente para análise de escores dicotômicos.

O ajuste aos dados dos dois modelos, o simples de Thurstone e o de cargas cruzadas, foi verificado por meio do índice da aproximação do erro médio-quadrático – *Root Mean-Square Error of Approximation* (RMSEA), e do índice de ajuste comparativo – *Comparative Fit Index* – CFI (Bentler, 1990). O RMSEA é um índice que define a falta de ajuste do modelo definido em relação aos graus de liberdade do modelo (Shumacker & Lomax, 2004). O qui-quadrado do modelo definido, considerado pela estimativa da máxima verossimilhança é subtraído dos graus de liberdade do modelo.

Essa diferença é dividida pelo tamanho da amostra menos 1 e multiplicado pelos graus de liberdade do modelo. A raiz quadrada do valor resultante indica o valor do RMSEA. O CFI, por sua vez, é um indicador que transforma os qui-quadrados de um modelo qualquer

em uma escala de 0 a 1. Este último valor representando o ajuste perfeito. Um modelo de linha de base, de ajuste nulo, é contrastado ao modelo definido e esse contraste indica o índice do CFI. Assim, o CFI é obtido pela diferença entre o qui-quadrado da máxima verossimilhança com os graus de liberdade do modelo definido. Essa diferença é dividida pela diferença entre o qui-quadrado da máxima verossimilhança do modelo nulo.

Esse valor é então subtraído a 1, indicando o valor do CFI do modelo analisado (Shumacker & Lomax, 2004). Um bom ajuste aos dados é indicado se o RMSEA é igual ou inferior a 0,06 (Browne & Cudeck, 1993) e o CFI igual ou superior a 0,95 (Hu & Bentler, 1999).

Resultados

A Tabela 1 apresenta o grau de dificuldade dos itens em termos de cinco agrupamentos de percentuais de acerto dos estudantes: 0 a 20% de acerto, 21 a 40%, 41 a 60%, 61 a 80% e 81 a 100%. Arbitrariamente esses grupos determinarão a classificação dos itens em muito difíceis, difíceis, medianos, fáceis e muito fáceis, respectivamente. Pelas frequências absolutas apontadas na Tabela 1, é possível constatar que os domínios de matemática (MT) e ciências da natureza (CN) apresentaram um grau de dificuldade relativamente semelhante, assim como ocorreu entre os domínios de ciências humanas (CH) e linguagens (LC).

Houve uma dificuldade maior no domínio de CN, com uma concentração de nove itens no agrupamento dos itens muito difíceis. Os itens difíceis foram os mais frequentes em MT e CN, enquanto os medianos foram os mais frequentes em LC e CH. Houve um número muito baixo de itens muito fáceis e muito difíceis em LC e CH. Por sua vez, houve um número muito baixo de itens muito fáceis em CN e MT. No global dos 180 itens, teve um número muito baixo de itens muito fáceis e uma predominância de itens difíceis e medianos.

Ambos os modelos analisados apresentaram adequado grau de ajuste aos dados. O modelo de cargas simples de Thurstone, por meio da análise fatorial confirmatória, apresentou um RMSEA adequado e um CFI limítrofe ($\chi^2[15924]=288691,48$; CFI=0,94; RMSEA=0,016). Por sua vez, o modelo das cargas fatoriais cruzadas via ESEM apresentou um bom ajuste tanto pelo CFI quanto pelo RMSEA, indicando um bom ajuste do modelo ($\chi^2[15396]=88183,565$; CFI=0,984; RMSEA=0,008).

No que diz respeito às cargas dos itens no modelo de cargas simples (análise fatorial confirmatória), a variável latente de ciências da natureza apresentou betas variando entre -0,07 e 0,66 ($M=0,36$; $DP=0,18$), enquanto a variável latente de ciências humanas apresentou betas

² Interessados em utilizar o pacote podem entrar em contato por e-mail: hfg9s@eservices.virginia.edu

variando entre 0,06 e 0,74 ($M=0,44$; $DP=0,15$). Já a variável latente de linguagens apresentou betas entre 0,14 e 0,77 ($M=0,48$; $DP=0,15$), enquanto a variável latente de

matemática apresentou betas entre 0,00 e 0,68 ($M=0,46$; $DP=0,16$). As correlações entre as variáveis latentes variaram entre 0,75 e 0,92 ($M=0,85$; $DP=0,07$).

Tabela 1

Frequência Absoluta de Acertos aos Itens, Distribuída em Cinco Agrupamentos

Domínios	Agrupamentos de percentuais de acerto dos estudantes				
	0-20%	21-40%	41-60%	61-80%	80-100%
Matemática	5	22	14	4	0
Ciências da Natureza	9	22	11	2	1
Ciências Humanas	1	14	18	10	2
Linguagens	1	10	19	14	1
Total	16	68	62	30	4

Fonte: elaboração dos autores

A Figura 2 apresenta o padrão dos betas de cada uma das variáveis latentes do modelo de quatro fatores de cargas simples (análise fatorial confirmatória). Em cada círculo da figura, restringiu-se o valor mínimo dos betas que aparecem no gráfico. O círculo superior à esquerda apresenta os betas superiores à 0,80, enquanto o superior à direita apresenta betas superiores a 0,70. O círculo central à esquerda apresenta betas superiores à 0,60 e o central à direita betas superiores à 0,50. A Figura 2 também apresenta o modelo de quatro fatores de cargas simples com valores superiores a 0,30, representado pelo círculo inferior ao centro. Por fim, a Figura 3 apresenta a média e o intervalo de 95% de confiança das cargas padronizadas dos itens nos fatores do ENEM. Em suma, as cargas padronizadas apresentaram as seguintes médias nesse modelo: ciências da natureza ($M=0,357$; $DP=0,177$), ciências humanas ($M=0,443$; $DP=0,151$), linguagens ($M=0,479$; $DP=0,147$) e matemática ($M=0,460$; $DP=0,165$).

No que diz respeito ao modelo de cargas cruzadas, via ESEM, os 45 itens de quase todos os domínios apresentaram uma carga fatorial média maior justamente na variável latente correspondente ao domínio-alvo (Tabela 2). Os 45 itens de LC tiveram uma carga fatorial média de 0,41 justamente em relação à variável latente LC; já os 45 itens de CH apresentaram uma carga fatorial média de 0,248 em relação à variável latente de CH. Os 45 itens de CN apresentaram uma carga fatorial média de 0,142 em relação à variável latente de CN e de 0,189 na variável latente de MT. Por último, os 45 itens de MT apresentaram uma carga fatorial média de 0,334 em relação à variável latente de MT. A concentração média das

cargas dos itens nos fatores pode ser visualizada na Figura 4 e na Tabela 2.

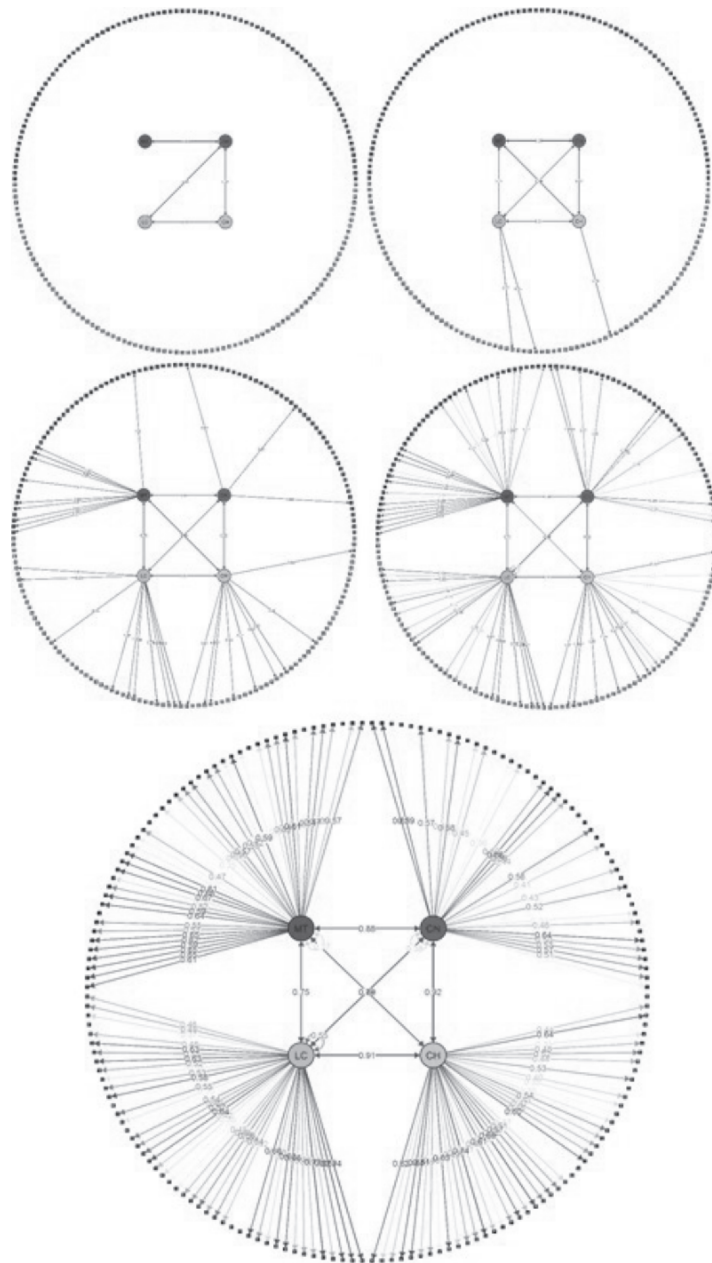
A Figura 5 apresenta o padrão de cargas fatoriais de cada uma das variáveis latentes do modelo de quatro fatores de cargas cruzadas (ESEM). Em cada círculo da figura, restringiu-se o valor mínimo dos betas que aparecem no gráfico. O círculo superior à esquerda apresenta as cargas superiores a 0,80, enquanto o superior à direita apresenta as cargas superiores a 0,70. O círculo central à esquerda apresenta as cargas superiores a 0,60 e o círculo central à direita cargas superiores a 0,50. A Figura 5 também apresenta o modelo de quatro fatores de cargas cruzadas com valores superiores a 0,30, representado pelo círculo inferior ao centro.

A Tabela 2 mostra a frequência absoluta de cargas fatoriais das variáveis latentes com valor igual ou superior a 0,30 nos quatro agrupamentos de itens. No agrupamento de itens de CN, sete itens apresentaram carga fatorial igual ou superior a 0,30 em CH, oito em CN e dois em MT. No agrupamento de itens de CH, 10 apresentaram carga fatorial igual ou superior a 0,30 em LC, 13 em CH e dois em CN. O agrupamento de itens de LC apresentou carga fatorial igual ou superior a 0,30 em 36 de seus itens na variável latente de LC e um item na variável latente de CN. Já o agrupamento de itens de MT apresentou 25 itens com carga fatorial igual ou superior a 0,30 na variável latente de MT e 11 itens junto à variável latente de CN.

Os domínios se correlacionam, e a Tabela 2 apresenta os valores dessas correlações. As correlações entre os domínios são moderadas, conforme o critério proposto por Cohen (1988), excetuando-se as correlações entre CN e CH, e entre CN e LC, que são correlações fracas.

Figura 2

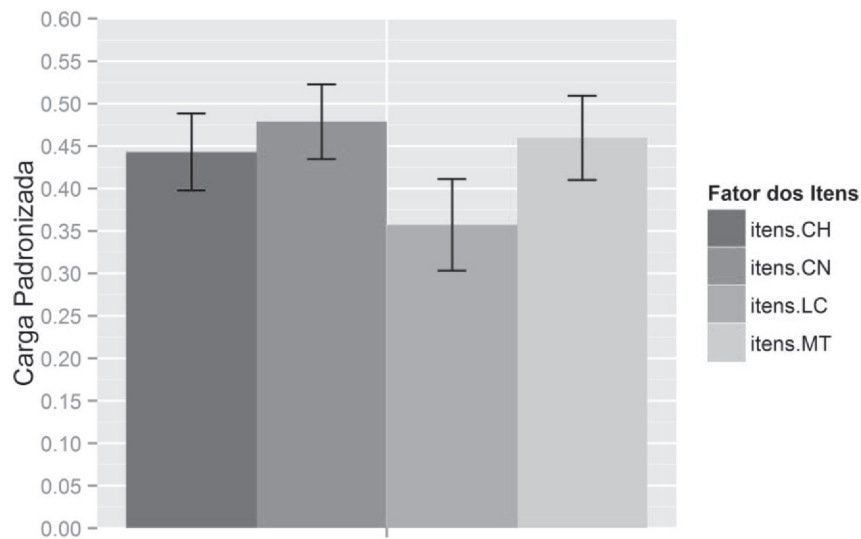
Padrão de cargas de correlação entre as variáveis latentes no modelo de cargas simples



Nota. Padrão de cargas de correlação entre as variáveis latentes no modelo de cargas simples (análise fatorial confirmatória), e de regressão das variáveis latentes para os itens, com valores mínimos de 0,80 (superior esquerdo), 0,70 (superior direito), 0,60 (centro esquerdo), 0,50 (centro direito) e 0,30 (inferior).

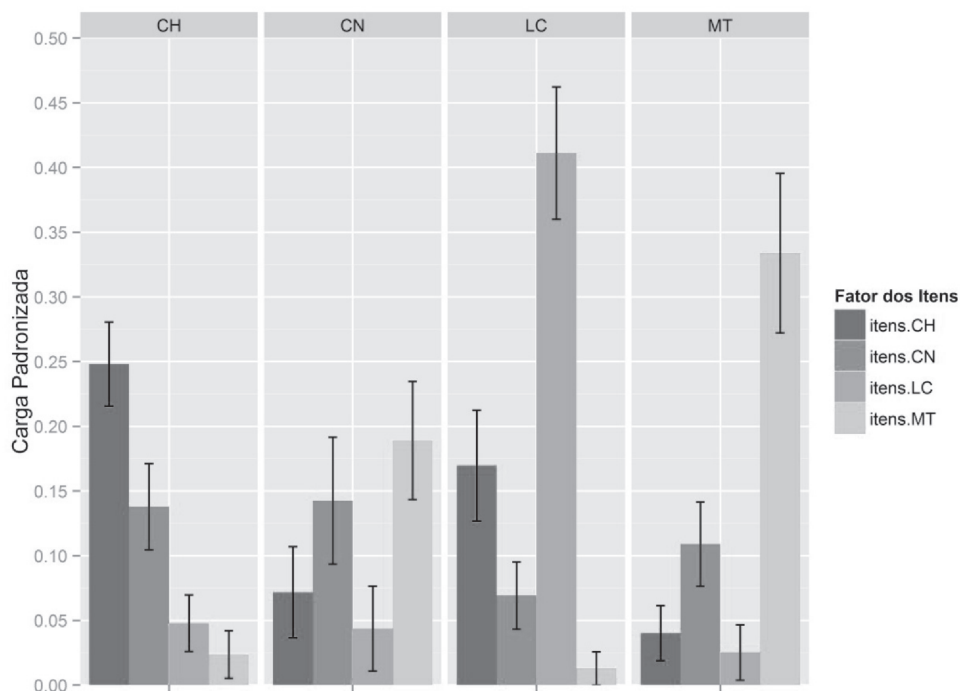
Fonte. elaboração dos autores

Figura 3
Média e intervalo de 95% de confiança das cargas fatoriais por fatores do ENEM



Legenda. CN (ciências da natureza); CH (ciências humanas); LC (linguagens); e MT (matemática).
Fonte. elaborado pelos autores

Figura 4
Média e intervalo de 95% de confiança das cargas fatoriais por fatores do ENEM no modelo de cargas cruzadas (ESEM)



Legenda. CN (ciências da natureza); CH (ciências humanas); LC (linguagens); e MT (matemática).
Fonte. elaborado pelos autores

Tabela 2
Correlações entre as Variáveis Latentes (Domínios) e Cargas Fatoriais do Modelo de Cargas Cruzadas

Estatística	Domínios	CH		CN		LC		MT	
		M	DP	M	DP	M	DP	M	DP
Correlação entre as variáveis latentes	CH	1,000							
	CN	0,352		1,000					
	LC	0,665		0,281		1,000			
	MT	0,554		0,375		0,647		1,000	
Frequência absoluta das cargas fatoriais superiores a 0,30	CH	13		2		10		0	
	CN	7		8		0		2	
	LC	0		1		36		0	
	MT	0		11		0		25	
Média e desvio padrão das cargas fatoriais	CH	0,248	0,107	0,072	0,117	0,170	0,141	0,040	0,070
	CN	0,138	0,110	0,142	0,161	0,069	0,086	0,109	0,108
	LC	0,048	0,072	0,044	0,109	0,411	0,170	0,025	0,070
	MT	0,024	0,060	0,189	0,150	0,012	0,043	0,334	0,205

Legenda. CN (ciências da natureza); CH (ciências humanas); LC (linguagens); MT (matemática); M (média); e DP (desvio padrão).
Fonte. elaborado pelos autores

Discussão

Até a elaboração de modelos mais abrangentes, como a modelagem de equação estrutural exploratória, a análise fatorial confirmatória se limitava ao modelo simples de Thurstone. No entanto, este é demasiadamente restritivo e pode levar a ajustes pobres em função da restrição das cargas fatoriais. Esse último aspecto não delimita o cerne da validade da presença das variáveis latentes, de modo que o desacoplamento da restrição das cargas fatoriais na análise das variáveis tem se mostrado um avanço considerável nos últimos anos na psicometria. O modelo de medida do ENEM, utilizado para o cálculo dos escores, é um modelo simples, baseado no postulado do item-relacionado-a-um-único-fator. Neste estudo, avaliamos as evidências de validade estrutural desse modelo por meio da comparação de seu ajuste aos dados ao ajuste de um modelo com cargas fatoriais cruzadas.

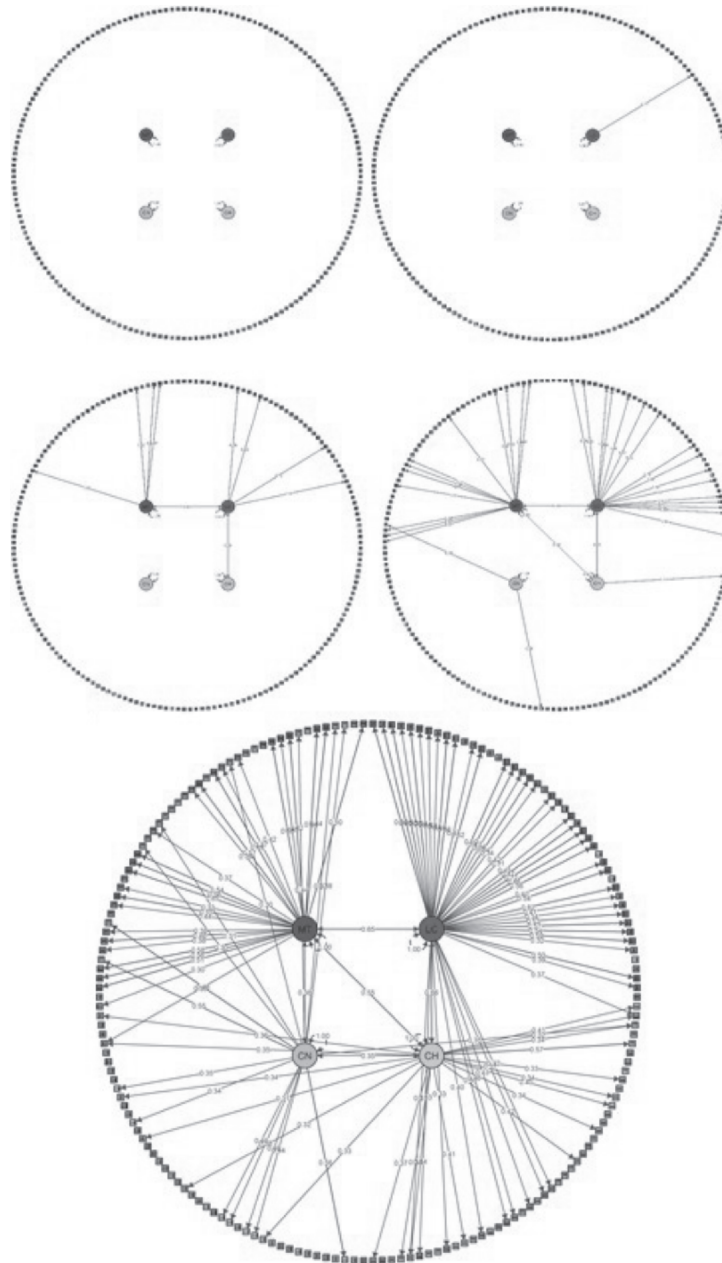
Os resultados deste estudo apontaram que ambos os modelos analisados sustentam empiricamente a presença dos quatro domínios do ENEM. O satisfatório ajuste do modelo de cargas simples de Thurstone, verificado por meio de análise fatorial confirmatória, e o bom ajuste do modelo de cargas fatoriais cruzadas, via modelagem de equação estrutural exploratória, trazem evidências favoráveis para a validade dos quatro domínios postulados pelo modelo teórico do ENEM: matemática, linguagens, ciências da natureza e ciências humanas. Os resultados deste estudo sustentam, portanto, a presença empírica de quatro variáveis latentes que representam esses domínios.

O modelo de cargas fatoriais cruzadas, no entanto, apresentou um ajuste mais robusto aos dados do que o modelo de cargas simples, atendendo aos pontos de corte preconizados na literatura para os índices de ajuste. Esse resultado indica claramente que certos domínios têm participação no desempenho de itens relacionados a outros. Quanto aos betas, a média das cargas fatoriais padronizadas diminui quando se permite que um mesmo item carregue em todos os fatores do modelo. Com essa permissão, implementada via modelagem de equação estrutural exploratória, diferentes fatores irão competir pela explicação da variância de um mesmo item, o que pode ocasionar uma diminuição das cargas fatoriais, quando comparadas com modelos que não permitem o cruzamento das cargas.

O fato de o modelo de cargas cruzadas mostrar-se mais robusto em seu ajuste aos dados possui implicações psicométricas para o ENEM, tanto para o modelo teórico que está expresso em sua matriz de referência e que sustenta a elaboração dos itens, quanto para o modelo de medida que operacionaliza o cálculo das notas das provas. Por um lado, esse resultado é uma evidência desfavorável à validade fatorial do modelo de medida adotado para o cálculo dos escores, que parte do postulado restritivo da pureza das cargas fatoriais, ou seja, de que cada item do exame é relacionado a um único fator. Consequentemente, esse resultado coloca em questão também a qualidade dos escores calculados. Do ponto de vista da validade consequencial, esse resultado aponta a necessidade de se investigar se a forma atualmente

Figura 5

Padrão de cargas de correlação entre as variáveis latentes no modelo de cargas simples



Nota. Padrão de cargas de correlação entre as variáveis latentes no modelo de cargas simples (análise fatorial confirmatória), e de regressão das variáveis latentes para os itens, com valores mínimos de 0,80 (superior esquerdo), 0,70 (superior direito), 0,60 (centro esquerdo), 0,50 (centro direito) e 0,30 (inferior).

Fonte. elaboração dos autores

adotada para o cálculo dos escores representa algum tipo de prejuízo não intencional para a interpretação ou os usos que se faz dos resultados do ENEM, como na seleção de candidatos ao ingresso no Ensino Superior. Por outro lado, esse resultado é uma evidência favorável ao pressuposto da natureza interdisciplinar das habilidades e competências avaliadas pelo ENEM, conforme planejado na concepção teórica do exame.

Os resultados deste estudo também apontam para outros desafios enfrentados pelo ENEM do ponto de vista psicométrico. O número de itens muito fáceis e muito difíceis precisa ser melhor balanceado em relação às outras faixas de dificuldade. Os itens muito fáceis foram poucos em todos os domínios e precisam ser ampliados. Por sua vez, os domínios de ciências humanas e linguagens foram mais carentes em itens muito difíceis, em comparação com os domínios de matemática e ciências da natureza.

Além de um relativo desequilíbrio no grau de dificuldade dos itens, os agrupamentos de itens apresentaram uma carga fatorial média fraca ou moderada em seus domínios-alvo (Cohen, 1988). Os agrupamentos de itens de ciências da natureza, ciências humanas e matemática apresentaram cargas fatoriais médias fracas em relação a seus respectivos domínios-alvo no modelo de cargas cruzadas, e cargas fatoriais médias moderadas no modelo de cargas simples. A única exceção foi o agrupamento de itens em linguagens, apresentando uma carga fatorial média moderada no modelo de cargas cruzadas. De maneira geral, a quantidade de itens com cargas fatoriais padronizadas elevadas é muito pequena, como pode ser visto nas Figuras 3 e 5. É possível que essas cargas fatoriais fracas a moderadas envolvam o próprio desafio do ENEM de elaborar itens mais amplos, interdisciplinares e focados na resolução de problemas, de uma forma abrangente, ao invés de focar em conteúdos ou habilidades bem específicos. Essa é mais uma questão a ser investigada.

O presente estudo apresenta limitações em suas evidências, na medida em que investigou a estrutura fatorial exclusivamente da prova de 2011. Outras provas podem ter características distintas da prova analisada, mesmo que os cadernos de provas do ENEM sejam montados a partir de um mesmo banco de itens previamente calibrados (Inep, 2011).

No entanto, as evidências obtidas neste estudo servem de ponto de partida para análises posteriores e comparações com os dados das provas aplicadas em outros anos. O fato de utilizar uma amostra dos estudantes que fizeram o ENEM de 2011, selecionando os cadernos 120, 124, 125, e 129, assim como a língua inglesa, não é um empecilho. Excetuando os itens da língua inglesa, todos os outros itens são iguais ao restante dos estudantes que fizeram a prova de 2011. Apesar da escolha desses

cadernos ter sido arbitrária, ela justifica-se pelo método randômico adotado para a distribuição dos cadernos aos participantes. Além disso, o tamanho da amostra selecionada foi mais que suficiente para uma estimativa precisa dos dados.

Concluindo, os domínios do ENEM parecem ser válidos para explicar o desempenho dos estudantes na prova de 2011. No entanto, este estudo apresenta evidências favoráveis ao pressuposto teórico da interdisciplinaridade do ENEM, que orienta a elaboração de itens para o exame, enquanto coloca em questão o modelo de medida adotado para o cálculo dos escores, baseado no pressuposto de que os itens de cada domínio avaliado são explicados exclusivamente por um único domínio. Além disso, identificou-se alguns desafios relacionados à qualidade dos itens para a construção de uma prova mais equilibrada em termos de graus de dificuldade e com itens mais intensamente relacionados aos domínios-alvo.

Agradecimentos

Não há menções.

Financiamento

Todas as fontes de financiamento para elaboração e produção do estudo (coleta, análise e interpretação dos dados, bem como, escrita dos resultados no presente no manuscrito) foram fornecidas pelo projeto de pesquisa 'Edital Chamada Pública Inep/Dired nº 05/2012 do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)'.

Contribuições dos autores

Declaramos que todos os autores participaram da elaboração do manuscrito. Especificamente, o(s) autor(es) Cristiano Mauro Assis Gomes e Hudson Fernandes Golino participaram da redação inicial do estudo – conceitualização, investigação, visualização, o(s) autor(es) Cristiano Mauro Assis Gomes e Hudson Fernandes Golino participaram da análise dos dados, e o(s) autor(es) Cristiano Mauro Assis Gomes, Hudson Fernandes Golino e Alexandre José de Souza Peres participaram da redação final do trabalho – revisão e edição.

Disponibilidade dos dados e materiais

Todos os dados e sintaxes gerados e analisados durante esta pesquisa serão tratados com total sigilo devido às exigências do Comitê de Ética em Pesquisa em Seres Humanos. Porém, o conjunto de dados e sintaxes que apoiam as conclusões deste artigo estão disponíveis mediante razoável solicitação ao autor principal do estudo.

Conflito de interesses

Os autores declaram que não há conflitos de interesses.

Referências

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438. <https://doi.org/10.1080/10705510903008204>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Brasil, Ministério da Educação (MEC). *Portaria MEC nº 438*, de 28 de maio de 1998. Institui o Exame Nacional do Ensino Médio – ENEM. http://www.crmariocovas.sp.gov.br/pdf/diretrizes_p0178-0181_c.pdf
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long. (Orgs.), *Testing structural equation models* (pp. 136-162). Sage.
- Cohen, J. (1998). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Costa, C. E. S. (2015). *Análise da dimensionalidade e modelagem multidimensional pela TRI no ENEM (1998-2008)*. [Dissertação de mestrado, Universidade Federal de Santa Catarina], Universidade Federal de Santa Catarina. <https://repositorio.ufsc.br/xmlui/handle/123456789/169527>
- Epskamp, S. (2014). *semPlot: Path diagrams and visual analysis of various SEM packages' output* (R package version 1.0.1) [Software]. <http://CRAN.R-project.org/package=semPlot>
- Golino, H. F. (2014). *ENEM: an implementation of functions to help automatic downloading, importing, cleaning and scoring of the Brazilian's National High School Exam (ENEM)* [Software]. <https://github.com/hfgolino/ENEM>
- Gomes, C. M. A. (2010). Avaliando a avaliação escolar: notas escolares e inteligência fluida. *Psicologia em Estudo*, 15(4), 841-849. <https://www.scielo.br/j/pe/a/YpmzFYHdYkR35YSdjPy89zr/?lang=pt#>
- Gomes, C. M. A., & Borges, O. (2009). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, 20(42), 73-88. <https://doi.org/10.18222/eaec204220092060>
- Gomes, C. M. A., & Jelihovschi, E. (2020). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education*, 43(2), 201-221. <https://doi.org/10.1080/1743727X.2019.1654992>
- Gomes, C. M. A., Almeida, L. S., & Núñez, J. C. (2017). Rationale and applicability of Exploratory Structural Equation Modeling (ESEM) in psychoeducational contexts. *Psicothema*, 29(3), 396-401. <https://doi.org/10.7334/psicothema2016.369>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2016). Investigando a validade estrutural das competências do ENEM: quatro domínios correlacionados ou um modelo bifatorial? *Boletim na Medida (INEP-Ministério da Educação)*, 5(10), 33-38. <http://portal.inep.gov.br/documents/186968/494037/BOLETIM+NA+MEDIDA+-+N%C2%BA+10/4b8e3d73-d95d-4815-866c-ac2298dff0bd?version=1.1>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2018). Análise da fidedignidade composta dos escores do Enem por meio da análise fatorial de itens. *European Journal of Education Studies*, 5(8), 331-344. <https://doi.org/10.5281/zenodo.2527904>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2020). Fidedignidade dos escores do Exame Nacional do Ensino Médio (ENEM). *Psico*, 51(2), e31145. <https://doi.org/10.15448/1980-8623.2020.2.31145>
- Gomes, C. M. A., Amantes, A. & Jelihovschi, E. G. (2020). Applying the regression tree method to predict students' science achievement. *Trends in Psychology*, 28, 99-117. <https://doi.org/10.9788/s43076-019-00002-5>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020). Comparing the predictive power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1), 87-96. <http://dx.doi.org/10.15689/ap.2020.1901.17737.10>
- Hallquist, M., & Wiley, J. (2014). *MplusAutomation: automating Mplus model estimation and interpretation* (R package version 0.6-3) [Software]. <http://CRAN.R-project.org/package=MplusAutomation>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) (2012). *Microdados do ENEM – 2011* [Data file]. <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2005). *Exame Nacional do Ensino Médio (ENEM): fundamentação teórico-metodológica*. <http://portal.inep.gov.br/documents/186968/484421/ENEM+-+Exame+Nacional+do+Ensino+Médio+funda+menta%ç%o+te%o%r+ico-metodol%o%g+ica/449eea9e-d904-4a99-9f98-da804f3c91f5?version=1.1>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2011). *Nota Técnica sobre o uso da Teoria de Resposta ao Item no ENEM*. http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2014). *Exame Nacional do Ensino Médio (ENEM): relatório pedagógico 2009-2010*. https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/relatorio_pedagogico_enem_2009_2010.pdf
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2018). *Exame Nacional do Ensino Médio - ENEM: Escalas de proficiência 1998/2018*. https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/enem_escala_de_proficiencia_1998_2008.pdf
- Muthén, L. K., & Muthén, B.O. (2014). *Mplus (Version 7)* [Software]. Los Angeles, Estados Unidos: Muthén & Muthén. <https://www.statmodel.com/>
- Nojosa, R. T. (2002). Teoria da Resposta ao Item (TRI): modelos multidimensionais. *Estudos em Avaliação Educacional*, 25, 123-166. <https://doi.org/10.18222/eaec02520022193>
- Phelps, R. P. (Eds.). (2008). *Correcting fallacies about educational and psychological testing*. American Psychological Association.
- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(2), 151-159. <https://doi.org/10.1590/S0102-37722001000200007>

- R Development Core Team (2013). *R: A language and environment for statistical computing* (R version 3.0.0) [Software]. The R Project for Statistical Computing. Vienna, Austria. <http://www.r-project.org>
- Shumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Lawrence Erlbaum Associates.
- Travitzki, R. (2013). *ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar*. [Tese de Doutorado, Universidade de São Paulo], Universidade de São Paulo. <https://www.teses.usp.br/teses/disponiveis/48/48134/tde-28062013-162014/pt-br.php>
- Vieira, N. N. (2016). *As provas das quatro áreas do ENEM vistas como prova única na ótica de modelos da Teoria da Resposta ao Item Uni e Multidimensional* [Dissertação de mestrado, Universidade Federal de Santa Catarina], Universidade Federal de Santa Catarina. <https://repositorio.ufsc.br/xmlui/handle/123456789/169668>

recebido em maio de 2018
aprovado em fevereiro de 2021

Sobre os autores

Cristiano Mauro Assis Gomes é doutor em Educação pela Universidade Federal de Minas Gerais (UFMG) e professor do Departamento de Psicologia, do Programa de Pós-Graduação em Psicologia, Cognição e Comportamento e do Programa de Pós-Graduação em Neurociências da UFMG. Bolsista de produtividade do CNPq, nível 2.

Hudson Fernandes Golino é doutor em Neurociências pela Universidade Federal de Minas Gerais (UFMG) e professor do Departamento de Psicologia da University of Virginia.

Alexandre José de Souza Peres é doutor em Psicologia Social, do Trabalho e das Organizações pela Universidade de Brasília (UnB) e professor do Programa de Pós-Graduação em Psicologia e do Câmpus de Paranaíba da Universidade Federal de Mato Grosso do Sul (UFMS).

Como citar este artigo

Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2021). Evidências desfavoráveis ao postulado de cargas fatoriais simples do Exame Nacional do Ensino Médio (ENEM). *Avaliação Psicológica*, 20(3), 308-320. <http://dx.doi.org/10.15689/ap.2021.2003.15777.05>