

**CARLA FERNANDA FERREIRA-RODRIGUES**



**ESTUDOS COM O ENEM A PARTIR DE UMA ABORDAGEM  
PSICOMÉTRICA DA INTELIGÊNCIA**

**Apoio:**



**ITATIBA  
2015**

**CARLA FERNANDA FERREIRA-RODRIGUES**

**ESTUDOS COM O ENEM A PARTIR DE UMA ABORDAGEM  
PSICOMÉTRICA DA INTELIGÊNCIA**

Tese apresentada ao Programa de Pós-Graduação  
*Stricto Sensu* em Psicologia da Universidade São  
Francisco, Área de Concentração - Avaliação  
Psicológica, para obtenção do título de Doutora.

ORIENTADOR: PROF. DR. RICARDO PRIMI

ITATIBA  
2015

P 157.931      Ferreira-Rodrigues, Carla Fernanda.  
F44e              Estudos com o Enem a partir de uma abordagem  
psicométrica da inteligência / Carla Fernanda Ferreira-  
Rodrigues. -- Itatiba, 2015.  
95 p.


Tese (Doutorado) – Programa de Pós-Graduação  
*Stricto Sensu* em Psicologia da Universidade São  
Francisco.  
Orientação de: Ricardo Primi.

1. Avaliação educacional. 2. Rendimento escolar.  
3. Vestibular. 4. Enade. I. Primi, Ricardo. II. Título.




**UNIVERSIDADE SÃO FRANCISCO**  
**PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU**  
**EM PSICOLOGIA**

Carla Fernanda Ferreira Rodrigues defendeu a tese “ESTUDOS COM O ENEM A PARTIR DE UMA ABORDAGEM PSICOMÉTRICA DA INTELIGÊNCIA” aprovada pelo Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco em 27 de agosto de 2015 pela Banca Examinadora constituída por:

  
\_\_\_\_\_  
Prof. Dr. Ricardo Primi  
Presidente

  
\_\_\_\_\_  
Prof. Dra. Elaine Toldo Pazello  
Examinadora

  
\_\_\_\_\_  
Prof. Dr. Daniel Bartholomeu  
Examinador

  
\_\_\_\_\_  
Prof. Dr. Nelson Hauck Filho  
Examinador

  
\_\_\_\_\_  
Prof. Dra. Claudette Maria Medeiros Vendramini  
Examinadora

*“Quando eu vim do sertão, seu moço, do meu Bodocó...só  
trazia coragem e a cara...eu penei, mas aqui cheguei...”*

(Pau de Arara - Luiz Gonzaga)

*Minha vida é andar por este país  
Pra ver se um dia descanso feliz  
Guardando as recordações  
Das terras onde passei  
Andando pelos sertões  
E dos amigos que lá deixei*

*Chuva e sol  
Poeira e carvão  
Longe de casa  
Sigo o roteiro  
Mais uma estação  
E a alegria no coração*

*Minha vida é andar por esse país  
Pra ver se um dia descanso feliz  
Guardando as recordações  
Das terras onde passei  
Andando pelos sertões  
E dos amigos que lá deixei*

*Mar e terra  
Inverno e verão  
Mostre o sorriso  
Mostre a alegria  
E a saudade no coração*

(A vida do viajante - Luiz Gonzaga)

Dedico esse trabalho à minha mãe que sempre colocou a minha formação educacional como prioridade.

## Agradecimentos

Ao final desse trabalho é impossível não relembrar tudo e todos que passaram no meu caminho desde que saí de Petrolina-PE para cursar a faculdade em João Pessoa-PB, depois o mestrado em Natal-RN e, mais recentemente, o Doutorado em Itatiba-SP, na Universidade São Francisco que sempre sonhei estudar desde quando comecei a me interessar pela Avaliação Psicológica.

Inicialmente, agradeço aos meus pais, *Maria Ferreira Rodrigues e Antonio Hosterno Rodrigues* que nunca mediram esforços para investir na minha educação e apoiaram a minha decisão de fazer mestrado e depois doutorado em universidades diferentes da que fiz a graduação e, assim, ter uma formação voltada à área de Avaliação Psicológica. À minha irmã *Ilka Juliana* pela amizade, amor e companheirismo, por sempre estar disposta a ler rapidamente qualquer trabalho meu e me indicar o que não está compreensível. Sem vocês eu não teria chegado até aqui!

Foi na Universidade Federal da Paraíba (UFPB), onde cursei a faculdade, que dei os primeiros passos e fiz as primeiras reflexões sobre o meu futuro profissional. Foi muito bom ter cursado a disciplina de Psicometria com o Prof. Dr. Valdiney Veloso Gouveia e, assim, conhecer um lado da Psicologia que eu não sabia bem como era (a medida!), mas que eu buscava encontrar no curso. Foi nessa disciplina que iniciei o meu interesse formal pela avaliação psicológica.

Na UFPB, agradeço muito ao Prof. Dr. Francisco José Batista de Albuquerque por ter me mostrado a magia da “universidade submersa”, ou seja, dos núcleos de pesquisa. Se segui na área acadêmica é por ele ter me mostrado a importância e como se fazer pesquisa, seja em ambiente urbano, rural ou agrário.

Na Universidade Federal do Rio Grande do Norte (UFRN), fui orientada no mestrado pelo Prof. Dr. João Alchieri e com ele aprendi a não só valorizar as análises estatísticas, mas também o lado qualitativo dos fenômenos psicológicos.

Apesar de toda importância da UFPB e da UFRN na minha formação, é na USF que me sinto em casa e nessa etapa do doutorado tenho muito o que agradecer:

Agradeço ao meu orientador, Prof. Dr. *Ricardo Primi*, que é, sem dúvida, a pessoa mais inteligente que conheço, exemplo de pesquisador, sempre atualizadíssimo e com mil ideias para o projeto, sempre com uma dica para passar, livros, textos ou *sites* para mostrar, que demonstra por meio do seu comportamento que sempre temos algo para aprender, que é preciso estudar, se atualizar e estar aberto a coisas novas. Por respeitar o meu tempo de escrita, afinal não é fácil escrever algo e saber que o Ricardo Primi vai ler, rs. Ricardo, você é o meu maior modelo de pesquisador, espero que nossa parceria não termine com o doutorado pois tenho muito a aprender contigo.

Aos Professores Doutores do Programa de Pós-Graduação em Psicologia da USF, especialmente aos que cursei disciplinas: *Acácia Aparecida Angeli dos Santos, Ana Paula Porto Noronha, Anna Elisa de Villemor-Amaral, Claudette Maria Medeiros Vendramini e Ricardo Primi* agradeço por todos os ensinamentos. Até hoje fico meio sem acreditar que sou aluna de vocês, pois sempre os admirei lendo seus artigos, capítulos de livros, testes psicológicos ou assistindo suas apresentações em congressos. Quando me tornei aluna, passei a admirá-los ainda mais. Tenho muito orgulho de tê-los como professores e referências para minha vida profissional.

Gostaria de agradecer ao *Lucas de Francisco Carvalho* que, no início do doutorado, era meu colega de laboratório (LabAPE) e depois tornou-se professor do programa, pela confiança em me convidar para desenvolver trabalhos com ele, pelo convite para fazer parte do apoio editorial da revista Psico-USF e para ministrar minhas primeiras aulas, na Pós-Graduação *Lato Sensu* da USF.

Agradeço aos professores da minha banca de qualificação pela leitura cuidadosa e contribuições ao trabalho: *Dr. Daniel Bartholomeu, Dra. Marjorie Cristina Rocha da Silva, Dra. Acácia Aparecida Angeli dos Santos e Dr. Cristian Zanon*.

Muito obrigada também aos professores da banca de Arguição Final pela disponibilidade e pelas contribuições que deram ao trabalho: *Dra. Elaine Toldo Pezolla, Dr. Daniel Bartholomeu, Dra. Claudette Maria Medeiros Vendramini e Dr. Nelson Hauck Filho*.



Aos meus colegas de turma de Doutorado Ana Cristina, Eni, Fernanda Lopes, Fernanda Otatti, Lariana, Marco Antônio, Rodolfo Ambiel e Thatiana pelas contribuições ao meu trabalho e convívio nas disciplinas Seminários I e II e Produção de Textos I e II.

Aos meus colegas que passaram pelo Laboratório de Avaliação Psicológica e Educacional (LabAPE) nesse período: Marjorie, Daniel, Sanyo, Lucas, Priscila, Fernanda Lopes, Fernando Pessotto, Alexandre, Everson, Marie, André, Anderson e Jonatha.

As amigas que conquistei no doutorado e que pude dividir momentos de alegria, tensão, e sempre ter ótimas conversas sobre a vida, o meio acadêmico e o futuro: *Raquel Tavella, Juliana Bueno, Jocemara Mognon, Roberta Ramazotti, Karen Lamas, Thatiana Lima, Fernanda Lopes, Pamella Pavan, Lucila Cardoso e Giselle Pianowski.*

A Carla Leite que fez da sua casa, minha casa quando cheguei em Itatiba para fazer a seleção do doutorado e nos 2 anos e meio seguintes quando dividimos apartamento, ou melhor, administramos o “museu do Fabiano”.

Agradeço também a *Roberta Ramazotti Ferraz de Campos* e a *Jocemara Ferreira Mognon* por me convidarem para morar com elas nesse último ano. Rô e Jô, obrigada pela acolhida, pelo convívio, amizade, companheirismo, pela troca de ideias e por mudarem minha rotina de sono, me mostrando que sou capaz de dormir antes da meia noite e acordar no máximo às 07:00 h da madrugada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Aos meus familiares e amigos que torceram, me ligaram, mandaram mensagens, queriam saber como eu estava e que iam rapidamente me ver quando eu ia a Petrolina. Obrigada pelo incentivo, carinho e orgulho!

## RESUMO

Ferreira-Rodrigues, C. F. (2015). *Estudos com o Enem a partir de uma abordagem psicométrica da inteligência*. Tese de Doutorado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia. Universidade São Francisco, Itatiba-SP.

O objetivo desta tese foi verificar algumas propriedades psicométricas do Enem, exame que se propõe avaliar as habilidades e competências que os estudantes possuem ao final do ensino médio. Para tanto, o trabalho foi dividido em quatro estudos. Um estudo teórico foi realizado e nele é mostrado como se deu a evolução da avaliação da inteligência a partir da abordagem psicométrica, as particularidades da inteligência no adulto, as principais teorias sobre avaliação da inteligência no adulto, além de uma síntese dos achados dos estudos psicométricos realizados com o Enem. O estudo II, verificou se há evidências de validade baseada na estrutura interna da prova de 2006 do Enem. Os resultados demonstram que o modelo teórico do Enem não é encontrado nos itens da prova e os itens se agruparam em dois fatores que são correlacionados e possuem principalmente características de G<sub>q</sub>, G<sub>f</sub>, G<sub>v</sub> (Fator 1) e G<sub>rw</sub>, G<sub>c</sub> e G<sub>f</sub> (Fator 2). O estudo III, investigou a capacidade do Enem em prever desempenho futuro, no caso desse estudo, o desempenho no Enade. O coeficiente de validade preditiva na parte objetiva de todas áreas (Administração, Ciências Contábeis, Direito e Psicologia) foi em torno de 0,40, sendo esse índice semelhante aos encontrados em outros estudos de evidência de validade preditiva que utilizaram o desempenho no exame americano SAT. O estudo IV verificou a estabilidade dos escores dos estudantes que realizaram o Enem em pelo menos dois anos. A estabilidade dos escores dos estudantes é alta e as partes das provas tanto do antigo modelo teórico (2007 e 2008), quanto do novo modelo teórico do Enem (2009) são fortemente e significativamente correlacionadas. Ademais, tanto no estudo III quanto no estudo IV, foi evidenciado que a redação precisa ser melhor estudada em outras pesquisas, visto que ela acrescenta pouca informação tanto em termos de relação com as partes objetivas da prova, como quando relacionada com a redação de outros anos.

Palavras-chave: avaliação educacional, rendimento escolar, vestibular, Enade

## ABSTRACT

Ferreira-Rodrigues, C. F. (2015). Studies with the Enem from a psychometric approach of the intelligence. *Doctoral Thesis*, Programa de Pós-Graduação *Stricto Sensu* em Psicologia. Universidade São Francisco, Itatiba, SP.

The objective of this thesis was to investigate some psychometric properties of the Enem, take it proposes to assess the skills and competencies that students have at the end of high school. To this end, the work was divided in four studies. A theoretical study was carried out and it is shown how the evolution of intelligence assessment from the psychometric approach, was intelligence peculiarities in adults, the main theories of intelligence assessment in adults as well as a summary of the findings of psychometric studies performed with the Enem. The study II, checks for evidence of validity based on the internal structure of the proof 2006 Enem. The study II, checks for evidence of validity based on the internal structure of the proof 2006 Enem. The results show that the theoretical model of the Enem is not found on test items and the items were grouped into two factors that are correlated and mostly have characteristics of Gq, Gf, Gv (Factor 1) and Grw, Gc and Gf (Factor 2). The III study investigated the ability of Enem in predicting future performance in the case of this study, the performance Enade. The III study investigated the ability of Enem in predicting future performance in the case of this study, the performance Enade. The predictive validity coefficient in the objective part of all areas was around 0.40, and this index similar to those found in other studies of predictive validity evidence used performance in the US survey SAT. Study IV found the stability of the scores of students who took the Enem in at least two years. The stability of the scores of students is high and the parts of the evidence from both the old theoretical model (2007 and 2008), as the new theoretical model of Enem (2009) are strongly and significantly correlated. Moreover, both Study III as the IV study, it was shown that the wording needs to be better studied in other research, since it adds little information both in relation to the objective pieces of evidence, such as when related to the writing of others years.

Keywords: Educational evaluation, academic achievement, student admission criteria

## RESUMEN

Ferreira-Rodrigues, C. F. (2015). *Estudios del Enem desde un enfoque psicométrico de la inteligencia*. Tesis Doctoral, Programa de Estudios de Posgrado en Psicología, Universidad San Francisco, Itatiba, São Paulo.

El objetivo de esta tesis ha sido investigar algunas propiedades psicométricas del Enem, tómelo propone evaluar las habilidades y competencias que los estudiantes tienen al final de la escuela secundaria. Para ello, el trabajo se divide en cuatro estudios. Un estudio teórico se llevó a cabo y se muestra cómo fue la evolución de la evaluación de la inteligencia desde el enfoque psicométrico, peculiaridades de inteligencia en los adultos, las principales teorías de la evaluación de la inteligencia en los adultos, así como un resumen de las conclusiones de los estudios psicométricos realizado con el Enem. El estudio II, no hay evidencia de validez basado en la estructura interna de la prueba 2006 Enem. Los resultados muestran que el modelo teórico del Enem no se encuentra en los elementos de prueba y los elementos se agruparon en dos factores que están correlacionados y en su mayoría tienen características de Gq, Gf, Gv (Factor 1) y Grw, Gc y Gf (Factor 2). El estudio III investigó la capacidad del Enem para predecir el rendimiento futuro en el caso de este estudio, el rendimiento Enade. El coeficiente de validez predictiva en la parte objetiva de todas las áreas fue de alrededor de 0,40, y este índice similar a los encontrados en otros estudios de evidencia predictiva validez utilizados rendimiento en la investigación estadounidense con el SAT. Estudio IV encontró la estabilidad de las puntuaciones de los estudiantes que tomaron el Enem en al menos dos años. La estabilidad de las puntuaciones de los estudiantes es alta y las partes de las pruebas, tanto del viejo modelo teórico (2007 y 2008), como el nuevo modelo teórico de Enem (2009) se correlacionan fuertemente y de manera significativa. Por otra parte, tanto el Estudio III como el estudio IV, se demostró que la redacción debe ser estudiado mejor en otras investigaciones, ya que añade poca información tanto en relación con las piezas objetivas de pruebas, como cuando se refieren a la redacción de los demás años.

Palabras clave: evaluación educacional, rendimiento escolar, vestibular

## SUMÁRIO

<b>Lista de Figuras.....</b>	<b>xiii</b>
<b>Lista de Tabelas.....</b>	<b>xiv</b>
<b>APRESENTAÇÃO.....</b>	<b>15</b>
<b>Estudo 1: A Inteligência no jovem adulto e o Enem.....</b>	<b>19</b>
<b>Estudo 2: O que o Enem mede? Análise da estrutura fatorial dos itens de 2006.....</b>	<b>39</b>
<b>Estudo 3: O Enem consegue prever o desempenho no Enade? Evidências de Validade Preditiva utilizando uma abordagem multinível.....</b>	<b>60</b>
<b>Estudo 4: Quão estáveis são os resultados do Enem em estudantes que fizeram o exame em 2007, 2008 e 2009?.....</b>	<b>82</b>
<b>Considerações Finais.....</b>	<b>92</b>

**LISTA DE FIGURAS****Estudo II**

Figura 1. Modelo CHC e sua relação hipotética com as cinco competências do Enem.....	44
Figura 2. Visualização dos <i>eigenvalues</i> dos fatores extraídos na análise fatorial.....	48
Figura 3. Exemplos de itens do fator 1 da prova amarela do Enem de 2006.....	51
Figura 4. Exemplos de itens do fator 2 da prova amarela do Enem de 2006.....	52
Figura 5. Exemplo de item que não se destacou em nenhum fator da análise fatorial.....	53
Figura 6. Mapa de itens da prova amarela do Enem 2006.....	55

**Estudo IV**

Figura 1. Modelo de três fatores com as cargas padronizadas.....	90
------------------------------------------------------------------	----

## LISTA DE TABELAS

### **Estudo II**

Tabela 1. Solução de fatores, cargas fatoriais dos itens da prova amarela do Enem de 2006 .....	49
----------------------------------------------------------------------------------------------------	----

### **Estudo III**

Tabela 1. Número de instituições de ensino superior com dados do Enem e Enade por ano.....	67
Tabela 2. Modelos da análise multinível do Enem para área de Psicologia.....	73
Tabela 3. Modelos da análise multinível do Enem para área de Administração.....	74
Tabela 4. Modelos da análise multinível do Enem para área de Ciências Contábeis.....	74
Tabela 5. Modelos da análise multinível do Enem para área de Direito.....	75

### **Estudo IV**

Tabela 1. Principais diferenças entre o antigo e o novo Enem.....	83
Tabela 2. Participantes dos Enade 2012 e número de vezes que fizeram o Enem.....	85
Tabela 3. Distribuição dos Estudantes que realizaram o Enem em 2007, 2008 e/ou 2009.....	86
Tabela 4. Correlação entre os Enem de 2007, 2008 e 2009.....	88

## APRESENTAÇÃO

O ato de avaliar está presente na vida das pessoas, existindo pelo menos dois tipos de avaliação: uma mais geral, de senso comum, que as pessoas fazem no cotidiano, e outra com características científicas que os pesquisadores e psicólogos devem fazer (Pasquali, 2001). Assim, independentemente do tipo, as avaliações estão presentes em todas as áreas da ação humana (Mello et al., 2001). Na área educacional, foco da presente pesquisa, um dos tipos é a avaliação em larga escala que possui, principalmente, duas finalidades: (1) a seleção, como é feita nos vestibulares e concursos, classificando e hierarquizando os candidatos e (2) a verificação de como se encontram os sistemas de ensino (Luna, 2011), gerando assim, informações quanto a situação educacional do país, o que os alunos aprenderam ou deveriam ter aprendido a partir do que é estipulado de assunto e habilidades no currículo (Castro, 2009).

No Brasil, a avaliação em larga escala no âmbito educacional começou a ser realizada, principalmente, a partir dos anos 90 do século passado, como forma de diminuir problemas educacionais e aumentar os padrões de desempenho (Becker, 2010; Vianna, 2003). No contexto de criação e consolidação desse tipo de avaliação na educação brasileira, insere-se a Psicometria, área responsável pela mensuração em ciências psicossociais e que utiliza a medida para estudar os fenômenos naturais (Pasquali, 2001). Nas avaliações educacionais em larga escala, a psicometria, frequentemente, está inserida e os estudos nessa área, podem utilizar critérios básicos da psicometria para avaliação da qualidade desses instrumentos de medida, sendo tais critérios chamados de validade, precisão e normatização.

De maneira geral, validade corresponde ao quanto as evidências encontradas correspondem a interpretação proposta nos escores do teste que foi construído para uma



determinada finalidade (American Educational Association, American Psychological Association, National Council on Measurement in Education – AERA, APA & NCME, 2014). Os estudos de evidências de validade trazem informações importantes sobre o alcance e as limitações dos resultados encontrados nos instrumentos. Nessa tese, serão enfatizadas duas evidências de validade, a baseada na estrutura interna e a baseada na relação com variáveis externas, de maneira específica, a preditiva.

A precisão ou fidedignidade, refere-se à estabilidade dos escores dos testes dos examinandos de forma que os escores sejam consistentes e com uma quantidade mínima de erro (Urbina, 2007), ou seja, que a flutuação nos escores dos examinandos, em momentos diferentes, seja mínimo (Zanon & Hauck Filho, 2015). Na presente tese, não se falará em precisão do Enem, uma vez que no intervalo de um ano os estudantes estudam para melhorar o seu desempenho, bem como as provas são diferentes. Assim, o que será mostrado é a estabilidade da pontuação relativa do indivíduo de um ano para outro, ou seja, se a posição relativa do estudante em relação a ele mesmo, no intervalo de um exame para outro, muda.

No exterior, a exemplo do que é encontrado em institutos como o ETS (Educational Testing Service) que desenvolve, administra e avalia 50 milhões de testes em mais de 180 países anualmente, vários estudos, análises e políticas educacionais são desenvolvidas (ver <http://www.ets.org/>). No Brasil, essa tradição não é tão frequente. Isso pode ser observado realizando uma busca em bases de dados científicas, em que poucos estudos empíricos são encontrados com as provas educacionais desenvolvidas no país.

No programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco (USF), os professores Dr. Ricardo Primi e Dra. Claudette Maria Medeiros Vendramini, nos últimos dez anos, vêm desenvolvendo estudos com as provas educacionais

do Ministério da Educação (MEC), especialmente o Enem e o Exame Nacional de Desempenho dos Estudantes (Enade), contribuindo com o aumento do número de estudos dessas avaliações.

A presente pesquisa foi realizada tendo em vista a disponibilidade de informações sobre os estudantes brasileiros, a partir das avaliações educacionais realizadas nesse país, devido a relevância social de estudar essa prova a qual causa impacto na vida de milhares de estudantes que a realizam anualmente, além da necessidade de se avaliar tais medidas. Ademais, do ponto de vista psicológico, é preciso verificar as habilidades cognitivas que esses sistemas de avaliação dizem mensurar.

Nesse sentido, a tese que ora se apresenta, em uma abordagem dentro da psicologia da inteligência, buscou investigar algumas propriedades psicométricas do Enem. Para tanto, foram realizados quatro estudos independentes, porém, relacionados a mesma temática. Inicialmente, foi realizado um estudo teórico intitulado de *A inteligência no jovem adulto e o Enem*, que enfatiza o desenvolvimento dos estudos na área da avaliação da inteligência e especificamente da inteligência no adulto, sua relação com o Enem, problematiza a situação atual das pesquisas empíricas com esse exame e o que ainda tem para avançar. O estudo II, *O que o Enem mede? Análise da estrutura fatorial dos itens de 2006*, configura-se como um estudo de evidências de validade baseada na estrutura interna da prova por meio da análise fatorial de itens e de uma análise teórica de todos os itens do exame utilizando uma visão psicométrica da inteligência.

O estudo III chamado de *O Enem consegue prever o desempenho no Enade? Evidências de Validade Preditiva utilizando uma abordagem multinível*, verificou a relação do Enem com variáveis externas. De forma mais específica, foi verificado se há evidências de validade de critério preditiva do Enem utilizando o desempenho do aluno no Enade. Por

fim, no estudo IV “*Quão estáveis são os resultados do Enem em estudantes que fizeram o exame em 2007, 2008 e 2009?*” foram selecionados alunos que realizaram o Enem em pelo menos dois anos para verificar a estabilidade dos escores dos estudantes, como também foram correlacionadas a prova do modelo teórico antigo do Enem e a do novo Enem.

## ESTUDO I

TÍTULO EM PORTUGUÊS: **A INTELIGÊNCIA NO JOVEM ADULTO E O ENEM**

TÍTULO EM INGLÊS: **THE INTELLIGENCE IN YOUNG ADULTS AND ENEM**

### **Resumo:**

A avaliação da inteligência no adulto possui especificidades se comparada à avaliação da inteligência na criança. Por isso, o estudo apresenta as particularidades da inteligência na fase adulta, mostrando como se deu a evolução dos estudos psicométricos até chegar ao modelo CHC, como ocorre a transição de Gf para Gc com o passar da idade. Apresenta dois modelos mais atuais de avaliação da inteligência no adulto (TWA e PPIK) que integram outros construtos como características de personalidade e interesses na explicação do desenvolvimento intelectual no adulto. Por fim, entende-se que as avaliações que são realizadas ao final do Ensino Médio para entrada na universidade podem mensurar habilidades cognitivas. Assim, são sintetizados os achados dos estudos psicométricos com a prova do Enem para entender melhor como é realizada a avaliação cognitiva nessa prova.

Palavras-chave: testes psicológicos, desenvolvimento do adulto, avaliação psicológica

### **Abstract:**

The intelligence assessment in adults has specificities if compared with the intelligence assessment in children. Therefore, the study presents the particularities of the intelligence in adulthood, showing how psychometric studies had evolved before reaching the CHC model, as the transition from Gf to Gc occurs with the increasing age, presents two most current models of the intelligence assessment in adults (TWA and PPIK) which integrate other constructs such as personality traits and interests in the explanation of the intellectual development in adults. Finally, it is understood that evaluations that are conducted at the end of high school to university entrance can measure cognitive abilities. Thus, the findings of psychometric studies with proof of the Enem paragraph entender best how and performed a cognitive assessment that test.

Keywords: psychological Testing, adult development, psychological assessment

Os primeiros estudos da avaliação da inteligência moderna datam do início do século XX com os trabalhos de Alfred Binet, que tinha um grande interesse pelas diferenças individuais. Binet foi convidado a fazer a avaliação de crianças para verificar aquelas aptas a frequentarem a escola tradicional (Ackerman, 2013). Com seu colaborador Theodore Simon, identificou a existência de três formas de avaliar a inteligência. A primeira corresponde ao método médico, que analisa as características biológicas (anatomia, fisiologia e sinais patológicos), a segunda trata-se do método pedagógico o qual focaliza o conhecimento adquirido, por exemplo, a leitura e a escrita e, por fim, o método psicológico que avalia a inteligência de forma direta, indicando o nível de inteligência de crianças (ex. memória, raciocínio), sendo este último considerado por Binet como o mais importante e direto dos três (Binet & Simon, 1905).

É importante destacar que embora Binet considerasse os testes psicológicos como um método de avaliação direta, a visão atual em Psicologia compreende que os instrumentos psicológicos mensuram de forma indireta características psicológicas, já que são atributos os quais não podem ser aferidos de forma direta como quando se mede, por exemplo, o tamanho de uma mesa ou a distância entre duas cidades. Áreas como as neurociências e a psicologia cognitiva, por exemplo, mensuram mecanismos neurais e processos mentais, respectivamente, utilizando exames de neuroimagem e testes cognitivos como ferramentas para conseguir inferir algo do intelecto.

Para o método psicológico, Binet criou a escala Binet-Simon composta por 30 testes que mediam a capacidade de compreensão, raciocínio, julgamento e invenção (Binet & Simon, 1905). A escala avaliava a idade mental de crianças e foi construída para verificar o desempenho na escola e a deficiência mental, tendo obtido um grande sucesso em prever o desempenho escolar. Como o ponto central era a avaliação da idade mental das crianças,

pressuponha-se que as crianças mais velhas deveriam acertar os itens que as crianças de idade inferior erravam (Ackerman, 1996, 2013; Weschler, 1973).

Com o início da Primeira Guerra Mundial, o exército dos Estados Unidos utilizou testes de inteligência para selecionar os seus soldados, iniciando a partir desse momento os estudos e avaliação da inteligência no adulto. Os testes utilizados nesse processo foram baseados nos desenvolvidos por Binet. Contudo, tinham duas características específicas: 1) a linguagem dos itens foi adaptada para os adultos e 2) ao contrário das crianças, os testes passaram a ser aplicados a um grande número de pessoas ao mesmo tempo. Essa mudança na forma de aplicação exigia que o respondente soubesse ler e escrever e permitiu o uso de questões de múltipla escolha (Yoakum & Yerkes, 1920).

Mesmo com as mudanças descritas, os resultados nos testes dos adultos foram comparados com os resultados de crianças nos testes Stanford-Binet e verificado que adultos apresentavam em média uma idade mental referente a adolescentes de 13 anos. Com isso, notou-se que a utilização do conceito de idade mental após a adolescência era pouco adequado (ver Lippmann, 1922), surgindo a necessidade de estudar as especificidades da inteligência na vida adulta.

Conrad (1930) analisou o resultado no teste Army Alpha, que foi baseado na escala Binet-Simon e utilizado no exército, de pessoas de 10 a 60 anos e observou, além do escore geral, os resultados em 8 subtestes, verificando que aos 20 anos de idade os escores atingiam o nível máximo do teste, mas as pessoas com idades mais avançadas não diminuían seus escores em testes verbais e que requeriam conhecimento cultural. Começou-se a acreditar que alguns aspectos da inteligência no adulto tendiam a diminuir, mas outras características pouco se alteravam.

Para Ackerman (2013), Weschler foi o primeiro estudioso que efetivamente avançou no estudo da avaliação da inteligência no adulto, pois modificou o formato do item (antes com conteúdo mais voltado para crianças), retirou o foco da avaliação centrada na idade mental, já que algumas habilidades pareciam ter melhor desempenho na vida adulta, além de ter buscado outros lugares de avaliação fora da sala de aula. Porém, não estabeleceu normas por idade, podendo os resultados serem confundidos com os efeitos do envelhecimento. Além disso, seus achados demonstraram que o auge da inteligência continuava aos 20 anos de idade sofrendo um declínio com o passar da idade.

Atualmente, com o avanço do estudo da inteligência, nota-se que os testes aplicados até esse período, como o Army Alpha, usado no exército, e a própria escala Binet-Simon não mediam adequadamente a inteligência no adulto. Ademais, como afirma Weschler (1973), deve-se atentar para que o estudo da inteligência sempre esteve mais voltado para as crianças e a psicometria, no entanto, desenvolveu grande parte dos seus estudos com adultos.

Do ponto de vista psicométrico, uma grande contribuição no estudo da inteligência foi dada por Cattell (1943, 1963) que propôs que o fator *g* de Spearman, entendido como uma aptidão mais geral que as pessoas possuem, pode ser mais bem explicado a partir da divisão de *g* em dois fatores que possuem a mesma importância. Esses foram chamados de inteligência cristalizada (*Gc*) que corresponde ao conhecimento e competências adquiridas ao longo da vida por meio da escolarização tendo forte influência cultural e educacional e a inteligência fluida (*Gf*) que corresponde a uma capacidade que exige adaptação a novas situações. *Gf* desenvolve-se na infância até o início da fase adulta e depois começa a haver um declínio, sendo importante para aquisição de conhecimento (*Gc*).

Com a definição dessas duas habilidades, Cattell (1963) criou a teoria de investimento Gf-Gc com o objetivo de descobrir o motivo de algumas pessoas saberem mais do que outras. O autor percebeu que uma baixa inteligência fluida limita a capacidade da inteligência cristalizada, ou seja, restringe a capacidade de adquirir e guardar novos conhecimentos. Por isso, os dois fatores são altamente correlacionados, visto que Gf influencia em parte a aquisição de Gc e é dessa relação que o fator g emerge (Cattell, 1943). O mais interessante da teoria Gf-Gc de Cattell é a capacidade de descrever os dois fatores, demonstrar como Gf-Gc surgiu a partir do fator g de Spearman e como tais achados resistiram ao tempo (Schneider & McGrew, 2012).

Para Cattell (1987), grande parte da aprendizagem é atribuída ao esforço, sendo que a capacidade de algumas pessoas se envolverem mais no processo de aprendizagem depende do investimento social, incluindo a qualidade da educação disponível, do investimento da família que corresponde aos recursos e expectativas familiares e o investimento pessoal, ou seja, aquele relacionado aos interesses e objetivos pessoais. Para Carroll (1963), no entanto, a aprendizagem varia de pessoa para pessoa visto que depende da quantidade de tempo que uma pessoa passa engajada na tarefa *versus* a quantidade de tempo necessária para essa pessoa conseguir aprender. Dessa forma, no engajamento do tempo estão relacionadas as oportunidades que a pessoa tem, mais a motivação e a persistência na tarefa, ao passo que a quantidade de tempo necessário depende da capacidade prévia da pessoa e da qualidade do ensino.

Posteriormente, Horn, aluno de Cattell, ampliou a teoria Gf-Gc e incluiu nela novos fatores aos dois já existentes, quais sejam, inteligência visual (Gv), velocidade de processamento (Gs), memória de curto prazo (Gsm) e armazenamento e recuperação a longo prazo (Glr) (Horn & Cattell, 1966). Horn continuou estudando com Cattell de 1965



ao início da década de 1990, e juntos ampliaram o número de fatores para incluir (Ga) processamento auditivo, (CDS) rapidez para decisão correta e (Gq) conhecimento quantitativo (McGrew & Flanagan, 1998). Além dessas capacidades cognitivas, ainda descreveram Grw que corresponde à capacidade de leitura e escrita (Schneider & McGrew, 2012).

John B. Carroll, por sua vez, fez uma reanálise dos mais importantes estudos realizados em quase um século de investigação dos testes mentais, utilizando vários métodos de análise fatorial (Carroll, 1993; McGrew, 2009; McGrew, 1997; McGrew & Flanagan, 1998; Primi, 2003). A ele é atribuída pela primeira vez uma organização taxonômica coerente da estrutura das habilidades cognitivas humanas (McGrew, 2009), pois propôs o modelo de três estratos das habilidades cognitivas. Para ele, em um nível mais amplo, denominado estrato III, encontra-se o fator g. No estrato II, são visualizadas oito habilidades: inteligência fluida (Gf); inteligência cristalizada (Gc); memória e aprendizagem geral (Gy); percepção visual (Gv); percepção auditiva (Gu); capacidade de recuperação (Gr); rapidez cognitiva (Gs) e velocidade de processamento de reação (Gt) e o estrato I, é composto por inúmeras habilidades (mais de 70) que subdividem os fatores do estrato II (Carroll, 1993; McGrew & Flanagan, 1998).

A partir das definições do modelo dos três estratos de Carroll e da teoria Gf-Gc de Horn-Cattell, observou-se que há similaridade entre as duas propostas teóricas, especialmente, quanto às habilidades amplas (estrato II no modelo de Carroll) (McGrew, 1997). Com isso, McGrew e Flanagan (1998) propuseram um modelo de integração das teorias Gf-Gc e dos três estratos, denominado de Teoria de Cattell-Horn-Carroll (CHC) das habilidades cognitivas (McGrew, 2009). Surge, então, o modelo CHC como uma taxonomia que organizou os resultados do passado e direciona os estudos para novas

investigações, além de explicar as diferenças cognitivas das pessoas (Schneider & McGrew, 2012).

Após sua criação, já foram incluídos ao modelo outras habilidades, sendo elas, a capacidade de conhecimento geral (Gkn), habilidades cinestésicas (Gr), olfativas (Go) e psicomotora (Gp) (McGrew, 2005). O modelo CHC representa a melhor forma de descrever a estrutura da inteligência e grande parte dos testes que avaliam a inteligência atualmente são baseados nesse modelo, mesmo que de maneira implícita em alguns testes ou aqueles que foram revisados e que não possuíam a influência desse modelo, também conseguem ser explicados por essa teoria (Keith & Reynolds, 2010).

#### *Integração de outros construtos ao estudo da inteligência*

A avaliação no estudo da inteligência do adulto avançou muito ao longo de pouco mais de um século. Nos últimos 20 anos, para além das questões teóricas sobre definição, mensuração e estrutura da inteligência humana, verifica-se o aumento do número de pesquisas que investigam a relação e importância de outros traços na explicação do desenvolvimento da inteligência nessa fase da vida (Ackerman, 1996), os chamados traços de investimento.

Os traços de investimento podem explicar algumas diferenças individuais na busca por conhecimento, ou seja, em buscar por oportunidades de aprendizagem (vonn Stumm, 2012). Em um estudo de meta-análise recente (vonn Stumm & Ackerman, 2013) foram agrupados 8 categorias de traços de investimento, tais traços, segundo Ackerman (1996) podem determinar quando, onde e como as pessoas investem sua capacidade cognitiva. Um núcleo central de investimento foi caracterizado por compreender pessoas que são envolvidas (cognitivamente) em atividades, buscando, participando e compreendendo o meio ambiente. Duas escalas foram relacionadas a esse núcleo, quais sejam: a escala de

engajamento intelectual típico e a escala de necessidade de cognição. A *curiosidade intelectual*, por sua vez, compreende a necessidade em adquirir conhecimento e se envolver em atividades cognitivamente estimulantes. O *pensamento abstrato* envolve o interesse por atividades de resolução de problemas, e exploração de questões teóricas. Nesse traço, as informações são livres para se chegar à resolução ou compreensão da questão, havendo assim um envolvimento intelectual para compreender e resolver “quebra-cabeças”. A *procura por novidades*, por sua vez, é um traço que se refere à necessidade de buscar emoções, que podem estar relacionadas a assumir riscos.

Outro traço de investimento encontrado é a *abertura* que abrange a capacidade de ser sensível e perceber o seu próprio ambiente, ou seja, é caracterizado por aquelas pessoas que tendem a ter uma imaginação mais aberta. A *absorção* é vista como um traço de investimento que se opõe às características de abertura por envolver uma maior introspecção do indivíduo em suas próprias experiências. O penúltimo traço é a *ambiguidade* que é a tolerância à incerteza, ao desconhecido. Por fim, a *curiosidade social* é uma categoria que não se encaixa em nenhuma outra e pode envolver o interesse por assuntos paralelos (vonn Stumm & Ackerman, 2013).

Dessa forma, personalidade, inteligência e interesses, embora já tenham sido considerados como não relacionados (Ackerman & Beier, 2003), nos últimos 20 anos, têm sido vistos como aspectos que contribuem para o desenvolvimento intelectual no adulto (ver Ackerman, 1996; Ackerman & Heggestad, 1997; Ackerman & Roffhus, 1999). Além de reunir informações sobre os perfis de habilidades específicas dos participantes, os estudos procuram compreender como os estudantes fazem escolhas importantes em suas vidas, levando-se em consideração a integração dessas características (Lubinski & Benbow, 2000; Wai, Lubinski, & Benbow, 2005). Duas teorias surgiram com esse propósito

integrativo e, por isso, essas serão discutidas a seguir como forma de mostrar o panorama atual no estudo da inteligência do adulto.

Ackerman (1996) propôs a teoria PPIK (inteligência-como-**P**rocesso, **P**ersonalidade, **I**nteresses, inteligência-como-conhecimento [na sigla, a letra K corresponde a *Knowledge*]) para demonstrar como os interesses e a personalidade podem provocar mudanças na inteligência durante toda a vida (Ackerman, 1996) e ajudar ou atrapalhar a aprendizagem (Ackerman, 2014). No modelo PPIK, a inteligência é dividida em dois tipos seguindo o pensamento dos teóricos Cattell e Hebb. Inteligência-como-processo está diretamente ligada à ideia de inteligência fluida (Gf), obtendo melhores resultados nos testes que avaliam esse tipo de inteligência, o adulto jovem. Ao passo que a inteligência-como-conhecimento está relacionada à ideia de inteligência cristalizada. Nesse sentido, os testes para avaliar o conhecimento nos adultos devem ser pensados tanto para avaliar o conhecimento de cada profissão, pois o conhecimento muda de acordo com cada contexto profissional, quanto em termos mais comuns, pois as pessoas, independentemente da profissão, também possuem conhecimentos gerais (Ackerman, 1996).

A personalidade também é incluída na teoria, sendo compreendida como a forma com que as pessoas investem e orientam seus traços de personalidade no seu desenvolvimento intelectual (Ackerman, 1996; vonn Stumm & Ackerman, 2013). Tem-se que, algumas das características de personalidade são mais estudadas por contribuírem para o investimento intelectual, especialmente a habilidade verbal, sendo elas: a abertura à experiência e o intelecto (Ackerman, 1996; Ackerman & Heggstad, 1997; Beier, Campbel, & Crook, 2010; Chamorro-Premuzic & Furnham, 2006). Os dois são aspectos relacionados que correspondem a um domínio dos cinco grandes fatores de personalidade (DeYoung, Quilty, & Peterson, 2007).

As facetas da abertura à experiência correspondem à busca pela fantasia (imaginação fértil), ações (interesse por novidades), apreciação da arte (estética), sentimentos (emoções), ideias (curiosidade intelectual) e valores. Os traços de intelecto, por sua vez, correspondem ao interesse por se engajar em atividades intelectualmente estimulantes (vonn Stumm & Ackerman, 2013).

Por fim, no modelo PPIK, os interesses também influenciam o desenvolvimento da inteligência no adulto. Para avaliar esse componente, é utilizado o modelo hexagonal de interesses profissionais de Holland (1959) que abrange os tipos: Realista, Investigativo, Artístico, Social, Empreendedor e Convencional. Dos seis interesses, os que estão relacionados à inteligência e a fatores de personalidade são os tipos: investigativo, realista e artístico (Ackerman, 1996). No modelo PPIK, os interesses realista e investigativo correlacionam-se tanto com a inteligência-como-processo (capacidade de raciocínio, matemática, ciências físicas) como com as habilidades de processo que estão mais próximas do conhecimento. Os interesses artísticos estão mais ligados à inteligência-como-conhecimento (habilidades verbais/cristalizadas como literatura) e o interesse investigativo também apresenta correlações significativas com habilidades cristalizadas, especialmente do tipo das ciências sociais e físicas (Rolfhus & Ackerman, 1996).

A teoria PPIK é útil para compreender porque pessoas com perfis cognitivos parecidos tem conhecimentos diferentes. Além disso, demonstra que as diferenças nos atributos não cognitivos pessoais influenciam o desenvolvimento de conhecimentos e habilidades específicas, como também leva em consideração as diferentes oportunidades que cada indivíduo possui ao longo da vida no contexto em que está inserido. Dessa forma, verificar e valorizar as dimensões de interesses, personalidade e inteligência-como-processo

permite oferecer um maior suporte na trajetória desenvolvimental com as habilidades cristalizadas (Lubinski & Benbow, 2000).

Lubinski e Benbow (2000) também desenvolveram um modelo de avaliação da inteligência no adulto, a chamada Teoria do Ajustamento do Trabalho (TWA), desenvolvida para pessoas adultas e no mundo do trabalho, sendo útil para organizar achados psicométricos nas dimensões de habilidades e interesses e, assim, facilitar o ajustamento vocacional e educacional (Achter, Lubinski & Benbow, 1996). De acordo com a TWA o ajuste educacional e vocacional envolve duas dimensões de correspondência, a satisfatoriedade (competência) e a satisfação (realização). A satisfatoriedade é determinada pela correspondência entre as habilidades e as demandas de habilidades do ambiente. A satisfação é determinada pela correspondência entre as necessidades pessoais e recompensas fornecidas pelo ambiente. Assim, na medida em que satisfatoriedade e satisfação co-ocorrem, a pessoa e o ambiente são ditos como em harmonia e, desse modo, ambos são motivados para continuar a interagir um com o outro, para se beneficiarem (Ackerman & Beier, 2003).

A TWA coloca igual ênfase na avaliação da pessoa e na avaliação do ambiente. O desenvolvimento ideal ocorre quando necessidades das pessoas são atendidas e suas habilidades são adequadamente desafiadas. Estudantes que são fortes principalmente em raciocínio verbal versus quantitativo ou raciocínio espacial tendem a seguir áreas mais relacionadas as suas habilidades (Achter, Lubinski, Benbow, & Eftekhari-Sanjani, 1999).

Essa visão integrativa das teorias de avaliação da inteligência no adulto pode ajudar a compreender as diferenças individuais na vida cotidiana das pessoas em relação, por exemplo, à realização profissional, pessoal, saúde e longevidade (vonn Stumm, Chamorro-Premuzic, & Ackerman, 2011). Não obstante, no contexto brasileiro, mesmo no processo

de escolha de carreira, por exemplo, geralmente, nas avaliações, habilidades, personalidade e interesses são considerados de maneira isolada, não sendo vistos, muitas vezes, como partes integrantes do indivíduo.

*O Exame Nacional do Ensino Médio como forma de avaliação da inteligência no jovem adulto*

Em termos práticos, na testagem do adulto, tem-se que os exames que são realizados ao final do ensino médio, como o *Scholastic Assessment Test* (SAT), para entrada nas universidades americanas, podem representar o nível da inteligência na vida adulta (vonn Stumm & Ackerman, 2013). Os exames de avaliação educacional implementados no Brasil também têm como foco as habilidades cognitivas (Primi et al., 2001).

No Brasil, a avaliação realizada nessa etapa do ensino é denominada Exame Nacional do Ensino Médio (Enem) e tem como proposta medir as estruturas mentais que são utilizadas para construção do conhecimento, não se restringindo apenas à memorização, uma vez que para compreensão do mundo atual são necessárias várias estruturas. Este é um aspecto que diferencia o Enem das avaliações que são focadas nos conteúdos memorizados como é o caso dos vestibulares tradicionais (Macedo, 2005).

O tipo de conhecimento voltado para o acúmulo de conhecimentos permanece. Contudo, cada vez mais cresce a necessidade do aluno conseguir interpretar e solucionar os problemas (Macedo, 2005). Dessa forma, Primi et al. (2001) apontam que “a resposta do aluno a uma questão deste tipo depende menos da recordação de conhecimentos prévios e mais da recombinação de conhecimentos já existentes de maneira nova” (p.153). As características principais dos itens que compõem o Enem são a interdisciplinaridade, a formulação de situação-problema e a contextualização.

Considerando a relevância de se desenvolver estudos psicométricos dos testes que são construídos, observa-se que não há pesquisas de evidências de validade nos documentos oficiais do Enem. No entanto, este é o aspecto mais básico com relação aos escores de testes e sua utilidade, sendo que somente a partir das evidências de validade encontradas é que se pode fazer alguma inferência com os resultados do teste (Urbina, 2007).

Pensando na questão das propriedades psicométricas em relação à prova, Primi et al. (2001) enfatizaram a importância da realização de estudos que verifiquem exatamente quais as habilidades cognitivas são mensuradas pela prova. Além disso, mostraram a tentativa de o exame dar maior ênfase ao raciocínio em seus itens, assim como ocorre em outros exames como o SAT. Embora haja algumas semelhanças entre os conceitos do Enem e as teorias de inteligência, existem algumas divergências no que se refere ao aspecto estrutural-conceitual, também em relação a sustentação empírica da medida de cinco competências, além da confusão de termos (definição dos termos habilidade e competência para o Enem e para as teorias de inteligência), além da grande abrangência teórica versus o que realmente se consegue operacionalizar.

Após a publicação do estudo desenvolvido por Primi et al. (2001), encontram-se alguns poucos trabalhos que buscaram verificar alguma propriedade psicométrica do Enem. Tais estudos, em síntese, não conseguem confirmar a proposta teórica utilizada pelo Enem (Bartholomeu, 2011; Gomes, 2005; Gomes & Borges, 2009) e com isso, a Teoria CHC vem sendo utilizada para explicar o que o Enem avalia (Bueno, 2013; Gomes, 2005; Gomes & Borges, 2009) e esse exame consegue prever algo em torno de 16% o desempenho do aluno no primeiro ano do curso universitário (Sousa, 2006; Sousa, Vendramini, & Silva, 2013).



### *Considerações Finais*

Os achados dos estudos da inteligência permitem inferir que a ênfase na idade adulta deve ser o conhecimento, ou seja, em Gc e que essa capacidade tende a se manter estável ao longo da vida, da mesma forma que o conhecimento decorrente da ocupação (trabalho) do indivíduo. Por outro lado, Gf ou a inteligência-como-processo tende a diminuir na vida adulta, mas o tamanho desse declínio não pode ser atribuído apenas à idade, pois outros atributos, como a escolaridade, também influenciam na intensidade da diminuição dessa capacidade. Os estudantes ao final do Ensino Médio, época que realizam a prova do Enem, estão em fase de transição entre a adolescência e o início da vida adulta. Dessa forma, a influência de Gf e Gc ainda apresenta-se de maneira equilibrada, ou seja, em transição de uma capacidade mais fluída para um conhecimento mais cristalizado. Ademais, além da definição dos fatores que compõem a inteligência é importante compreender de maneira global quais são os outros aspectos que estão envolvidos na inteligência no adulto.

A partir do que foi apresentado nesse estudo, nota-se, que no Brasil, não há um tipo de avaliação educacional ou mesmo clínica enfatizando a visão atual das teorias do desenvolvimento intelectual nos adultos, ou seja, que busquem avaliar além dos aspectos cognitivos, os traços de investimentos. Ademais, o número de estudos de evidências de validade com a prova do Enem ainda é incipiente e centrado em sua maioria em publicações referentes a dissertações e teses.

Diante desse panorama novas questões podem ser levantadas, para o desenvolvimento de novos estudos, tais como: A prova do Enem mede mais conhecimento do que raciocínio? Houve alguma mudança prática no novo Enem (a partir de 2009) em relação ao modelo antigo (1998 - 2008) ou essa mudança foi apenas na estrutura teórica? O

Enem consegue prever o desempenho do aluno na universidade? Poderia se pensar em um estudo também que relacionasse o desempenho dos estudantes no Enem (capacidade cognitiva), com os interesses desses alunos (curso universitário escolhido) e características de personalidade, seja incluindo alguns itens no questionário socioeconômico do Enem ou selecionando uma amostra de estudantes que fizeram o Enem e aplicando testes de personalidade e interesses nesses alunos, para assim, ter uma visão integrada do perfil dos estudantes como propõe as teorias atuais de inteligência no adulto.

### Referências

- Achter, J. A., Lubinski, D., & Benbow, C. P. (1996). Multipotentiality among intellectually gifted: "It was never there and already it's vanishing." *Journal of Counseling Psychology, 43*, 65-76.
- Achter, J. A., Lubinski, D., Benbow, C. P., & Eftekhari-Sanjani, E. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities. *Journal of Educational Psychology, 91*, 777-789.
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227-257.
- Ackerman, P. L. (2013). Assessment of Intellectual Functioning in Adults. *APA Handbook of Testing and Assessment in Psychology: Vol. 2. Testing and Assessment in Clinical and Counseling Psychology*, 119-132.
- Ackerman, P.L. (2014). Adolescent and Adult Intellectual Development. *Current Directions in Psychological Science, 23*(4) 246-251. doi: 10.1177/0963721414534960

- Ackerman, P. L. & Beier, M. E. (2003). Intelligence, personality, and interests in the career choice process. *Journal of Career Assessment*, *11* (2), 205-218.
- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 218–245.
- Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging*, *14*, 314–330. doi:10.1037/0882-7974.14.2.314
- Bartholomeu, D. (2011). *O uso do Enem e Enade em um delineamento longitudinal para avaliação dos cursos do ensino superior*. Tese de Doutorado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Bueno, J. M. P. (2013). *Análise das habilidades avaliadas pelo Enem por meio da teoria CHC: um estudo de validade*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Beier, M. E., Campbell, M., & Crook, A. E. (2010). Developing and demonstrating knowledge: Ability and non-ability determinants of learning and performance. *Intelligence*, *38*, 179-186.
- Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, *11*, 191–244.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

- Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. New York: Elsevier.
- Chamorro-Premuzic, T., & Furnham, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology*, *10*, 251–267. doi:10.1037/1089-2680.10.3.251.
- Conrad, H. S. (1930). General-information, intelligence, and the decline of intelligence. *Journal of Applied Psychology*, *14*, 592–599. doi:10.1037/h0069963.
- DeYoung C.G., Quilty L.C., Peterson J.B. (2007). Between facets and domains: Ten aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896.
- Ewing, M., Huff, K., Andrews, M. and King, K. (2006) Assessing the reliability of skills measured by the SAT. Office of Research and Analysis, *Research Notes*, *24*, 1-8.
- Gomes, C. M. A. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio*. Tese de doutorado, Programa de Pós-Graduação em Educação: Conhecimento e Inclusão Social, da Faculdade de Educação da Universidade Federal de Minas Gerais, Belo Horizonte.
- Gomes, C. M. A., & Borges, O. N. (2009). O Enem é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, *20*(42), 73-88.
- Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*(1), 35-45.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253-270.

- Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, 47 (7), 635-650.
- Lippmann, W. (1922). The mental age of Americans. *New Republic*, 32, 213-215.
- Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, 55, 137-150.
- Macedo, L. (2005). Competências e habilidades: Elementos para uma reflexão pedagógica. Ministério da Educação. *Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica*. Brasília: Inep.
- McGrew, K.S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 151-179). New York: Guilford Press.
- McGrew, K.S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D.P. Flanagan & P.L. Harrison (Eds), *Contemporary intellectual assessment* (2nd ed., pp. 136-181). New York: Guilford Press.
- McGrew, K.S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10.
- McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.
- Patterson, B. F. & Mattern, K. D. (2011). *Validity of the SAT for predicting first-year grades: 2008 SAT Validity Sample*. College Board Statistical Report. New York, NK: The College Board.

- Patterson, B. F. & Mattern, K. D. (2012). *Validity of the SAT for predicting first-year grades: 2009 SAT Validity Sample*. College Board Statistical Report. New York, NK: The College Board.
- Primi, R. (2003). Inteligência: avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação Psicológica*, 2(1), 67-77.
- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. de F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(2), 151-159.
- Rolfhus, E. L., Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology*, 88, 174-188.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 99-144). New York: Guilford.
- Souza, A. M. (2006). *Validade preditiva de um processo seletivo em relação ao desempenho de universitários de Psicologia*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Sousa, A. M., Vedramini, C. M. M., & Silva, M. C. R. (2013). Validade preditiva de um processo seletivo em relação ao desempenho de universitários de Psicologia. *Encontro Revista de Psicologia*. (16) 24, 55-68.
- von Stumm, S., & Ackerman, P. L. (2013). Investment and intelligence: A review and meta-analysis. *Psychological Bulletin*, 139, 841–869.
- von Stumm, S., Chamorro-Premuzic, T., & Ackerman, P. L. (2011). Re-visiting Intelligence-Personality Associations: Vindicating Intellectual Investment. In T.

- Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *Handbook of Individual Differences*. UK: Wiley-Blackwell.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- Yoakum, C.S. & Yerkes, R. M. (1920). *Army Mental Tests*. New York: Henry Holt and Company.
- Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, 97(3), 484-492.
- Weschler, D. (1973). *La medida de la inteligencia del adulto*. Buenos Aires: Ediciones Huascar.

## ESTUDO II

### TÍTULO EM PORTUGUÊS: O QUE O ENEM MEDE? ANÁLISE DA ESTRUTURA FATORIAL DOS ITENS DE 2006

### TÍTULO EM INGLÊS: WHAT ENEM MEASURES? ANALYSIS OF THE FACTORIAL STRUCTURE OF 2006 ITEMS

#### Resumo

O objetivo do estudo foi verificar o que o Enem avalia a partir de uma visão psicométrica da inteligência. Para tanto, foram realizadas uma análise teórica e uma análise fatorial dos itens da prova amarela de 2006. Foram encontrados dois fatores que explicam 20,6% e 2,9% da variância (Eig1/Eig2=4,54, Rotação Promax  $r [f1 \text{ vs } f2]= 0,67$ ) que estão correlacionados. Portanto, há um fator geral subjacente composto principalmente por características de Gc, Gq, Gf, Gv e Grw. Os itens do fator 1 possuem a característica de ter que identificar e aplicar conceitos de matemática, relacionar informações e fazer deduções, além de aplicar conhecimentos básicos de matemática em problemas do cotidiano. O fator 2, envolve decodificação e compreensão em leitura (vocabulário e rapidez, automatização), relacionar ideias e fazer deduções de conhecimento geral e cultura. Conclui-se que a prova consegue ser explicada por algumas habilidades cognitivas do modelo CHC.

Palavras-chave: análise fatorial; inteligência, avaliação educacional

#### Abstract

The aim of the study was to determine what Enem evaluates. For that, we carried out a theoretical and a factor analyses about the items of the yellow test of 2006. There were two factors that explain 20.6% and 2.9% of the variance (Eig1/Eig2=4,54, Promax rotation  $[f1 \text{ vs } f2]= 0,67$ ) and that are correlated. So, there is an underlying general factor composed mainly of Gc, Gq, Gf, Gv characteristics. Factor 1 items have the characteristic of having to identify and to apply math concepts, to relate information and to make deductions, besides to apply basic mathematical knowledge in everyday problems. The factor 2 involves decoding and reading comprehension (vocabulary and rapidity, automatization), to relate ideas and to make deductions about general and cultural knowledge. The conclusion was that the test can be explained by some cognitive skills of the CHC model.

Kewwords: factor analysis, intelligence, educational evaluation



O Exame Nacional do Ensino Médio (Enem) foi criado com o intuito de verificar as estruturas mentais responsáveis pelo conhecimento, centrando não apenas na memorização de conceitos, como ocorre nos vestibulares tradicionais, mas dando ênfase também a capacidade do aluno raciocinar ao responder os itens. Para atender a essa proposta, os itens são construídos de maneira que sejam interdisciplinares, envolvendo situações-problemas e considerando sempre a contextualização (Macedo, 2005).

Da criação em 1998 até 2008 o Enem foi formulado, a cada ano, como uma prova única, de realização individual, onde participavam estudantes que estavam concluindo o ensino médio ou que já haviam terminado essa etapa da escolaridade. A prova, durante esse período, era composta por 63 questões objetivas e uma redação. Tais questões eram elaboradas a partir de uma matriz de 5 competências e 21 habilidades inter-relacionadas as competências, ou seja, cada competência possuía várias habilidades relacionadas, ou seja, cada habilidade podia fazer parte de mais de uma competência (Inep, 2001).

A partir de 2009, a estrutura foi modificada e as cinco competências passaram a se chamar eixos cognitivos. As competências e habilidades tornaram-se específicas para cada área do conhecimento (Linguagens, códigos e suas tecnologias; Matemática e suas tecnologias; Ciências da Natureza e suas tecnologias; Ciências Humanas e suas tecnologias) e não são mais interligadas. Cada área do conhecimento possui 30 habilidades o que muda em relação a cada área é o número de competências (MEC, 2009).

Nota-se que os termos competências e habilidades fazem parte da estrutura do que esse exame pretende avaliar, sendo necessário compreender esses conceitos na perspectiva do Enem. As competências são vistas como “modalidades estruturais da inteligência, ou melhor, ações e operações que utilizamos para estabelecer relações com e entre objetos, situações, fenômenos e pessoas que desejamos conhecer”, e, as habilidades “decorrem das

competências adquiridas e referem-se ao plano imediato do ‘saber fazer’. Por meio das ações e operações, as habilidades aperfeiçoam-se e articulam-se, possibilitando nova reorganização das competências” (Brasil, 2008, p.18). Embora as definições de habilidade e competência sejam baseadas em Perrenoud e Piaget, tais conceitos não são descritos a partir de um modelo teórico que dê suporte para desenvolver as questões do exame.

Não obstante, os termos competência e habilidade para o Enem, se aproximam dos conceitos de inteligência cristalizada e inteligência fluida, respectivamente. No modelo proposto por Cattell (1943, 1963), a inteligência cristalizada (Gc) é desenvolvida a partir de experiências culturais e educacionais, estando presente na maioria das atividades escolares. Refere-se à extensão e a profundidade das informações adquiridas ao longo do processo de escolarização, sendo baseada primariamente na linguagem (Horn & Cattell, 1966, 1967; McGrew & Flanagan, 1998; Primi et al., 2001; Primi, 2003). A inteligência fluida (Gf) está associada a componentes não-verbais, pouco dependentes de conhecimentos previamente adquiridos e da influência de aspectos culturais. Gf corresponde as operações mentais que as pessoas utilizam frente a uma tarefa relativamente nova e que não pode ser executada automaticamente (McGrew, 1997).

Com o passar do tempo a teoria Gf-Gc de Cattell foi ampliada para incluir novas habilidades cognitivas. Posteriormente, foi proposta a teoria dos três estratos de Carroll (1993) e mais recentemente, o modelo CHC, que integra as ideias teóricas das habilidades cognitivas de Cattell, do seu aluno Horn e de Carroll (Carroll, 1993; McGrew & Flanagan, 1998; McGrew, 2009; Primi, 2003). O modelo CHC enfatiza a natureza multidimensional da inteligência. Reconhece a existência do fator *g*, mas, enfatiza as capacidades amplas. Além disso, vem gradualmente sendo usado como conceito-chave entre profissionais e pesquisadores no entendimento da inteligência (Primi, 2003).

Apesar dessa associação entre os termos habilidade e competência para o Enem e alguns tipos de habilidade do modelo psicométrico de avaliação da inteligência, destaca-se que em 15 anos de existência, o Enem passou por reformulações em sua estrutura, mas, não encontra-se, nos documentos oficiais do exame, estudos que comprovem que a proposta teórica da prova é encontrada empiricamente.

Primi et al. (2001) levantaram essa questão da validade de construto do exame. Os autores enfatizaram não existir suporte teórico para as 5 competências e 21 habilidades do Enem que pudesse abranger todos os aspectos da inteligência como proposto (ver Inep, 1999). Apontaram existir divergência entre as definições de habilidade e competência para o Enem e para outras teorias, como a psicométrica e do processamento da informação, por exemplo, nas quais as definições de habilidade e competências são consideradas de maneira contrária (ver Primi et al., 2001).

Apesar desse panorama é importante compreender que o Enem não foi criado com o objetivo de testar empiricamente o seu modelo. Idealmente, é importante fazê-lo, mas, na prática, não é tão fácil viabilizar, nos exames educacionais, essa perspectiva de construir um teste educacional com base nos critérios psicométricos (análise de itens, validade, fidedignidade) para depois utilizá-lo. Contudo, os estudos que se propõem a testar e estudar esses modelos são de grande importância para compreensão e avanço na qualidade desses instrumentos. O que não se pode é centrar apenas na questão de que o exame não consegue medir o que se propõe. Partindo desse olhar, o presente estudo buscou verificar a estrutura interna do Enem de 2006 (antigo Enem) a partir da abordagem psicométrica da inteligência, ou seja, contribuir para o entendimento do que ele mede por meio de uma visão psicométrica.

Na Figura 1, uma possível relação, ao menos do ponto de vista teórico, entre o modelo CHC e as competências do Enem foram estabelecidas para este estudo. A competência dominar linguagens parece estar mais relacionada à ideia de inteligência cristalizada (Gc), a capacidade de leitura e escrita (Grw) e ao conhecimento quantitativo (Gq). A competência enfrentar situações-problema, também por definição, se relacionaria com a inteligência fluida (Gf) e ao processamento visual (Gv); construir argumentação com Gf; compreender fenômenos com Gc e conhecimento quantitativo. Por fim, a competência de elaborar propostas pode relacionar-se à capacidade de Armazenamento e Recuperação Associativa a Longo Prazo (Glr). Teoricamente essas relações podem ser estabelecidas, mas, na prática, há sustentação empírica que permita fazer essas inferências? Para tentar responder essa questão, inicialmente, foi feito um levantamento das publicações sobre o Enem na base de dados BVS-PSI (artigos e teses) utilizando a palavra Enem como descritor. Apenas 4 publicações que buscaram avaliar a estrutura interna do Enem foram encontradas, sendo que duas referem-se a publicações decorrentes de um mesmo estudo.

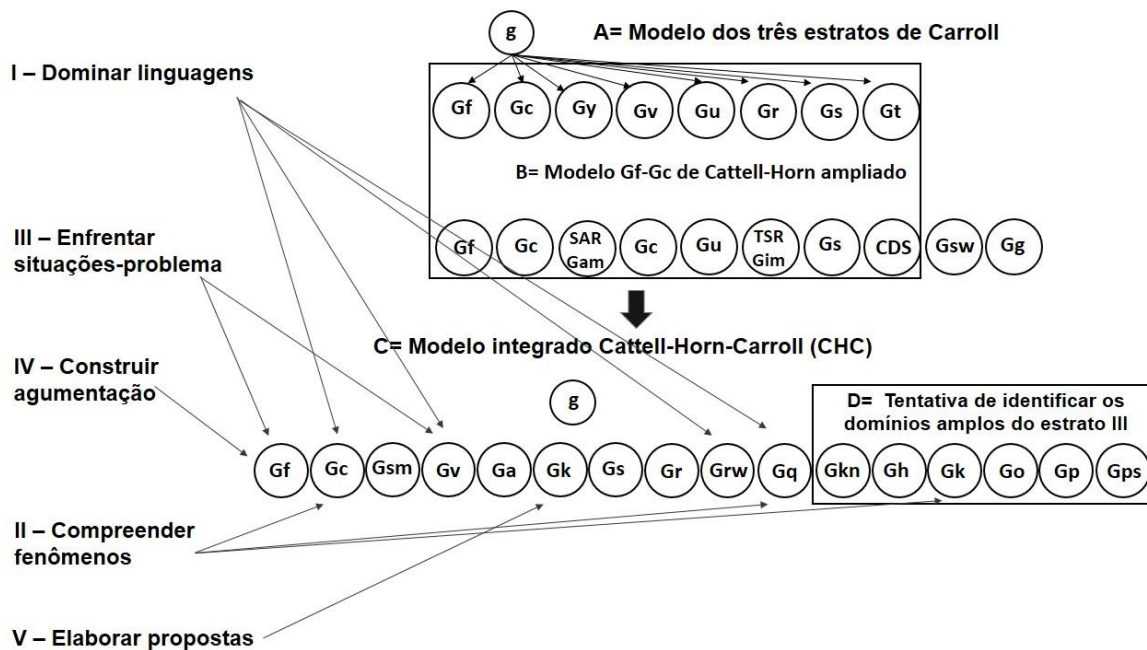


Figura 1. Modelo CHC e sua relação hipotética com as cinco competências do Enem

A prova do Enem de 2001 foi aplicada a 160 estudantes de uma escola da rede federal de ensino, realizada uma análise fatorial que encontrou uma solução de dois fatores que explicou 13,41% da variância. O primeiro fator corresponde a competência verbal e o segundo a uma competência quantitativa. Os dois fatores foram novamente rodados pela rotação *oblimin* e pelas matrizes de correlação e fatorial e verificada a presença de um fator geral que explicou 67,60% da variância. O autor afirma que grande parte da prova de 2001 do Enem pode ser entendida por um fator relacionado a escolarização que ele chama de Fator Geral Escolar e que muitos itens poderiam ser eliminados. Assim, conclui que a prova do Enem de 2001 não corresponde ao modelo teórico adotado pelo Enem (Gomes, 2005; Gomes & Borges, 2009).

Na mesma pesquisa (Gomes, 2005; Gomes & Borges, 2009), também foram realizadas três análises de regressão múltipla entre os fatores de provas de inteligência que medem o fator geral (G), o fator da Habilidade Visual Abrangente (Gv), o fator Memória

(Gy), o fator da Fluência (Gr), o fator da Velocidade Cognitiva (Gs) e o fator de Inteligência Cristalizada (Gc) e cada um dos três fatores encontrados na análise fatorial da prova do Enem de 2001 (Competência Verbal, Competência Quantitativa e a Competência Geral). Constatou-se que a inteligência geral explica a competência verbal. A competência geral e o fator g juntamente com a competência verbal estão mais relacionados a competência qualitativa. Os autores inferem que o fator g no Enem corresponde a inteligência fluida e que o Enem pode ser utilizado como um teste de inteligência que tem como foco principal a resolução de problemas (Gf-Gc) tendo 52% da variância explicada por essa habilidade, em seguida, a rapidez cognitiva (Gs) explica 5% da variância e a compreensão verbal (V) que contribui apenas com 3% da variância (Gomes & Borges, 2009).

Bueno (2013) realizou um estudo de validade de construto com a prova do Enem, sendo que a amostra foi composta por 63.236 estudantes do Estado de São Paulo que fizeram o exame de 2007. Inicialmente, a autora realizou uma análise de conteúdo da prova. As questões foram lidas e classificadas em quatro habilidades (*Gc* – Inteligência Cristalizada, *Gq* – Conhecimento Quantitativo, *Gf* - Inteligência Fluida e *Grw* - leitura e escrita) do CHC que estão relacionadas a escolarização e se encaixavam em cada item. Os itens agruparam-se em 3 grupos: Grupo A) 48 itens que tinham mais características de *Grw* e *Gf*; Grupo B) 6 itens com características de *Gq* e *Grw*; e Grupo C) 9 itens voltados a uma maior capacidade de *Gc*, *Grw* e *Gf*. A autora sugere que a prova do Enem de 2007 é mais voltada para o raciocínio e resolução de problemas.

Além disso, Bueno (2013) realizou 10 análises fatoriais com rotação *PROMAX* até encontrar o modelo mais adequado que foi o de 3 fatores. O fator 1 é visto como mais geral e abrange itens que avaliam *Grw* e *Gf* do grupo A, *Gq* do grupo B e *Gc* do grupo C. O

segundo fator abrange itens do grupo A (*Grw* e *Gf*) e o terceiro fator também abrange itens do grupo A (*Grw* e *Gf*), apenas um item avalia *Gq*. A pesquisadora compara seus achados com o de Gomes (2005) e conclui que no seu estudo não foi encontrado um fator que avalie o conhecimento quantitativo e que os fatores 2 e 3 estão relacionados a competência verbal e o fator 1 abrange uma competência mais geral.

O último estudo de evidência de validade fatorial com o Enem é o de Muner (2013) sendo o primeiro publicado com o novo modelo teórico do Enem. A autora utilizou a prova do ano de 2010 para realização da AFE (Análise Fatorial Exploratória) e a AFC (Análise Fatorial Confirmatória). Foram realizadas AFEs com rotação *Varimax* para cada uma das áreas de conhecimento. Itens que possuíam carga fatorial menor que 0,30 era retirados e uma nova AFE era realizada. Após esse procedimento, a prova de Ciências da Natureza ficou composta por 3 fatores (Fator 1= *Gc*; Fator 2= *Gq*; Fator 3=*Grw*), Ciências Humanas (Fator 1= *Grw*; Fator 2= *Gc*) e a de Matemática e suas tecnologias por dois fatores (Fator 1=*Gq* e Fator 2= *Gf*) e Linguagens e códigos por 1 fator (*Grw*). As AFCs foram realizadas apenas com os itens que sobraram das AFEs e indicaram bons índices de ajuste para a prova do Enem de 2010, mostrando que essa avalia tanto habilidades cognitivas quanto acadêmicas e que um número inferior de questões seria capaz de mensurar o que foi encontrado.

O presente estudo e os quatro encontrados na literatura buscaram verificar o que o Enem mede, sendo que todos partiram de uma visão psicométrica da inteligência. Compreende-se, por tanto, que os responsáveis pela prova do Enem e os pesquisadores que investigaram o que essa prova avalia possuem formas diferente de olhar para o exame. Partindo desse entendimento, tem-se que: 1) o estudo de Gomes (2005; Gomes & Borges, 2009) não utilizou o banco de dados do Enem, pois a prova do Enem foi aplicada a um

grupo de estudantes, que cursavam o terceiro ano do Ensino Médio, dentro de um projeto maior que envolvia a aplicação de outros testes; 2) os demais estudos (Bueno, 2013; Muner, 2009) utilizaram o banco de dados do Enem e seguiram uma metodologia na qual foram selecionados itens que mensuram de maneira mais pura cada fator e no presente estudo foi utilizado o banco de dados do exame e método de análise que avalia todos os itens da prova, ou seja, a prova de maneira global.

## **2. MÉTODO**

### ***2.1. Fonte de Dados***

O banco de dados do Enem de 2006 é composto por 2.784,192 participantes, sendo, 62,2% do sexo feminino, 32,1% possuíam renda de até R\$ 700,00 seguido de 26,7% com renda de até R\$ 1750,00. Em uma escala de 0 a 100, a nota média no Enem foi de 36,90 (DP=13,3). Do total de participantes em 2006, foram selecionados 666,042 sujeitos que responderam a prova amarela para realização das análises com os itens da prova.

### ***2.2. Instrumento***

*Exame Nacional do Ensino Médio de 2006* - A prova possui 63 questões de múltipla escolha e quatro tipos de provas (amarela, branca, rosa e azul), o que muda em cada uma é a ordem de apresentação dos itens na prova. O exame teve duração de cinco horas e possui questões de Português, Geografia, História, Biologia, Matemática, Física, Química e Redação. Os itens foram construídos a partir de uma matriz de referência que inclui 21 habilidades derivadas de 5 competências, entendidas pelo MEC como básicas para os alunos concluintes do Ensino Médio.

### ***2.3. Procedimentos***

Inicialmente, foi selecionado o banco de dados do Enem de 2006 disponível no site do Inep (<http://portal.inep.gov.br/basica-levantamentos-acessar>). Foram realizadas



estatísticas descritivas para melhor caracterização dos participantes que realizaram o Enem nesse ano. Posteriormente, foi escolhida aleatoriamente a prova amarela para realizar uma análise fatorial por informação completa (*full information factor analysis – FIFA*) dos itens e assim avaliar a dimensionalidade da prova. Tal análise é baseada na Teoria de Resposta ao Item (TRI) (Bock, Gibbons & Muraik, 1986; Primi & Almeida, 1998). Uma análise teórica dos itens foi realizada e os achados comparados com o da análise fatorial. Por fim, foi realizada uma análise do mapa de itens da prova via TRI. Tais análises foram realizadas no programa SPSS versão 19 e *TestFact*.

### 3. Resultados e Discussão

Os resultados da extração dos fatores principais da análise fatorial exploratória dos 63 itens do Enem são apresentados na Figura 2. Verifica-se por meio do *scree plot* que há claramente um fator definido que explicou 20,6% da variância e um segundo fator com 2,9% da variância ( $Eigenvalue1/Eigenvalue2 = 4,54$ ). A correlação entre os fatores (rotação *Promax*) foi de 0,67.

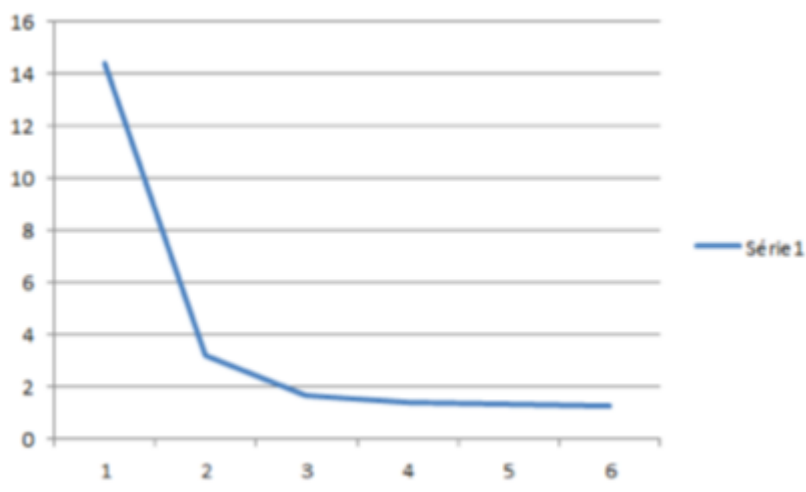


Figura 2. Visualização dos *eigenvalues* dos fatores extraídos na análise fatorial

Na tabela 1 pode-se visualizar o resultado da análise fatorial. Foi realizada também uma análise teórica de todos os itens. As questões foram lidas e verificado se possuíam

mais características do fator 1 ou do fator 2 resultante da análise fatorial, constando-se que há uma convergência entre as duas análises realizadas (teórica e empírica).

Tabela 1

*Solução de fatores, cargas fatoriais dos itens da prova amarela do Enem de 2006*

Item	Análise Fatorial	
	F1	F2
Q62	0.822	-0.058
Q10	0.788	-0.006
Q54	0.771	0.049
Q46	0.631	0.140
Q19	0.621	0.149
Q31	0.589	-0.066
Q59	0.544	-0.381
Q53	0.520	-0.042
Q29	0.510	0.138
Q20	0.472	0.206
Q49	0.447	0.156
Q51	0.405	0.257
Q23	0.396	0.134
Q28	0.394	-0.391
Q56	0.387	-0.341
Q63	0.356	0.209
Q37	0.342	0.181
Q58	0.329	0.276
Q60	0.328	-0.032
Q26	0.313	0.177
Q27	0.305	0.176
Q48	0.299	0.168
Q57	0.294	0.212
Q45	0.287	0.245
Q36	0.255	<b>0.489</b>
Q09	0.238	0.259
Q50	0.216	<b>0.464</b>
Q42	0.206	0.295
Q40	0.190	0.102
Q61	0.186	0.090
Q05	0.154	0.172
Q11	0.131	<b>0.627</b>
Q24	0.120	-0.183
Q44	0.119	<b>0.514</b>
Q12	0.107	0.264
Q01	0.101	0.287
Q55	0.100	0.170
Q25	0.098	<b>0.345</b>
Q08	0.097	0.252

Tabela 1

Continuação

Itens	Análise Fatorial	
	Fator 1	Fator 2
Q02	0.089	0.320
Q16	0.084	0.473
Q47	0.072	0.562
Q03	0.057	0.067
Q33	0.051	0.296
Q22	0.029	0.364
Q18	0.023	0.434
Q13	0.013	0.314
Q52	0.011	0.230
Q15	0.003	0.616
Q06	0.002	0.239
Q34	-0.006	0.032
Q21	-0.008	0.591
Q07	-0.013	0.399
Q38	-0.023	0.658
Q04	-0.027	0.269
Q39	-0.037	0.445
Q32	-0.039	0.406
Q17	-0.057	0.599
Q41	-0.061	0.576
Q30	-0.082	0.578
Q35	-0.118	0.314
Q14	-0.134	0.330

No Fator 1 foram agrupados 21 itens com cargas fatoriais variando de 0,30 a 0,76. Os itens 62 e 10, por exemplo, fazem parte desse fator (ver Figura 3). Para resolver a questão 62 o estudante tem que lembrar que é necessário utilizar uma fórmula para chegar a resposta. Assim, é preciso tanto um conhecimento quantitativo (Gq), quanto alguns fatores ligados ao raciocínio fluído, utilizar o raciocínio sequencial geral (RG) e quantitativo (RV) para resolução do problema. Na questão 10 é preciso descobrir implicações subjacentes a pergunta ou ligada a combinação lógica entre elas para se chegar a resposta. Assim, nesse item estão envolvidos o raciocínio fluído, requerendo especialmente o raciocínio sequencial geral (RG).

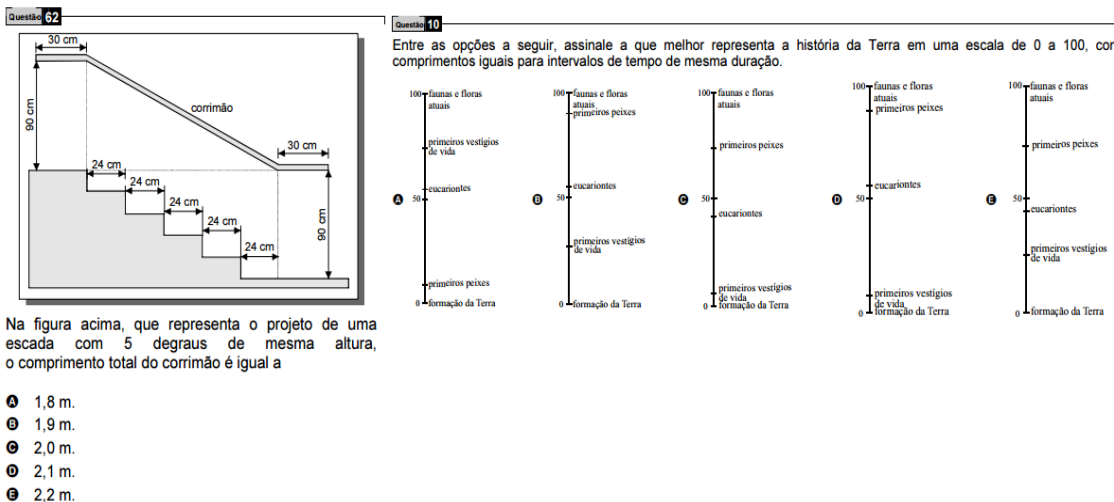


Figura 3. Exemplos de Itens do fator 1 da prova amarela do Enem de 2006

No fator 2 as cargas fatorias variaram de 0,30 a 0,59 em 22 itens (Tabela 1). Dois itens desse fator podem ser visualizados na Figura 4. Para chegar a resposta é necessário a capacidade de leitura e escrita (Grw), especialmente, a decodificação da leitura (RD), compreensão da leitura (RC) e compressão da linguagem verbal (V). Como a compreensão de textos e a leitura estão muito envolvidas na resolução dessas questões, consequentemente, é importante o conhecimento cristalizado (Gc), envolvendo informação geral (KO), informação sobre a cultura (K2) e conhecimento léxico (VL).

Embora a metodologia utilizada no presente estudo seja diferente de outros estudos que também verificaram a estrutura interna da prova, observa-se semelhança nos resultados especialmente no que se refere a encontrar características de Grw, Gc, Gf e Gq (Bueno, 2013; Muner, 2013) nos itens do Enem. Nos dois fatores parece ser necessário em alguma intensidade a inteligência fluida, pois para resolução dos itens é necessário a recombinação de conhecimentos adquiridos. A influência da inteligência fluida nos itens da prova do Enem também foi encontrada no estudo de Gomes (2005), evidenciando, desse modo, que a correlação moderada dos dois fatores encontrada no presente estudo refere-se a inteligência fluida. Contudo, não se pode afirmar que o Enem avalie mais inteligência fluida.

Se comparado aos vestibulares tradicionais, o Enem, realmente, avalia mais inteligência fluida (raciocínio) (Primi et al. 2001), pois as perguntas nos vestibulares tradicionais são diretas ou o aluno tem o conhecimento para responder ou não conseguirá acertar os itens. No Enem, em muitas questões, são fornecidas informações no próprio item para se chegar a resposta. Assim, é necessária leitura e a capacidade de raciocínio para ativar o conhecimento que se tem e chegar a resposta. Ademais, Gc é efeito de Gf, quando se mede Gc se avalia Gf também, pois os dois fatores são altamente correlacionados (Carroll,1993), dessa forma, não se pode afirmar simplesmente que o Enem mensura mais Gf visto que Gc também faz parte dessa habilidade.

<p><b>Questão 38</b></p> <p>À produção industrial de celulose e de papel estão associados alguns problemas ambientais. Um exemplo são os odores característicos dos compostos voláteis de enxofre (mercaptanas) que se formam durante a remoção da lignina da principal matéria-prima para a obtenção industrial das fibras celulósicas que formam o papel: a madeira. É nos estágios de branqueamento que se encontra um dos principais problemas ambientais causados pelas indústrias de celulose. Reagentes como cloro e hipoclorito de sódio reagem com a lignina residual, levando à formação de compostos organoclorados. Esses compostos, presentes na água industrial, despejada em grande quantidade nos rios pelas indústrias de papel, não são biodegradáveis e acumulam-se nos tecidos vegetais e animais, podendo levar a alterações genéticas.</p> <p><small>Celênia P. Santos et al. <i>Papel: como se fabrica? In: Química nova na escola</i>, n.º 14, nov./2001, p. 3-7 (com adaptações).</small></p> <p>Para se diminuir os problemas ambientais decorrentes da fabricação do papel, é recomendável</p> <p><b>A</b> a criação de legislação mais branda, a fim de favorecer a fabricação de papel biodegradável.</p> <p><b>B</b> a diminuição das áreas de reflorestamento, com o intuito de reduzir o volume de madeira utilizado na obtenção de fibras celulósicas.</p> <p><b>C</b> a distribuição de equipamentos de desodorização à população que vive nas adjacências de indústrias de produção de papel.</p> <p><b>D</b> o tratamento da água industrial, antes de retorná-la aos cursos d'água, com o objetivo de promover a degradação dos compostos orgânicos solúveis.</p> <p><b>E</b> o recolhimento, por parte das famílias que habitam as regiões circunvizinhas, dos resíduos sólidos gerados pela indústria de papel, em um processo de coleta seletiva de lixo.</p>	<p><b>Questão 15</b></p> <p>No princípio do século XVII, era bem insignificante e quase miserável a Vila de São Paulo. João de Laet dava-lhe 200 habitantes, entre portugueses e mestiços, em 100 casas; a Câmara, em 1606, informava que eram 190 os moradores, dos quais 65 andavam homiziados*.</p> <p><small>*homiziados: escondidos da justiça</small></p> <p><small>Nelson Werneck Sodré. <i>Formação histórica do Brasil</i>. São Paulo: Brasiliense, 1964.</small></p> <p>Na época da invasão holandesa, Olinda era a capital e a cidade mais rica de Pernambuco. Cerca de 10% da população, calculada em aproximadamente 2.000 pessoas, dedicavam-se ao comércio, com o qual muita gente fazia fortuna. Cronistas da época afirmavam que os habitantes ricos de Olinda viviam no maior luxo.</p> <p><small>Hildegard Féist. <i>Pequena história do Brasil holandês</i>. São Paulo: Moderna, 1998 (com adaptações).</small></p> <p>Os textos acima retratam, respectivamente, São Paulo e Olinda no início do século XVII, quando Olinda era maior e mais rica. São Paulo é, atualmente, a maior metrópole brasileira e uma das maiores do planeta. Essa mudança deveu-se, essencialmente, ao seguinte fator econômico:</p> <p><b>A</b> maior desenvolvimento do cultivo da cana-de-açúcar no planalto de Piratininga do que na Zona da Mata Nordestina.</p> <p><b>B</b> atraso no desenvolvimento econômico da região de Olinda e Recife, associado à escravidão, inexistente em São Paulo.</p> <p><b>C</b> avanço da construção naval em São Paulo, favorecido pelo comércio dessa cidade com as Índias.</p> <p><b>D</b> desenvolvimento sucessivo da economia mineradora, cafeicultora e industrial no Sudeste.</p> <p><b>E</b> destruição do sistema produtivo de algodão em Pernambuco quando da ocupação holandesa.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 4. Exemplos de Itens do fator 2 da prova amarela do Enem de 2006

Vinte itens não se destacaram em nenhum dos dois fatores, mas por meio da análise teórica nota-se que esses possuem algumas características principalmente do fator 2. Na figura 5, há dois exemplos. Observa-se que no item 33 para resolvê-lo é necessário ler o

enunciado e utilizar as habilidades de compreensão e leitura (Grw), fluida (Gf) e conhecimento cristalizado (Gc) para conseguir chegar a resposta que corresponde a letra D. Na questão 42, embora tenha uma tabela com informações para ajudar na resolução da questão, grande parte da informação necessária para responder o item está nas alternativas de resposta que por eliminação faz com que o aluno chegue a resposta correta, nesse caso a letra A. Assim, para resolução do item 42 também é necessário a compreensão e leitura (Grw), a inteligência fluida (Gf) e o conhecimento cristalizado (Gc).

**Questão 33**

As florestas tropicais úmidas contribuem muito para a manutenção da vida no planeta, por meio do chamado sequestro de carbono atmosférico. Resultados de observações sucessivas, nas últimas décadas, indicam que a floresta amazônica é capaz de absorver até 300 milhões de toneladas de carbono por ano. Conclui-se, portanto, que as florestas exercem importante papel no controle

- A das chuvas ácidas, que decorrem da liberação, na atmosfera, do dióxido de carbono resultante dos desmatamentos por queimadas.
- B das inversões térmicas, causadas pelo acúmulo de dióxido de carbono resultante da não-dispersão dos poluentes para as regiões mais altas da atmosfera.
- C da destruição da camada de ozônio, causada pela liberação, na atmosfera, do dióxido de carbono contido nos gases do grupo dos clorofluorcarbonos.
- D do efeito estufa provocado pelo acúmulo de carbono na atmosfera, resultante da queima de combustíveis fósseis, como carvão mineral e petróleo.
- E da eutrofização das águas, decorrente da dissolução, nos rios, do excesso de dióxido de carbono presente na atmosfera.

**Questão 42**

A tabela a seguir apresenta dados relativos a cinco países.

país	saneamento básico (%)		taxa de mortalidade infantil (por mil)		
	esgotamento sanitário adequado	abastecimento de água	anos de permanência das mães na escola		
			até 3	de 4 a 7	8 ou mais
I	33	47	45,1	29,6	21,4
II	36	65	70,3	41,2	28,0
III	81	88	34,8	27,4	17,7
IV	62	79	33,9	22,5	16,4
V	40	73	37,9	25,1	19,3

Com base nessas informações, infere-se que

- A a educação tem relação direta com a saúde, visto que é menor a mortalidade de filhos cujas mães possuem maior nível de escolaridade, mesmo em países onde o saneamento básico é precário.
- B o nível de escolaridade das mães tem influência na saúde dos filhos, desde que, no país em que eles residam, o abastecimento de água favoreça, pelo menos, 50% da população.
- C a intensificação da educação de jovens e adultos e a ampliação do saneamento básico são medidas suficientes para se reduzir a zero a mortalidade infantil.
- D mais crianças são acometidas pela diarreia no país III do que no país II.
- E a taxa de mortalidade infantil é diretamente proporcional ao nível de escolaridade das mães e independe das condições sanitárias básicas.

Figura 5. Exemplo de item que não se destacou em nenhum fator da análise fatorial

Por meio do mapa de itens do Enem 2006 (Figura 6) via TRI visualiza-se que a amostra pode ser dividida em quatro grupos em relação a dificuldade do item. O primeiro grupo compreende os participantes que estão na média e conseguem acertar os itens mais fáceis, sendo que esses pertencem ao fator 2. O segundo grupo consegue acertar itens com dificuldade um pouco maior, mas continuam sendo itens pertencentes ao fator 2. O terceiro grupo abrange questões do Fator 2, além de itens que não se destacaram em nenhum dos

dois fatores e começa a aparecer alguns itens do fator 1. O quarto grupo compreende apenas os itens mais difíceis, sendo todos do fator 1.

Desse modo, para além do olhar normatizado dos resultados é importante verificar qualitativamente que há uma hierarquia de complexidade entre os itens. Os que são associados a Gf, Gq e Gv (fator 1) são mais complexos dos que os que envolve Grw (Fator 2). Com essa análise, pode-se verificar que o padrão de pontuação em determinados tipos de itens requer um determinado nível de habilidade do sujeito (menor habilidade tende acertar mais os itens do fator 2 e quanto maior habilidade há uma tendência a acertar itens do fator 2 e fator 1). Assim, pode-se estabelecer uma hierarquia dos itens tanto em relação ao tipo (conteúdo) do item, quanto em relação ao nível de habilidade do respondente.

Por meio do modelo de Rasch foram verificados os índices de ajuste dos itens (*infit* e *outfit*). Os valores de *infit* indicam se há discrepâncias próximas do nível de habilidade do sujeito, por exemplo, quando esse acerta/adere ou não um item próximo do seu nível de habilidade. Na prova amarela do Enem de 2006, os valores de *infit* variaram de 0,87 a 1,15. O *outfit* refere-se a uma discrepância distante do nível de habilidade do sujeito, ou seja, a pessoa não acerta/adere itens muito fáceis para sua habilidade ou o contrário, acerta/adere itens muito difíceis para o seu nível de habilidade (Baker, 2001), os valores de *outfit* variaram 0,85 a 1,28. Desse modo, os índices de ajuste dos itens do Enem de 2006 se mostraram adequados pois variaram de 0,70 a 1,30 (Linacre & Wright, 1994).

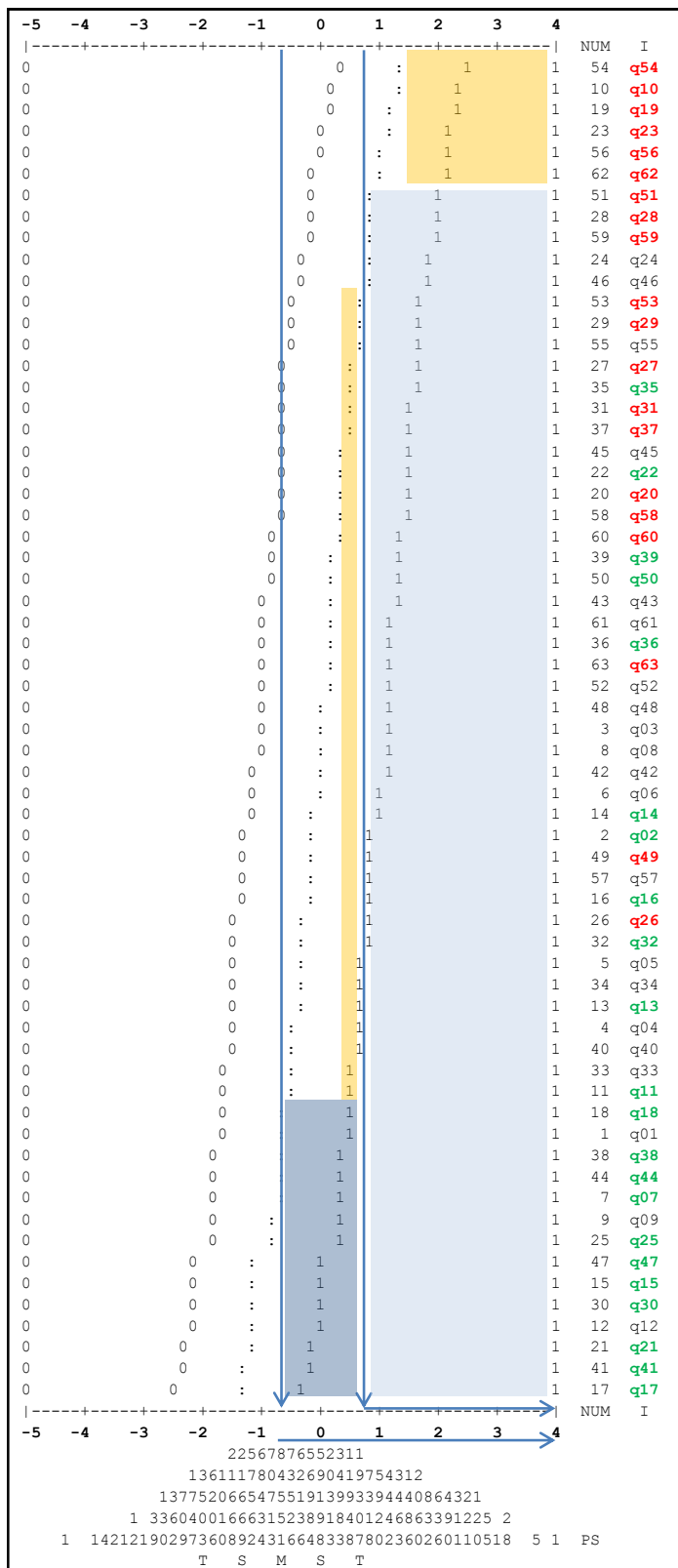


Figura 6. Mapa de itens da prova amarela do Enem 2006



#### **4. Considerações Finais**

O presente estudo buscou verificar o que o Enem avalia a partir de um olhar psicométrico da inteligência. Constata-se que os itens da prova podem ser compreendidos por meio de algumas habilidades do modelo CHC e correspondem, principalmente, a itens compostos por características de Gc, Gq, Gf, Gv e Grw.

Para além da verificação do que o Enem avalia, esse estudo evidenciou a necessidade de realizar novos estudos com essa prova, tanto do ponto de vista psicométrico quanto em relação a etapa de vida em que essa avaliação é realizada, pois o número de pesquisas nessa etapa da escolaridade é o menor em relação aos outros níveis de ensino. Esse estudo contribui para o entendimento do que as questões do Enem medem a partir de uma visão psicométrica e é o primeiro que utilizou os dados de todos os estudantes que realizaram o exame, ou seja, que não realizou recortes (apenas dados do Estado de São Paulo) para compor uma amostra como no caso de Bueno (2013) e Muner (2013) ou que aplicou a prova posteriormente a estudantes do terceiro ano do Ensino Médio (Gomes, 2005). No entanto, deve-se pensar em novos estudos que procurem, por exemplo, comparar a estrutura do antigo modelo e do novo modelo teórico do Enem.

## 5. Referências

- Brasil (2008). Ministério da Educação. *PDE: Plano de Desenvolvimento da Educação: SAEB: ensino médio: matrizes de referência, tópicos e descritores*. Brasília: MEC, SEB; Inep. 127 p.
- Baker, F. B. (2001). *The basics of Item Response Theory*. Washington, DC: ERIC.
- Bock, D., Gibbons, R. & Muraki, E. (1986). Full-information item factor analysis. *Applied Psychological Measurement*, 12(3), 261-280.
- Bueno, J. M. P. (2013). *Análise das habilidades avaliadas pelo Enem por meio da teoria CHC: um estudo de validade*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153–193.
- Cattell, R.B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312.
- Gomes, C. M. A. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio*. Tese de doutorado, Programa de Pós-Graduação em Educação: Conhecimento e Inclusão Social, da Faculdade de Educação da Universidade Federal de Minas Gerais, Belo Horizonte.
- Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, 6, n. 2, p. 167-179.

- Gomes, C. M. A., & Borges, O. N. (2009). O Enem é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, 20(42), 73-88.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- Fini, M. E. (2005). Erros e acertos na elaboração de itens para a prova do Enem. In: BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica*. Brasília, 101-106.
- Inep (2001). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. – Brasília:
- Linacre, J. & Wright, B. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(2), 370.
- Macedo, L. (2005). Competências e habilidades: Elementos para uma reflexão pedagógica. Ministério da Educação. *Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica*. Brasília: Inep.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp.151–180). New York, NY: Guilford Press.

- McGrew, K.S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10.
- McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.
- Ministério da Educação - MEC (2009). Matriz de Referência para o Enem 2009. Recuperado em: 10 de Julho de 2012. Disponível em: [http://portal.mec.gov.br/index.php?Itemid=310&id=13318&option=com\\_content&view=article](http://portal.mec.gov.br/index.php?Itemid=310&id=13318&option=com_content&view=article)
- Muner, L. C. (2013). *Análise fatorial exploratória e confirmatória do Enem 2010 com estudantes paulistas*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Primi, R. (2003). Inteligência: avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação Psicológica*, 2(1), 67-77.
- Primi, R. & Almeida, L. S. (1998) Considerações sobre a análise factorial de itens com resposta dicotômica. *Psicologia: Teoria, Investigação e Prática*, 3, 225-234.
- Primi, R., Muniz, M., & Nunes, C. H. S. (2009). Definições Contemporâneas de Validade de Testes Psicológicos. Em: Cláudio Simon Hutz (Org.). *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo, 2009, 1, 243-265.
- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. de F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(2), 151-159.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.

### ESTUDO III

TÍTULO EM PORTUGUÊS: **O ENEM CONSEGUE PREDIZER O DESEMPENHO NO ENADE?**

**EVIDÊNCIAS DE VALIDADE PREDITIVA UTILIZANDO UMA ABORDAGEM MULTINÍVEL**

TÍTULO EM INGLÊS: **THE ENEM CAN PREDICT THE PERFORMANCE IN ENADE? EVIDENCES**

**OF PREDICTIVE VALIDITY USING A MULTILEVEL APPROACH**

#### **Resumo**

Os estudos de validade preditiva são muito valiosos do ponto de vista econômico e prático. No Brasil há a disponibilidade de uma avaliação padronizada ao final do ensino médio (Enem) e ao final do ensino superior (Enade), com isso, buscou-se no estudo verificar a validade preditiva do Enem utilizando um delineamento longitudinal com o Enade. Foi selecionado o banco de dados do Enade de 2012 e resgatadas as informações desses estudantes quando fizeram a prova do Enem. Assim, 4 áreas (Administração, Ciências Contábeis, Direito e Psicologia) foram selecionadas em função de possuírem maior número de estudantes com notas nos dois exames. Foram elaborados 3 modelos de análise multinível e os resultados indicam que o Enem consegue prever o desempenho do aluno no Enade. A redação adiciona menos informação que a parte objetiva da prova e não há mudança em função dos cursos.

Palavras-chave: análise multivariada; avaliação educacional; validade do teste

#### **Abstract**

The predictive validity studies are very valuable from an economic and practical point of view. Once in Brazil there is the availability of a standardized assessment at the end of high school (Enem) and at the end of higher education (Enade), was sought with the study to verify the predictive validity of Enem using a longitudinal lineation with Enade. Enade's database from 2012 was selected and the information about these students was rescued when they took the Enem exam. Therefore, four courses (Administration, Accounting, Law and Psychology) were selected based on having more students with grades in both exams. Three models of multilevel analysis were developed and the results indicate that the Enem can predict student's performance in Enade. The essay adds less information than the objective part of the test and there is no change in function of the courses.

Keywords: multivariate analysis, educational evaluation, test validity

O Exame Nacional do Ensino Médio (Enem) foi criado em 1998 (Inep, 2013) e teve em sua primeira edição 157,2 mil inscritos e 115,6 mil participantes. A mais recente avaliação (2014) contou com 8.721,946 inscritos, 21,6% a mais que o ano anterior (Inep, 2014). Os dados demonstram o crescimento expressivo que o Enem conquistou ao longo de 15 anos de existência e isso pode ser atribuído as diferentes formas de utilização do desempenho do aluno no exame.

Atualmente, a nota do Enem pode ser utilizada: 1) para obtenção de certificação do ensino médio - isto implica dizer que a escola deixa de ser o único meio para se alcançar um título educativo; 2) como forma de entrada no ensino superior, pois a cada ano cresce o número de universidades que deixam de realizar os vestibulares tradicionais e passam a utilizar o desempenho no Enem; 3) para acesso a programas do Governo Federal, como o ciências sem fronteiras, bolsas de estudo do PROUNI (Programa Universidade para Todos) e financiamento estudantil do FIES (Fundo de Financiamento Estudantil) e; 4) como indicador da qualidade escolar (Inep, 2013; Zoghbi, Oliva, & Moriconi, 2010) sendo este último passível de questionamentos visto que a qualidade escolar não se restringe ao desempenho dos estudantes.

Apesar do crescimento, seu desenvolvimento vem sendo permeado de problemas. Isso pode ser observado em *sites* de notícia como os indicados nas referências desse artigo e que serviram de base para listagem dos acontecimentos a seguir. Encontra-se o caso dos estudantes que realizaram o Enem em 2007, 2008 e 2009 e que tiveram todos os seus dados disponibilizados a quem acessasse o site do Inep. Em 2009, a prova foi roubada o que obrigou o exame ser adiado para elaboração de um novo. Em decorrência disso, algumas universidades não utilizaram o desempenho no Enem em seus vestibulares naquele ano. Outros problemas incluem provas que foram impressas em determinada ordem e o gabarito

estava invertido, candidatos que conseguiram enviar informação da prova por mensagem de celular, além de estudantes que tiveram que fazer a prova em locais distantes de suas residências.

Além disso, a redação é a parte da prova que mais tem recebido críticas, principalmente, pelos casos de estudantes que escreveram hino de futebol e receita culinária e não tiveram suas redações anuladas. Somado a isso, em 2010 uma fiscal violou o pacote de provas e repassou o tema da redação para o filho e no ano de 2014, um estudante, pouco antes do início da prova, recebeu no celular uma foto da prova com o tema da redação e fez a denúncia.

Diante dos fatos listados acima, observa-se que os problemas divulgados nos meios de comunicação e que conseqüentemente chegam à população, correspondem a falha humana. Não obstante, a literatura científica também indica problemas, quanto à proposta teórica da prova, uma vez que não há suporte empírico nos manuais do Enem, como também nos estudos encontrados na literatura (Bartholomeu, 2011; Bueno, 2013; Gomes, 2005; Gomes & Borges 2007; 2009; Primi et al., 2001). Apesar disso, a utilização de uma avaliação em larga escala como o Enem é fundamental para verificar não apenas o perfil do egresso, mas visualizar como anda a educação nesse nível de ensino, traçar metas e gerar políticas públicas de melhoria para Educação.

O Enem, ao contrário dos vestibulares tradicionais que enfatizam mais a memorização dos conteúdos aprendidos, tem como foco o raciocínio e nesse sentido se aproxima de outras provas como o SAT (Scholastic Assessment Test), bastante conhecido e utilizado como forma de entrada nas universidades americanas. O SAT avalia além do conhecimento dos alunos no ensino médio, habilidades de raciocínio em testes de leitura, escrita e matemática (ver <http://sat.collegeboard.org/about-tests/sat>). A proposta é que no

exame sejam avaliadas habilidades importantes para um bom desempenho na universidade e na vida, tendo estudos demonstrado a capacidade dessa prova em prever desempenho no Ensino Superior (Ewing, Huff, Andrews, & King, 2005; Patterson & Mattern, 2011, 2012).

Em alguns estudos de predição com o SAT, destaca-se que esse exame correlaciona-se moderadamente com as notas na faculdade ( $\beta=0,43$ ), fortemente com o fator  $g$  ( $\beta=0,78$ ) e os chamados resíduos, ou seja, o que não corresponde ao fator  $g$  na prova, como subtestes de matemática e habilidade verbal, conseguem prever praticamente o mesmo ( $\beta=0,29$ ) que o fator  $g$  isolado ( $\beta=0,32$ ) no primeiro ano do curso universitário. Isso pode ocorrer uma vez que habilidades verbais e numéricas, mensuradas no SAT, também são importantes para o desempenho acadêmico (Coyle & Pillow, 2008).

Coely, Synder, Richmond, e Little (2015) relacionaram o desempenho no SAT, controlando os resíduos, ou seja, sem influência do fator  $g$  e mantendo a influência do fator  $g$  nessa prova, e o desempenho dos estudantes no primeiro ano universitário. Pela primeira vez em uma pesquisa esse efeito em cursos de áreas distintas foi avaliado. Os resultados demonstraram que os subtestes do SAT são associados a  $g$  ( $M r=0,84$ ;  $p\leq 0,01$ ) e ao desempenho na universidade ( $M r=0,46$   $p\leq 0,01$ ) e predizem melhor o desempenho em cursos de ciências naturais (exemplo: biologia, química) ( $M r= 0,34$ ;  $p\leq 0,01$ ) do que o desempenho em cursos de áreas verbais (literatura, inglês) e de matemática ( $M r= 0,28$ ;  $p\leq 0,01$ ). Após remover o fator  $g$ , verificou-se que os resíduos do SAT em matemática estão correlacionados ao desempenho acadêmico na área de matemática ( $r=0,11$ ;  $p\leq 0,01$ ) e negativamente com cursos de áreas verbais ( $r=-0,06$ ;  $p\leq 0,01$ ). O desempenho no SAT, tanto em leitura quanto em escrita, como era esperado, são relacionados negativamente com o desempenho na área de matemática e apenas os resíduos do SAT na área de escrita correlaciona-se com a verbal. Os resultados do estudo indicam que mesmo considerando



apenas os resíduos do SAT, o exame consegue prever o desempenho acadêmico na universidade em diferentes cursos, mesmo essa predição sendo pequena.

É importante enfatizar que os estudos de validade preditiva possuem um grande valor econômico e prático pois conseguem prever o que pode ocorrer no futuro, por exemplo, desempenho no trabalho ou aprendizagem futura (Schmidt & Hunter, 1998). Assim, e pensando na importância de verificar as propriedades psicométricas do Enem, a presente pesquisa teve como objetivo realizar um estudo de validade preditiva do exame. Para tanto, utilizou-se o Enade (Exame Nacional de Desempenho dos Estudantes) como critério externo. Tal exame mensura o desempenho dos estudantes universitários em relação ao que foi aprendido de conteúdo na universidade, sendo esse conteúdo correspondente aos previstos pelas diretrizes curriculares de cada curso de graduação (Inep, 2015).

Tanto o Enem quanto o Enade buscam mensurar o desempenho dos estudantes. Dessa forma, o objetivo das duas avaliações estão associados, permitindo a realização de estudos dessa natureza em que se relaciona os índices encontrados no teste com outras variáveis externas no intuito de encontrar validade para o teste (Primi, Muniz, & Nunes, 2009). Com isso, a pergunta central desse estudo é: O desempenho do aluno no Enem consegue prever qual será o desempenho do aluno posteriormente no Enade?

Foi realizada uma busca na BVS-PSI (Biblioteca Virtual em Saúde) e no Google Acadêmico (<https://scholar.google.com.br/>) para verificar se há estudos de evidências de validade preditiva com o Enem. Apenas dois estudos foram encontrados. Pansanato e Soares (1999) avaliaram a capacidade preditiva do resultado do Enem de 1998, em relação ao desempenho dos candidatos ao vestibular da UFMG de 1999. A correlação entre a nota do Enem e da primeira fase do vestibular da UFMG foi de 0,78. Posteriormente, foram comparados se a utilização da nota do Enem, substituindo a nota da primeira etapa do

vestibular causaria mudança na aprovação para segunda etapa do vestibular. Foi verificado que caso a nota fosse substituída pelo Enem a mudança no número de classificados para segunda fase seria mínimo, não causando impacto na nota final dos alunos. No outro estudo de evidências de validade preditiva, foi utilizada a nota no Enem de 2002, a nota no processo seletivo de entrada na universidade e as notas semestrais (2003, 2004 e 2005) de 129 estudantes universitários do curso de Psicologia. Foi encontrada uma correlação moderada e significativa ( $r=0,40$ ;  $p\leq 0,05$ ) entre o Enem e o desempenho dos alunos apenas no primeiro ano (2003) da faculdade. Contudo, em relação a redação e o desempenho dos alunos, foi encontrada correlação significativa no segundo ano (2004) ( $r=0,27$ ;  $p\leq 0,05$ ) e no terceiro ano (2005) ( $r=0,34$ ;  $p\leq 0,01$ ) do curso universitário. Por meio da análise de regressão linear observou-se que 25,5% da variância das notas dos alunos no primeiro ano de faculdade foi atribuído ao desempenho dos alunos no vestibular e 16,1% atribuído a nota no Enem, sendo que tanto a nota do vestibular quanto a do Enem foi contribuindo cada vez menos e de forma não significativa nos anos seguintes da faculdade. Os autores consideram as notas do Enem e do vestibular como boas preditoras do desempenho acadêmico especialmente no primeiro ano de curso (Souza, 2006; Souza, Vendramini, & Silva, 2013).

Tendo em vista a escassez de estudos de evidência de validade com o Enem e diante da disponibilidade dos dados de duas provas educacionais que os estudantes realizam em dois momentos de suas vidas, o que realmente caracteriza esse estudo como longitudinal, é que a pesquisa foi realizada. Para tanto, foi utilizada a análise multinível, ainda pouco utilizada no Brasil (Puente-Palácios & Laros, 2009), mas frequente na área de avaliação educacional (Andrade & Laros, 2007; Barbosa & Fernandes, 2000; Jesus & Laros, 2004; Valentini, 2013).

## **2. Método**

### ***2.1. Fonte de Dados***

Para compor a amostra de participantes do estudo, inicialmente, foi selecionado o banco de dados do Enade de 2012 e verificado quais os anos que mais se tinha informação do Enem desses participantes. Os anos de 2007, 2008 e 2009 foram os mais frequentes. Posteriormente, foi calculado o valor mínimo de alunos que cada instituição de ensino deveria ter para que essa entrasse na análise. Esse valor foi igual a 28 alunos, ou seja, cada curso universitário do Brasil que participou do Enade em 2012 deveria ter pelo menos 28 alunos com informação do Enem.

Conforme pode ser observado na Tabela 1 as áreas de Administração, Ciências Contábeis, Direito e Psicologia foram as selecionadas nesse estudo por possuírem maior quantidade de informação de Enade 2012 e Enem nos anos de 2007, 2008 ou 2009. Na área de Administração, 1546 instituições que oferecem o curso no Brasil participaram do Enade (2012) e o ano que mais possui informação de Enem nessa área é 2008 com 370 instituições com mais de 28 participantes cada uma contendo informação dos estudantes tanto no Enem quanto no Enade. Nesse sentido, para área de Administração e Ciências Contábeis foi selecionado o ano de 2008 e nas áreas de Direito e Psicologia o ano de 2007. Foram selecionadas apenas as quatro áreas que possuíam maior número de informação dos estudantes nos dois exames. Nesse sentido, outras áreas, que não foram selecionadas nesse estudo, possuíam maior informação no Enem de 2009, mas não foram incluídas nas análises por não atender a esse critério.

Tabela 1

*Número de instituições que participaram do Enade em 2012 por área e número de instituições com mais de 28 alunos com dados do Enem e Enade por ano*

	N Instituições Enade	Enem		
		2007	2008	2009
Administração	1546	259	370	66
Ciências Contábeis	860	62	130	13
Direito	960	319	71	14
Psicologia	371	72	9	2

O número de participantes em si não pode ser visualizado na Tabela 1, pois esse número variou em função do modelo de análise multinível utilizado. Assim, o número de estudantes de Administração flutuou de 24.938 a 48.464, os estudantes de Direito de 17.633 a 67.836. O curso de Psicologia contou com a participação de 2.477 a 7.905 estudantes e 4.564 a 18.929 estudantes de Ciências Contábeis.

## **2.2. Instrumentos**

***Exame Nacional do Ensino Médio*** - As provas do Enem dos anos de 2007 e 2008 correspondem ao modelo antigo do Enem composto por 63 questões de múltipla escolha e cinco tipos de provas, todas com os mesmos itens, o que muda é a ordem de apresentação dos itens na prova. O exame de 2007 e 2008 teve duração de cinco horas e possui questões de Português, Geografia, História, Biologia, Matemática, Física, Química e Redação. Os itens foram construídos a partir de uma matriz de referência com 21 habilidades derivadas de 5 competências, entendidas pelo MEC como básicas para os alunos concluintes do Ensino Médio.

***Exame Nacional de Desempenho dos Estudantes - Enade***: O Enade faz parte do Sistema Nacional de Avaliação da Educação Superior (SINAES) e tem como objetivo verificar a aprendizagem dos estudantes universitários (ingressantes e concluintes) em suas áreas de

formação. O exame é composto por 40 questões (10 de formação geral para todos os cursos e 30 específicas de cada área) discursivas e de múltipla escolha que pretendem verificar a capacidade do estudante analisar, sintetizar, criticar, deduzir, construir hipóteses, estabelecer relações, fazer comparações, detectar contradições, decidir e organizar as ideias.

### **2.3. Procedimentos**

#### **2.3.1. Banco de Dados**

O banco de dados utilizado para realização do estudo foi fornecido pelo Inep após análise do projeto de pesquisa e aprovação (Anexo I). O banco enviado contém os dados de todos os estudantes que fizeram o Enade em 2012 e os dados relativo ao Enem desses estudantes. Os anos com maior incidência de Enem foram 2007, 2008 e 2009. Por isso, esses foram os anos considerados para o cruzamento das informações. Como esperado, alguns estudantes que fizeram o Enade em 2012 não fizeram o Enem em nenhum ano ou fizeram em anos anteriores aos de maior frequência Assim, os dados dos estudantes sem informação do desempenho no Enem em um desses três anos foram excluídos. O banco de dados desse estudo possui um dado que será muito comentado ao longo do estudo, mas que pode gerar confusão para o leitor. Por isso, é importante esclarecer que o termo “área” refere-se a área de formação, ou seja, área de Administração, área de Ciências Contábeis, Direito e Psicologia e que cada área abrange n cursos, ou seja, o termo “curso” refere-se a instituição de ensino de uma determinada área. Por exemplo: Área - Psicologia; Cursos - Psicologia da USF (Itatiba-SP), Psicologia da USF (Campinas-SP), Psicologia da USP (São Paulo), Psicologia da UFPE, etc.

#### **2.3.2. Análise de Dados**

Foram realizadas estatísticas descritivas para selecionar e melhor caracterizar a amostra por meio do *Statistical Package for the Social Sciences* (SPSS), versão 19. Em

seguida, foram elaborados três modelos para realização da análise multinível que foi executada no programa *MPlus* (Muthén & Muthén, 2010).

Os modelos consideram dois níveis. Em uma dada área (Administração, Ciências Contábeis, Direito e Psicologia) os  $i$ 's estudantes concluintes estão alocados nos  $j$  cursos daquela área. Assim, o Modelo 0 (vazio) considera a nota no  $Enade_{ij}$  de um estudante  $i$  de um curso  $j$  e é prevista pela equação (Raudembush & Bryk, 2002):

$$Enade_{ij} = \beta_{0j} + e_{ij} \text{ (Nível 1)}$$

$$\beta_{0j} = \beta_0 + \mu_{0j} \text{ (Nível 2)}$$

Onde:

$\beta_{0j}$	Média dos concluintes do curso $j$ no Enade
$e_{ij}$	Desvio do desempenho do aluno $i$ em relação à média do curso $j$ (também chamado Resíduo do Nível 1.)
$\beta_0$	Média geral de todos os N concluintes no Enade
$\mu_{0j}$	Desvio da média do curso $j$ em relação à média global (também chamado resíduo do Nível 2.)
$\sigma_{\mu}^2$	Variância do parâmetro $\mu_j$ . Variância entre cursos. Dispersão do desempenho médio dos cursos
$\sigma_e^2$	Variância do parâmetro $e_{ij}$ . Variância entre os alunos dentro de seu curso. Dispersão do desempenho dos alunos dentro dos cursos
$ICC = \frac{\sigma_{\mu}^2}{\sigma_{\mu}^2 + \sigma_e^2}$	ICC Correlação intraclasse indicando a proporção de variância do desempenho que está associada aos cursos.

No Modelo 1 foram acrescentadas as notas dos estudantes na parte objetiva e redação do Enem (que responderam antes de entrar no curso) e foi estimado o coeficiente de validade preditiva do Enem para prever os resultados do Enade. Nesse modelo o coeficiente é um efeito fixo - idêntico para todos os cursos. O Modelo 1 é escrito:

$$Enade_{ij} = \beta_{0j} + \beta_1 Enem\_OBJ_{ij} + \beta_2 Enem\_RED_{ij} + e_{ij} \text{ (Nível 1)}$$

$$\beta_{0j} = \beta_0 + \mu_{0j} \text{ (Nível 2)}$$

Onde:

$\beta_{0j}$	Média dos concluintes do curso $j$ no Enade depois de controlado o efeito do Enem (capacidade prévia), isto é, média do curso considerando alunos com mesmo desempenho no Enem
$Enem\_OBJ_{ij}$	Nota do Enem na parte objetiva no ano anterior à entrada no curso
$Enem\_RED_{ij}$	Nota do Enem na redação no ano anterior à entrada no curso
$\beta_1$	Coefficiente não padronizado de validade preditiva da parte objetiva do Enem. Montante de aumento da nota no Enade para uma unidade de aumento na parte objetiva do Enem considerando mesmas notas na redação.
$\beta_2$	Coefficiente não padronizado de validade preditiva da redação do Enem. Montante de aumento da nota no Enade para uma unidade de aumento na nota da redação do Enem considerando mesmas notas na parte objetiva.

Os parâmetros restantes têm a mesma interpretação do que no Modelo 0. A diferença é que os resíduos no Nível 1 e 2 ( $\sigma_e^2$  e  $\sigma_\mu^2$ ) serão reduzidos proporcionalmente à magnitude do coeficiente de validade. Nesse modelo também foram calculados os coeficientes padronizados transformando em  $z$  as notas do Enade e Enem antes de estimar os parâmetros. Assim,  $\beta_1$  e  $\beta_2$  indicam quantas unidades de desvio padrão são aumentadas no Enade com o aumento de um desvio padrão no Enem deixando a outra nota constante (redação ou parte objetiva).

Por fim, foi testado um último modelo, Modelo 2, no qual os coeficientes de validade variaram em cada curso. Assim, os efeitos são agora randômicos. O modelo fica escrito:

$$Enade_{ij} = \beta_{0j} + \beta_{1j}Enem\_OBJ_{ij} + \beta_{2j}Enem\_RED_{ij} + e_{ij} \text{ (Nível 1)}$$

$$\beta_{0j} = \beta_0 + \mu_{0j} \text{ (Nível 2)}$$

$$\beta_{1j} = \beta_{10} + \mu_{1j} \text{ (Nível 2)}$$

$$\beta_{2j} = \beta_{20} + \mu_{2j} \text{ (Nível 2)}$$

Onde:

$\beta_{1j}$	Coeficiente não padronizado de validade preditiva da parte objetiva do Enem no curso $j$ .
$\beta_{2j}$	Coeficiente não padronizado de validade preditiva da redação do Enem no curso $j$ .
$\beta_{10}$	Média global dos coeficientes de validade da parte objetiva (sobre todos cursos)
$\beta_{20}$	Média global dos coeficientes de validade da redação (sobre todos cursos)
$\mu_{1j}$	Desvio do coeficiente de validade da parte objetiva no curso $j$ em relação a média global. Esse efeito é chamado o efeito de interação cruzada de níveis já que o curso é um moderador da relação ente Enem e Enade.
$\mu_{2j}$	Desvio do coeficiente de validade redação no curso $j$ em relação à média global.
$\sigma_{\mu 1}^2$ e $\sigma_{\mu 2}^2$	Variância dos parâmetros $\mu_{1j}$ e $\mu_{2j}$ . Dispersão da relação Enem X Enade nos cursos

A adequação dos modelos foram testados verificando-se os índices BIC (Critério de Informação Akaike) e AIC (Critério de Informação Baysiano). Quanto menor o valores desses índices melhor o ajuste do modelo.



### 3. Resultados

Para atingir os objetivos do estudo que compreendeu verificar se há validade preditiva para o Enem, foram realizados os procedimentos acima descritos. Conforme visualizado na Tabela 2, encontram-se os resultados das estimativas dos parâmetros e erros padrão para os três modelos estabelecidos na área de Psicologia. No modelo M0, sem nenhuma variável explicativa, é calculado apenas o intercepto que corresponde à média geral dos estudantes no Enade nos diferentes cursos. O valor do intercepto é maior na parte objetiva da prova do que na discursiva e demonstra que há diferença entre os cursos universitários de Psicologia pois a razão entre a variância e o erro padrão é superior ao valor de 1,96 (Laros & Marciano, 2008), na parte objetiva esse valor é de 4,76 e 3,88 na discursiva. O valor da variância entre cursos foi de 19,49 e 210,97 de responsabilidade da variância intra-curso. O valor da correlação intraclasse demonstra que 8% da variação das notas no Enade é devido ao curso de Psicologia no qual o aluno estuda.

No modelo 1 é calculado o intercepto, a variância dos resíduos no nível do Enem tanto para prova objetiva quanto discursiva (no caso do Enem a redação). O valor da correlação intraclasse no modelo 1 diminuiu em relação ao modelo 0 indicando que 4% da variação nas notas do Enade de Psicologia na parte objetiva e 2% da variação na parte discursiva decorre do desempenho no Enem. Nesse modelo foi estimado o coeficiente de validade preditiva da parte objetiva do Enem para prever os resultados do Enade e os resultados foram significativos. A validade para parte objetiva do Enem é de 0,40 e na redação cai para 0,25. O coeficiente de regressão para a prova do Enem objetiva é de 0,38, o que significa que a cada ponto mais alto no Enem, é esperado um aumento de 0,38 pontos no Enade. Na redação do Enem o coeficiente de regressão foi de 0,06, ou seja, a cada ponto mais alto na redação do Enem é esperado um aumento de 0,06 pontos no Enade.

O objetivo do modelo 2 foi deixar os coeficientes de validade variarem em cada curso. Nota-se que os valores do efeito fixo permanecem praticamente os mesmos que na parte objetiva do modelo 1. O que é acrescentado nesse modelo são as correlações entre os Enems (0,03 e 0,02). O modelo 2 pode ser considerado um pouco melhor, mas como a variabilidade é muito pequena, ele agrega muito pouca informação se comparado ao modelo 1.

Tabela 2

*Modelos da análise multinível do Enem para área de Psicologia*

	<b>M0objetiva</b>	<b>M0discursiva</b>	<b>M1objetiva</b>	<b>M1discursiva</b>	<b>M2objetiva</b>
<b>Efeito Fixo</b>					
Intercept	42,54 (0,55)	25,17 (0,42)	44,35 (0,46)	26,34 (0,40)	44,36
Enem07 Obj			0,38 (0,01)	0,26 (0,02)	0,39(0,01)
			[0,40]	[0,25]	
Enem07Red			0,06 (0,01)	0,06 (0,02)	0,06(0,03)
			[0,06]	[0,05]	
<b>Efeito Randômico</b>					
$\sigma_{\mu}^2$	19,49(4,09)	9,43 (2,43)	10,34 (2,53)	5,38 (1,86)	9,717
$\sigma_{\varepsilon}^2$	210,97	210,16	167,98	205,89	167,55
$\sigma_{\mu1}^2$					0,003
$\sigma_{\mu2}^2$					0,002
<b>Intraclasse</b>	0,08	0,04	0,04	0,02	0,04
<b>Loglikelihood</b>	-35219,805	-25605,050	-12093,375	-10134,368	-12093,160
<b>R square</b>			0,17	0,14	
<b>AIC</b>	70445,610	51216,100	24196,749	20278,736	24206,321
<b>BIC</b>	70466,741	51236,317	24226,826	20307,810	24266,474
<b>NAlunos</b>	7905	6243	3027	2477	3027
<b>NCursos</b>	72	72	72	72	72

Nota. O valor entre colchetes refere-se ao valor padronizado.

Os mesmos modelos foram utilizados nas áreas de Administração, Ciências Contábeis e Direito e os resultados são visualizados, respectivamente, nas tabelas 3, 4 e 5. A descrição detalhada dessas tabelas não serão apresentadas visto que os resultados foram similares a área de Psicologia em todas as áreas e modelos, mas destaca-se comparando-se os resultados de todas as áreas que em administração o coeficiente de validade preditiva foi o maior de todas as áreas, tanto na parte objetiva (0,56) do Enem como na redação (0,31).

Tabela 3

*Modelos da análise multinível do Enem para área de Administração*

	<b>M0objetiva</b>	<b>M0discursiva</b>	<b>M1objetiva</b>	<b>M1discursiva</b>	<b>M2objetiva</b>
<b>Efeito Fixo</b>					
Intercept	35,62 (0,29)	25,26 (0,25)	37,30 (0,22)	26,87 (0,20)	37,19 (0,21)
Enem08 Obj			0,56 (0,01)	0,31 (0,01)	0,27 (0,03)
Enem08Red			0,07 (0,01)	0,09 (0,01)	0,01 (0,03)
<b>Efeito Randômico</b>					
$\sigma_{\mu}^2$	30,42 (3,00)	20,59 (2,37)	14,27 (1,47)	9,08 (1,22)	12,84 (1,26)
$\sigma_{\varepsilon}^2$	234,20 (5,74)	215,77 (5,22)	186,51(3,99)	207,17 (4,58)	185,60 (3,96)
$\sigma_{\mu1}^2$					0,00 (0,02)
$\sigma_{\mu2}^2$					0,01 (0,01)
<b>Intraclasse</b>	0,11	0,08	0,05	0,03	0,05
<b>Loglikelihood</b>	-348881.429	-199416.850	-100880.627	-66391.053	-100825.397
<b>R square</b>			0,26	0,12	
<b>AIC</b>	697768.857	398839.700	201771.255	132792.105	201670.794
<b>BIC</b>	697796.873	398866.066	201811.876	132830.571	201752.035
<b>NAlunos</b>	83.993	48.464	67.400	16.207	67.400
<b>NCursos</b>	370	370	370	370	370

Tabela 4

*Modelos da análise multinível do Enem para área de Ciências Contábeis*

<b>Ciências Contábeis</b>	<b>M0objetiva</b>	<b>M0discussiva</b>	<b>M1objetiva</b>	<b>M1discussiva</b>	<b>M2objetiva</b>
<b>Efeito Fixo</b>					
Intercept	35,10 (0,45)	32,39 (0,36)	35,88 (0,41)	30,82 (0,19)	35,79 (0,41)
Enem08 Obj			0,40 (0,01)	0,15 (0,01)	0,37 (0,08)
Enem08Red			0,04 (0,01)	0,08 (0,01)	0,04 (0,09)
<b>Efeito Randômico</b>					
$\sigma_{\mu}^2$	23,91 (3,69)	14,12 (2,33)	17,73 (2,93)	6,57 (0,99)	16,14 (2,65)
$\sigma_{\varepsilon}^2$	228,24 (5,86)	226,97 (4,39)	204,86 (5,88)	229,07 (2,61)	203,44 (5,53)
$\sigma_{\mu1}^2$					0,01 (0,01)
$\sigma_{\mu2}^2$					0,01 (0,05)
<b>Intraclasse</b>	0,09	0,05	0,07	0,02	0,07
<b>Loglikelihood</b>	-78417.315	-51974.712	-26417.504	-18799,274	-26397,698
<b>R square</b>			0,13	0,06	
<b>AIC</b>	156840,630	103955,424	52845,008	37608,548	52815,396
<b>BIC</b>	156864,175	103977,738	52878,867	37640,678	52883,114
<b>Nalunos</b>	18929	12554	6450	4564	6450
<b>NCursos</b>	130	130	130	130	130

Tabela 5

*Modelos da análise multinível do Enem para área de Direito*

<b>Direito</b>	<b>M0objetiva</b>	<b>M0discussiva</b>	<b>M1objetiva</b>	<b>M1discussiva</b>	<b>M2objetiva</b>
<b>Efeito Fixo</b>					
Intercept	42,96 (0,24)	28,98 (0,19)	45,56 (0,21)	30,82 (0,19)	45,56 (0,21)
Enem07 Obj			0,34 (0,01)	0,15 (0,01)	0,34 (0,01)
Enem07Red			0,08 (0,01)	0,08 (0,01)	0,01 (0,01)
<b>Efeito Randômico</b>					
$\sigma_{\mu}^2$	17,41 (1,87)	10,091 (1,05)	10,37 (1,34)	6,57 (0,99)	10,19 (1,30)
$\sigma_{\epsilon}^2$	214,26 (2,63)	232,77 (1,76)	192,83 (3,97)	229,07 (2,61)	192,47 (4,01)
$\mu_{1j}$					0,01 (0,01)
$\mu_{1j}$					0,01 (0,01)
<b>Intraclasse</b>	0,07	0,04	0,04	0,02	0,04
<b>Loglikelihood</b>	-278727,446	-222768,383	-84513,169	-73067,417	-84507,012
<b>R square</b>			0,14	0,03	
<b>AIC</b>	557460,893	445542,766	169036,338	146144,834	169034,024
<b>BIC</b>	557488,267	445569,438	169076,055	146183,722	169113,458
<b>Nalunos</b>	67836	53683	20814	17633	20814
<b>Ncursos</b>	319	319	319	319	319

**4. Discussão**

O objetivo do presente estudo foi desenvolver e testar modelos explicativos do desempenho acadêmico no Enade a partir do desempenho do aluno no Enem e, assim, verificar o poder preditivo do Enem em relação ao desempenho futuro. Conforme demonstram os resultados, há evidências de validade preditiva para o Enem.

Primeiramente, destaca-se que os coeficientes de validade preditiva variaram de 0,34 a 0,54 na parte objetiva dependendo da área (Administração, Psicologia, Ciências Contábeis e Direito) e esses índices estão relativamente próximos ao que é encontrado na literatura em outros estudos dessa natureza utilizando o SAT (Coyle & Pillow, 2008; Coyle, Synder, Richmond, & Little, 2015; Ewing, Huff, Andrews, & King, 2005; Patterson & Mattern, 2011, 2012) e utilizando o Enem (Souza, 2006; Souza, Vendramini, & Silva, 2013).

No entanto, a parte discursiva do Enem que corresponde à redação adiciona muito pouca informação se comparada à parte objetiva, os coeficientes de validade variaram de 0,15 a 0,31. Desse modo, sugere-se que em outros estudos seja melhor controlada a precisão da redação do Enem e que novos estudos sejam realizados com a redação do exame com a finalidade de verificar, por exemplo, a influência dos avaliadores que a corrigem. Somado a isso, estudos que verifiquem a capacidade preditiva de cada uma das partes da prova do Enem como fez Coely et al. (2015) seria de grande valia para os estudos de validade preditiva na área educacional.

Ao analisar os três modelos, nota-se que há variabilidade entre as instituições de ensino (cursos), quando não se insere nenhuma variável preditora (Modelo 0), mas, quando o Enem é incluído (Modelo 1), essa variabilidade em todas as áreas cai para metade e se mantém (Modelo 2). Diante dos resultados encontrados, indiretamente, mesmo não sendo o objetivo do estudo, verifica-se que a prova do Enade que era aplicada para os alunos ingressantes pode ser substituída pela nota do Enem economizando recursos públicos, visto que o Enem consegue prever o desempenho no Enade.

É relevante apontar que talvez nenhum outro país tenha a disponibilidade de acesso ao perfil “cognitivo” dos estudantes como o Brasil possui, já que, praticamente, toda a população de estudantes do país, atualmente, faz o Enem. Aliado a isso, ainda dispomos de outra avaliação em que parte desses alunos ao final do Ensino Superior (Enade) realizam, o que permite o acompanhamento longitudinal dos estudantes brasileiros. Assim, é importante que os pesquisadores interessados pela área se apoderem mais desses dados, que desenvolvam novos estudos, que possam acompanhar o estudante mesmo depois de terminada a faculdade, afim de melhorar a informação que se tem sobre os sistemas de ensino e o impacto da educação em diferentes áreas da vida.

## 5. Referências

- Andrade, J. M., & Laros, J. A. (2007). Fatores associados as desempenho escolar: estudo multinível com dados do SAEB/2001. *Psicologia: Teoria e Pesquisa*, 23 (1), 33-42.
- Bartholomeu, D. (2011). O uso do Enem e Enade em um delineamento longitudinal para avaliação dos cursos do ensino superior. Tese de Doutorado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Bueno, J. M. P. (2013). *Análise das habilidades avaliadas pelo Enem por meio da teoria CHC: um estudo de validade*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Coyle, T. R., Pillow, D. R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence*, 36, 719-729.
- Coyle, T. R., Synder, A. C., Richmond, M. C., & Little, M. (2015). SAT non-g residuals predict course specific GPAs: Support for investment theory. *Intelligence*, 51, 57-66.
- Ewing, M., Huff, K. Andrews, M., & King, K (2005). Assessing the reliability of skills measured by the SAT. *Office of research and analysis*, RN, 24.
- Gomes, C. M. A. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio*. Tese de doutorado, Programa de Pós-Graduação em Educação: Conhecimento e Inclusão Social, da Faculdade de Educação da Universidade Federal de Minas Gerais, Belo Horizonte.
- Gomes, C. M. A., & Borges, O. N. (2009). O Enem é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, 20(42), 73-88.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York.

- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2001). *Exame Nacional do Ensino Médio (Enem): relatório pedagógico*. Brasília.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2013). *Exame Nacional do Ensino Médio (Enem): relatório pedagógico 2009-2010*. Brasília, INEP, 133p.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - Inep (2014). Edição deste ano terá mais de 8,7 milhões de participantes. Notícias. Retirado de: [http://portal.inep.gov.br/rss\\_enem/-/asset\\_publisher/oV0H/content/id/137108](http://portal.inep.gov.br/rss_enem/-/asset_publisher/oV0H/content/id/137108) em 06 de Setembro de 2014.
- Jesus, G. R., & Laros, J. A. (2004). Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, 3 (2), 21-31.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Pansanato, K. A., & Soares, J. F. (1999). Desempenho dos alunos no Enem e no vestibular da UFMG. In. *1ª Jornada Latino Americana de Estatística Aplicada*. São Carlos, SP.
- Patterson, B. F. & Mattern, K. D. (2011). *Validity of the SAT for predicting first-year grades: 2008 SAT Validity Sample*. College Board Statistical Report. New York, NK: The College Board.
- Patterson, B. F. & Mattern, K. D. (2012). *Validity of the SAT for predicting first-year grades: 2009 SAT Validity Sample*. College Board Statistical Report. New York, NK: The College Board.
- Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições contemporâneas de validade de testes psicológicos. In. C. S. Hutz (Org.). *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo, 243-265.

- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. de F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(2), 151-159.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
- Schmidt, F. L., Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124 (2), 262-274.
- Souza, A. M. (2006). *Validade preditiva de um processo seletivo em relação ao desempenho de universitários de Psicologia*. Dissertação de mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Souza, A. M., Vendramini, C. M. M., & Silva, M. C. R. (2013). Validade preditiva de um processo seletivo em relação ao desempenho de universitários de Psicologia. *Encontro Revista de Psicologia*. 16 (24), 55-68.
- Valentini, F. (2013). *A relação entre inteligência fluida, desempenho acadêmico e aprendizagem: uma abordagem multinível*. Tese de Doutorado, Programa Psicologia Social, do Trabalho e das Organizações, Universidade de Brasília, Brasília, DF.
- Zoghbi, A. C., Oliva, B. T., Moriconi, G. M. (2010). Aumentando a eficácia e a eficiência da avaliação do ensino superior: a relação entre o Enem e o Enade. *Estudos em Avaliação Educacional*, 21(45), 45-66.



**Link das notícias do Enem:**

- 1) <http://www.gazetadopovo.com.br/educacao/vida-na-universidade/vestibular/historico-de-problemas-do-enem-inclui-vazamento-e-gabarito-errado-zt7rf4ib7oo7gcus1sxeuku>
- 2) <http://www.gazetadopovo.com.br/educacao/vida-na-universidade/vestibular/dados-de-12-milhoes-de-inscritos-no-enem-vizam-na-internet-1zt9eothb5wu4z0asvf7b6tse>
- 3) <http://ultimosegundo.ig.com.br/enem/problemas+no+enem+e+sisu+e+as+respostas+do+mec/n1237958612015.html>
- 4) <http://g1.globo.com/pi/piaui/noticia/2014/12/policia-federal-confirma-vazamento-do-tema-da-redacao-do-enem-no-piaui.html>
- 5) <http://g1.globo.com/educacao/noticia/2010/11/cometi-um-erro-e-vou-pagar-diz-professor-indiciado-por-vazar-enem.html>

## ANEXO I – ACEITE PARA UTILIZAÇÃO DAS BASES DE DADOS DO ENEM E ENADE

Em 05/03/2014, às 15:41, Janaina Ferreira Ma escreveu:

Professor Ricardo Primi,

Obrigada pelo envio do termo de confidencialidade e sigilo. Sua assinatura implica que o senhor já está **autorizado** a utilizar as bases de dados disponibilizadas.

Qualquer dúvida, estou à disposição.

Cordialmente,

<image001.jpg>

Janaina Ferreira Ma

Pesquisadora-Tecnologista em Informações Educacionais

Coordenação Geral de Controle de Qualidade da Educação Superior | CGCQES

Diretoria de Avaliação da Educação Superior | DAES – INEP

Telefone: [+55 \(61\) 2022-3711](tel:+55(61)2022-3711)

## ESTUDO IV

TÍTULO EM PORTUGUÊS: **QUÃO ESTÁVEIS SÃO OS RESULTADOS DO ENEM DE ESTUDANTES QUE FIZERAM O EXAME EM 2007, 2008 E 2009?**

TÍTULO EM INGLÊS: **HOW STABLE ARE THE RESULTS OF ENEM IN STUDENTS WHO TOOK THE EXAM IN 2007, 2008 AND 2009?**

### **Resumo:**

Atualmente, cada vez mais estudantes realizam a prova do Enem, principalmente, para utilizar a nota no exame como forma de entrada na universidade. Com isso, aumentou o número de participantes que realizam a prova em vários anos. Assim, o objetivo do estudo foi verificar a estabilidade dos escores dos estudantes no Enem selecionando um banco de dados contendo informações de estudantes que realizaram o exame pelo menos duas vezes (diferentes anos). Os resultados demonstram que há estabilidade nos escores dos estudantes em diferentes anos, que as provas no modelo antigo e no novo Enem são altamente correlacionadas e que a redação correlaciona-se pouco com as outras partes da prova. É discutido também a especificidade desse estudo, já que entre as duas aplicações os estudantes treinam/estudam para melhorar o desempenho.

Palavras-chave: avaliação educacional, psicometria, Enem

### **Abstract**

Currently, more and more students perform the test of Enem, especially for use in the note taking as input the University form. With this, the number of students who undertake the race in several years. The objective of the study was to verify the stability of Enem scores selecting a database containing information of students who were examined at least twice. The results demonstrate a temporal stability, that the evidence in the old model and the new Enem are highly correlated and that the wording correlates somewhat with the other parts of the race. It also discussed the specificity of this study as between the two applications the students training/study to improve performance.

Keywords: test reliability, education assessment, psicometria, Enem

A cada ano, milhares de estudantes que estão em fase de conclusão ou concluíram o Ensino Médio realizam o Exame Nacional do Ensino Médio (Enem) no Brasil. Desde a sua criação em 1998 (Inep, 2002) o Enem foi reformulado duas vezes. As mudanças no exame referem-se ao número de questões, ao tempo de duração da prova, à nomenclatura das disciplinas, a pequenas alterações nas características das questões e ao aumento das formas de utilização da nota do estudante no exame (ver Tabela 1).

Tabela 1

*Principais diferenças entre o antigo e o novo modelo teórico do Enem*

	De 1998 até 2008	Após 2008
Número de Itens	63 questões com resposta de múltipla escolha mais uma redação	180 questões com resposta de múltipla escolha mais uma redação.
Duração	5 horas em único dia	Dois dias: 04h30min h no primeiro dia e 05h30min no segundo dia.
Disciplinas	Português, Geografia, História, Biologia, Matemática, Física, Química e Redação	Quatro áreas do conhecimento: Linguagens, códigos e suas tecnologias; Matemática e suas tecnologias; Ciências da Natureza e suas tecnologias; Ciências Humanas e suas tecnologias e a redação.
Características das Questões	Questões centradas no raciocínio; Interdisciplinaridade; Questões no formato de situações-problema.	Permanece a estrutura do modelo anterior e acrescenta o foco no conteúdo das áreas de conhecimento.
Resultados do Exame	Utilizado pelo Governo para análise do Ensino Médio e utilizado por universidades de forma parcial ou total para ingresso no Ensino Superior	Certificação do Ensino Médio; Aumento da utilização para ingresso em universidades públicas e particulares; Acesso a programas do Governo; Indicador de Qualidade Escolar.

Verificar as mudanças que ocorreram no exame ao longo de 17 anos de existência permite entender o motivo de a prova ser hoje realizada por praticamente toda população de estudantes do ensino médio brasileiro. É a possibilidade de utilizar o desempenho no Enem, principalmente, para acesso ao ensino superior que estimula cada vez mais os estudantes participarem desse processo. Em decorrência disso, muitos jovens fazem o Enem mais de uma vez, pois, caso eles não sejam aprovados no vestibular, no ano seguinte, provavelmente, tentarão novamente.

Assim, os bancos de dados do Inep, com o decorrer do tempo, passaram a possuir o desempenho de muitos estudantes no Enem em mais de um ano. Diante da disponibilidade desses dados e verificando a necessidade de desenvolver estudos com o Enem (Primi et al., 2001) pretendeu-se no presente estudo verificar se a posição relativa do estudante comparando com ele mesmo muda em diferentes avaliações (2007, 2008 e 2009).

Para entender melhor essa proposta, deve-se compreender que o objetivo desse estudo não foi verificar a precisão em si do Enem. A precisão ou também chamada fidedignidade refere-se à estabilidade dos escores dos testes dos examinandos, de forma que os escores dos examinandos, em momentos diferentes, sejam consistentes e com uma quantidade mínima de erro (Urbina, 2007; Zanon & Hauck Filho, 2015).

Quando se realizam estudos de precisão, algumas condições devem ser seguidas (ver Urbina, 2007). Por exemplo, se para verificar a precisão de um instrumento for utilizado o método de teste-reteste, deve-se dar um intervalo entre as duas aplicações para verificar a consistência dos escores de um determinado teste ao longo desse intervalo (Zanon & Hauck Filho, 2015). Contudo, voltando para o caso do Enem, o leitor familiarizado com essa propriedade psicométrica dos testes já deve ter atentado para o fato

que o intervalo entre o “teste-reteste” nesse exame não atende às condições esperadas para um estudo dessa natureza.

Os estudantes ao realizarem o Enem passam um período longo (1 ano) entre as duas avaliações e nesse intervalo são “treinados”, visto que grande parte dos estudantes passa um ano estudando, muitos fazendo cursinho, para se prepararem melhor para prova. Além disso, a prova não é a mesma. Assim, um estudo de precisão com o Enem não atende as condições para se inferir que ele é fidedigno ou não.

Pensando nessa questão e diante da disponibilidade da posição relativa do aluno de um ano para outro (nos intervalos), além da possibilidade de verificar a correlação entre as diferentes provas é que o estudo foi realizado.

## 2. Método

### 2.1. Fonte de Dados

Foi utilizado o banco de dados do Enade de 2012 que continha as informações dos estudantes de quando realizaram o Enem, pode-se verificar, conforme pode ser visualizado na Tabela 2, que 8,7% dos participantes no Enade de 2012 fizeram duas vezes o Enem e 1,8% dessa amostra realizaram o Enem 3 vezes. Assim, foram utilizadas as informações dos estudantes que possuíam pelo menos duas notas no Enem.

Tabela 2.

*Participantes dos Enade 2012 e número de vezes que fizeram o Enem*

	<i>f</i>	%
Nenhuma vez	296.811	63,2
Uma vez	123.328	26,3
Duas vezes	40.652	8,7
Três vezes	8.652	1,8
Total	469.443	100

Por meio do cruzamento da distribuição da quantidade de Enem por ano (Tabela 3) pode-se verificar melhor como se distribui a participação dos estudantes do Enade em 2012, nas provas do Enem de 2007, 2008 e 2009.

Tabela 3

*Distribuição dos Estudantes que realizaram o Enem em 2007, 2008 e/ou 2009*

Enem 2009				Enem 2008		Total
				Não fez Enem	Fez Enem	
<b>Não fez Enem</b>	Enem	N	296,811	48,822	345,633	
		%	85,9%	14,1%	100%	
	Enem	N	58,057	25,545	83,602	
		%	69,4%	30,6%	100%	
<b>Total</b>	N	354868	74,367	429,235		
	%	82,7%	17,3%	100%		
<b>Fez Enem</b>	Enem	N	16,449	9,827	26,276	
		%	62,6%	37,4%	100%	
	Enem	N	5,280	8,652	13,932	
		%	37,9%	62,1%	100%	
<b>Total</b>	N	21,729	18,479	40,208		
	%	54,0%	46,0%	100%		
<b>Total</b>	Enem	N	313,260	58,649	371,909	
		%	84,2%	15,8%	100%	
	Enem	N	63,337	34,197	97534	
		%	64,9%	35,1%	100%	
<b>Total</b>	N	376597	92846	469443		
	%	80,2%	19,8%	100%		

### 2.3. Instrumentos

*Exame Nacional do Ensino Médio (Antigo Enem)* – O primeiro modelo do Enem esteve em vigor da criação em 1998 até 2008. O exame era composto por 63 questões de múltipla escolha mais uma redação. Os itens eram construídos a partir de uma matriz de referência com 21 habilidades derivadas de 5 competências. As 63 questões eram divididas em questões de Português, Geografia, História, Biologia, Matemática, Física, Química.

*Exame Nacional do Ensino Médio (Novo Enem)* - O novo Enem utilizado a partir de 2009 é composto por 180 questões respondidas em dois dias consecutivos. No primeiro dia são aplicadas as provas da área de ciências humanas e suas tecnologias (45 questões) e da área de ciências da natureza e suas tecnologias (45 questões). No segundo dia acontece a prova com questões da área de Linguagem, códigos e suas tecnologias (45 questões) e de matemática e suas tecnologias (45 questões) mais a redação.

#### **2.4. Procedimentos**

Foi selecionado um banco de dados dos estudantes que realizaram o Enade de 2012. Nesse banco também havia informação dos mesmos estudantes antes da entrada no curso universitário quando fizeram o Enem. Após a verificação de que havia estudantes que fizeram o Enem mais do que uma vez, foi realizada uma correlação entre as notas do Enem nos diferentes anos. Por fim, foram estimadas as correlações entre as variáveis utilizando um modelo de equações estruturais.

#### **2.5. Análise de Dados**

Foi realizada uma correlação de Pearson para verificar a relação entre as partes das provas que compõem o Enem nos 3 anos, por meio do *Statistical Package for the Social Sciences* (SPSS), versão 19. Para estimar as relações entre as variáveis utilizando o modelo de equações estruturais foi utilizado o AMOS 16 (Arbuckle, 2007).

### **3. Resultados e Discussão**

Conforme visualizado na Tabela 3, foi verificada a relação entre as partes que compõem o Enem nos três anos selecionados, quais sejam, parte objetiva de 2007 e 2008, a redação de 2007, 2008 e 2009, além da parte objetiva de 2009 que passou a ser subdividida nas áreas: LPE (Linguagem, códigos e suas tecnologias), MAT (Matemática e suas



tecnologias), CAN (Ciências da Natureza e suas tecnologias), CHU (Ciências Humanas e suas tecnologias) (ver Tabela 4).

Tabela 4

*Correlação entre os Enem de 2007, 2008 e 2009*

	LPE.9	MAT.9	CAN.9	CHU.9	RED.9	OBJ.8	RED.8	OBJ.7	RED.7
LPE.9	r 1								
MAT.9	r 0,62	1							
CAN.9	r 0,62	0,68	1						
CHU.9	r 0,44	0,39	0,37	1					
RED.9	r 0,23	0,29	0,32	0,12	1				
OBJ.8	r 0,63	0,65	0,64	0,48	0,23	1			
RED.8	r 0,18	0,23	0,25	0,11	0,25	0,32	1		
OBJ.7	r 0,64	0,66	0,64	0,45	0,24	0,77	0,28	1	
RED.7	r 0,16	0,21	0,24	0,08	0,25	0,19	0,23	0,32	1

Nota: Todas as correlações foram significativas ( $p \leq 0,01$ )

As correlações em sua maioria podem ser consideradas de moderada a forte (0,44 a 0,77) (Dancey & Reidy, 2006). No entanto, a redação nos três anos apresenta correlação fraca, porém significativa ( $p \leq 0,01$ ) com todas as partes e anos do Enem. A prova de ciências humanas e suas tecnologias de 2009 merece destaque e uma análise mais detalhada por ser a que apresenta correlações mais baixas com as outras partes objetivas da prova do Enem (0,08 a 0,48).

Assim como o encontrado no estudo de evidência de validade preditiva dessa tese (ver estudo 3), verifica-se que a redação do Enem parece não trazer tanta contribuição nos estudos desenvolvidos com essa prova, necessitando que as pesquisas sejam desenvolvidas especificamente com a redação do Enem para verificar se isso deve-se à forma como a redação é corrigida ou a avaliação dos juízes que a corrigem.

Após analisar as correlações entre as partes do Enem, foi realizada uma nova análise na qual foram estimadas as correlações entre as variáveis utilizando um modelo de equações estruturais. Foi testado um modelo de 3 fatores que correspondem a prova

objetiva de 2009 (obj9), prova objetiva de 2007/2008 (obj78) e redação (red). Foi verificada a adequação do modelo segundo propõe Byrne (2001). Os resultados mostraram que o modelo de 3 fatores teve um bom ajuste, excetuando o valor do  $\chi^2$  que foi bem elevado e isso pode ter sido influenciado pelo tamanho da amostra.  $\chi^2=15131,87$ ,  $gl=24$ ,  $\chi^2/gl=130,49$ ,  $RMSEA=0,081$ ,  $CFI=0,96$ .

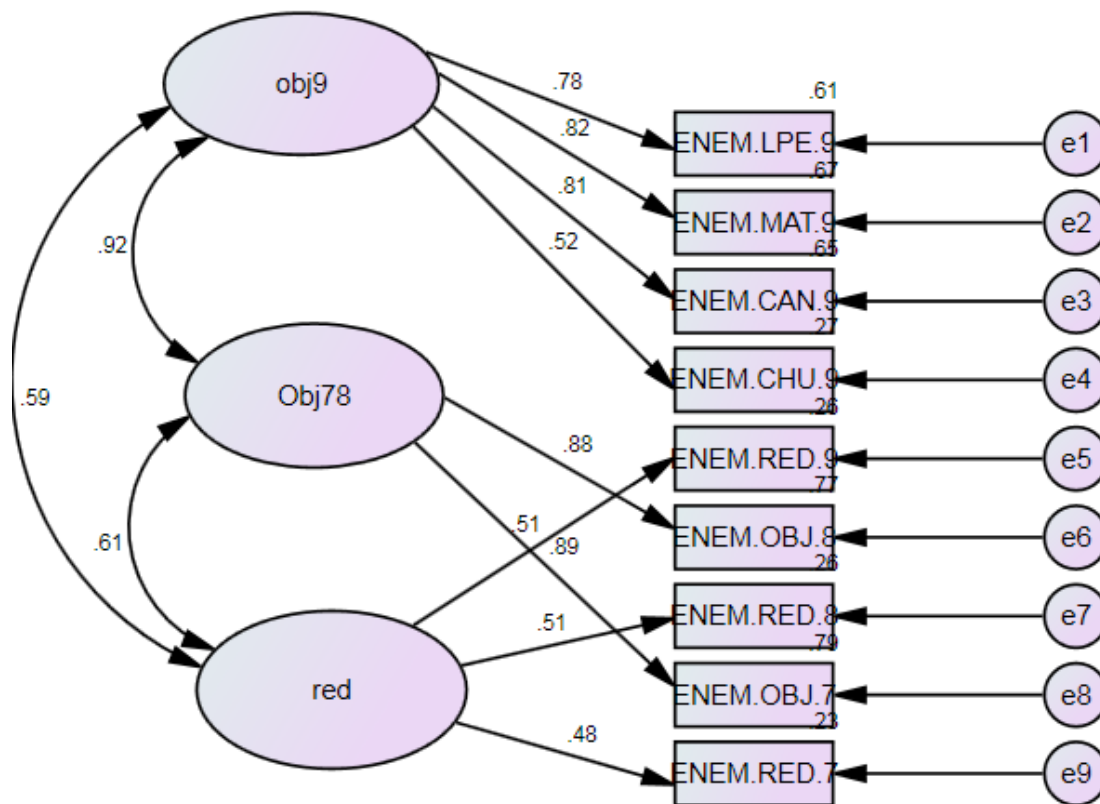


Figura 1. Modelo de três fatores com as cargas padronizadas

Na Figura 1 observa-se que as correlações mais elevadas foram entre as partes objetivas do Enem. Mesmo a estrutura do Enem tendo mudado em 2009 as provas objetivas estão altamente relacionadas o que demonstra que aparentemente continuam avaliando os mesmos aspectos. Nesse modelo, a parte objetiva e a redação estão moderadamente relacionadas, mais que as redações de forma separada (2007, 2008 e 2009).

Pode-se afirmar que há estabilidade nos escores dos examinandos que realizaram o Enem em diferentes anos. Seria interessante desenvolver outros estudos (os de caráter mais qualitativo se encaixariam bem) com os escores daqueles estudantes que realizaram o Enem pelo menos duas vezes, verificando o efeito daqueles que estudaram ao longo do ano, daqueles que não estudaram ou fizeram cursinho ao longo do ano, ou seja, que foram apenas realizar a prova sem “treino”. Além disso, comparar os escores dos estudantes nas provas a partir de 2009 que estão em uma mesma escala.

Uma limitação desse estudo é que não foi levado em consideração que muitos estudantes realizam o Enem como treineiros, ou seja, são estudantes que estão no primeiro ou segundo ano do ensino médio e que fazem o Enem para treinarem antes mesmo de chegarem ao final do ensino médio. Pode ser que parte da amostra desse estudo tenha feito isso e um grupo com essa característica é diferente de outro que já concluiu o ensino médio e está realmente focado em estudar para essa prova para melhorar seu desempenho no exame.

#### **4. Referências**

Arbuckle, JL (2007). Amos 16.0 Users Guide. Chicago: SPSS Inc.

Byrne, BM (2001). Structural Equation Modeling with AMOS. New Jersey: Lawrence Erlbaum.

Dancey, C.; Reidy, J. (2006). *Estatística Sem Matemática para Psicologia: Usando SPSS para Windows*. Porto Alegre: Artmed.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2002). *Documento Básico do Enem*. Brasília, INEP.

Pasquali, L. (2001). *Psicometria: teoria dos testes psicológicos*. Brasília: LabPAM.

- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. de F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa, 17*(2), 151-159.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- Zanon, C., Hauck Filho, N. (2015). Fidedignidade. In. C. S. Hutz, D. R. Bandeira, C. M. Trentini. *Psicometria*. Porto Alegre: Artmed.

## **Considerações Finais**

O objetivo geral da tese foi desenvolver estudos com o Enem a partir de uma visão psicométrica da inteligência. De maneira geral, os estudos desenvolvidos permitiram verificar: 1) a importância do Enem para que os pesquisadores possam verificar os alcances e limitações das avaliações educacionais em larga escala que são realizadas no país como forma de melhorar a Educação e a visão que se tem sobre o processo que vem sendo realizado; 2) A necessidade de interação entre os pesquisadores que desenvolvem estudos com as provas educacionais em larga escala e o Inep, que elabora e administra essas provas, para que os resultados não fiquem apenas no formato de dissertações, teses e/ou artigos científicos, mas que o conhecimento gerado na academia seja levado para os manuais do exame, para discussão do que é necessário melhorar e, conseqüentemente, para mudanças efetivas na Educação.

O primeiro estudo trouxe informações sobre como foram desenvolvidos os estudos de avaliação da inteligência e como essa é estudada atualmente a partir da visão psicométrica, mostrando que os estudos nas últimas duas décadas consideram que a inteligência continua a se desenvolver mesmo na vida adulta e que outros construtos como personalidade e interesses contribuem para isso (Ackerman, 1996; Ackerman & Heggestad, 1997; Ackerman & Roffhus, 1999; Lubinski & Benbow, 2000; Wai, Lubinski, & Benbow, 2005). O estudo mostrou que no Brasil o Enem se propõe avaliar as habilidades e competências dos estudantes ao final do Ensino Médio, sendo assim, essa prova pode fornecer alguns dados quanto as habilidades cognitivas do jovem adulto no país. Ademais, sintetizou o que os estudos psicométricos que já foram realizados com essa prova apontam.

Pensando nas teorias da inteligência na vida adulta que consideram alguns traços de investimento como importantes para o desenvolvimento da inteligência (Ackerman, 1996;

Lubinski & Benbow, 2000; von Stumm & Ackerman, 2013), talvez se pudesse pensar em incluir algumas questões sobre interesses e personalidade no questionário socioeconômico que o aluno preenche em casa. Dessa forma, seria possível desenvolver estudos a partir de uma visão mais integrativa da inteligência por meio do Enem, conforme são apresentadas nas teorias de inteligência no adulto (inteligência via resultado no Enem mais características de Personalidade e Interesses mensurados via questionário socioeconômico).

O segundo estudo trouxe a preocupação apontada na literatura de se ter uma prova realizada por milhares de estudantes e não se saber ao certo o que ela mede, pois nos manuais do Enem há apenas a proposta teórica que fundamenta as características que os itens devem abranger. Porém, foi enfatizado que o Enem e a maioria das provas educacionais não tem como foco testar o modelo teórico subjacente aos seus itens, tentando assim, alertar que embora seja importante testar a adequação da estrutura da prova, não fazê-lo não é uma característica apenas do Enem, mas das avaliações educacionais em larga escala de maneira geral frente as especificidades desse tipo de prova. Não obstante, os achados do estudo, a partir de uma visão psicométrica da inteligência, conseguem compreender que os itens do Enem de 2006 (antigo Enem) possuem características especialmente de inteligência cristalizada (Gc), conhecimento quantitativo (Gq), inteligência fluida (Gf), processamento visual (Gv) e capacidade de leitura e escrita (Grw).

O terceiro estudo demonstrou que apesar da falta de pesquisas empíricas nos manuais do exame, o Enem pode contribuir predizendo desempenho futuro dos estudantes e que o Brasil se destaca por dispor de duas provas padronizadas (Enem e Enade) que permitem o desenvolvimento de vários estudos entre essas duas etapas educacionais. Por fim, o quarto estudo permitiu visualizar a necessidade de se estudar melhor a redação do exame e que, mesmo havendo mudança na estrutura da prova, as correlações entre as partes

que compõem tanto o antigo quanto o novo Enem são altas. Também destaca-se que os escores dos estudantes que realizaram o exame em diferentes anos são altamente estáveis, apesar da maioria dos estudantes estudarem entre as duas aplicações e serem provas diferentes a cada ano. Não obstante, deve-se considerar as limitações dos quatro estudos apresentadas e com tais limitações pensar na possibilidade de novas pesquisas.

### Referências

- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22, 227–257.
- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121, 218–245.
- Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging*, 14, 314–330. doi:10.1037/0882-7974.14.2.314
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington.
- Becker, F. R. (2010). Avaliação educacional em larga escala: a experiência brasileira. *Revista Ibero-americana de Educação*, 53(1), 1-11.
- Castro, M. H. G. (2009). Sistemas de Avaliação da Educação no Brasil avanços e novos desafios. *São Paulo Perspectiva*. São Paulo, 23 (1), 5-18.
- Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, 55, 137-150.

- Luna, E. A. dos. A. (2011). Avaliação em larga escala e produção textual: reflexões sobre o Enem. *Estudos em Educação e Linguagem*, 1 (1).
- Mello, J. C., Leta, F., Fernandes, A., Vaz, M., Campos, M. H. & Barbejat, M. (2001). Avaliação qualitativa e quantitativa: uma metodologia de integração. *Ensaio: Avaliação Políticas Públicas Educacionais*, 09 (31), 237-251.
- Primi, R., Santos, A. A. A., Vendramini, C. M., Taxa, F., Muller, F. A., Lukjanenko, M. de F., & Sampaio, I. S. (2001). Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(2), 151-159.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- Vianna, H. M. (2003). Avaliações nacionais em larga escala: análises e propostas. *Estudos em Avaliação Educacional*, 27, 41-76.
- von Stumm, S., & Ackerman, P. L. (2013). Investment and intelligence: A review and meta-analysis. *Psychological Bulletin*, 139, 841–869.
- Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, 97(3), 484-492.
- Zanon, C., & Hauck Filho, N. (2015). Fidedignidade. In C. S. Hutz, D. R. Bandeira, C. M. Trentini. *Psicometria*. Porto Alegre: Artmed.