

EVIDÊNCIAS SOBRE A ESTRUTURA INTERNA DO ENEM E AS COMPETÊNCIAS DE CIÊNCIAS NATURAIS¹

 <https://doi.org/10.56238/sevened2025.011-067>

Rodrigo Travitzki

Grau de formação mais alto: Doutor em Educação pela Universidade de São Paulo

Instituição acadêmica: Unicamp, Campinas

email: r.travitzki@gmail.com

ORCID: <https://orcid.org/0000-0002-0107-682X>

Ricardo Primi

Grau de formação mais alto: Doutor em Psicologia Escolar e do Desenvolvimento

Humano pela Universidade de São Paulo

Instituição acadêmica: Universidade São Francisco, Campinas

email: ricardo.primi@usf.edu.br

ORCID: <https://orcid.org/0000-0003-4227-6745>

RESUMO

Neste trabalho, investiga-se a estrutura interna do ENEM, com especial atenção à prova de Ciências Naturais. Foram analisados os microdados de 2014 e 2015. A análise fatorial sugere que um fator é suficiente para descrever a variância dos 175 itens. O modelo bifator sugere heterogeneidade sistemática, nos dois anos, entre os subfatores, especialmente nas provas de Matemática e de Linguagens e Códigos. Por outro lado, a média das 4 provas se mostrou colinear ao fator geral. A prova de Ciências Naturais apresentou considerável multidimensionalidade interna, o que poderia ser uma característica própria da alfabetização científica. As competências de Ciências Naturais descritas na Matriz de Referência se mostraram coerentes com o comportamento empírico dos itens, sugerindo que a estrutura interna da prova reflete seus fundamentos teóricos.

Palavras-chave: ENEM. Alfabetização científica. Avaliação educacional. Validade. Modelo bifator.

¹ Este trabalho foi parcialmente apresentado em 2018 no VI CONBRATRI, organizado pela Associação Brasileira de Avaliação Educacional (ABAVE).

1 INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) é uma importante referência para a educação brasileira e carecem estudos de validade das provas a partir de 2009, quando o Exame passou por transformações estruturais. Há evidências de que as primeiras edições do ENEM avaliavam mais a capacidade de raciocínio (inteligência fluida) do que o conhecimento propriamente dito (inteligência cristalizada) (PRIMI et al., 2001). Sobre as edições recentes há poucos estudos de validade (GOMES, GOLINO, SOUZA PERES, 2020; TRAVITZKI, 2017).

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

Em relação às Ciências Naturais, há evidências de que crianças de 4 a 6 anos possuem elementos básicos do raciocínio científico, como relação entre hipótese e evidência, ou mesmo percepção de covariância (KOERBER et al., 2005). Segundo Stephen Norris e Linda Phillips, a alfabetização científica tem uma particularidade em relação aos outros tipos de alfabetização, ou letramento,² pois dependente mais de conhecimento prévios específicos. Nesse sentido, os autores consideram que há dois aspectos da alfabetização científica: 1) o fundamental (capacidade de compreender, interpretar, analisar e criticar qualquer texto) e 2) o derivado (dependente de conteúdos específicos de cada ciência). No entanto, concluem, o sentido fundamental é pouco aplicado na educação científica (NORRIS; PHILLIPS, 2003), o que pode estar relacionado à importância relativamente maior dos conteúdos específicos neste tipo de alfabetização, às vezes sobrevalorizada.

Este artigo visa contribuir para a compreensão da estrutura interna do ENEM, aplicando técnicas psicométricas no exame como um todo e também particularmente na prova de Ciências Naturais. Os primeiros resultados se referem à estrutura interna em termos gerais, sem levar em conta a Matriz de Referência do ENEM, com foco principalmente no modelo bifator. Na sequência, são apresentados resultados relativos à coerência entre as competências de Ciências Naturais da Matriz de Referência e o comportamento empírico dos itens.

2 METODOLOGIA

Realizou-se inicialmente a análise fatorial dos 175 itens³ de cada ano, a partir das correlações tetracóricas. A análise fatorial é um dos fundamentos psicométricos para investigação da inteligência

² O termo original em inglês é *literacy*, que é traduzido na literatura brasileira ora como “alfabetização”, ora como “letramento” (SASSERON, CARVALHO, 2011), que neste trabalho são considerados sinônimos.

³ Tendo em vista que 5 itens da prova de Linguagens e Códigos são de língua estrangeira e, portanto, serão retirados da análise.

(PRIMI et al., 2001). No contexto das avaliações educacionais, é comum observar-se um fator predominante na prova. Uma questão importante, neste caso, é verificar em que medida este fator predominante corresponde ao construto planejado, e em que medida corresponde a habilidades mais genéricas, relacionadas à resolução de provas, tais como a inteligência fluida e a velocidade de processamento – na nomenclatura da Teoria de Cattell-Horn-Carroll (PRIMI, 2003).

Para buscar respostas a esta questão, utilizou-se também modelos bifator, que foram inicialmente propostos como alternativa à estrutura oblíqua simples de Thurstone, contemplando diferentes níveis de complexidade no comportamento psicológico (SCHMID; LEIMAN, 1957). Com efeito, tal tipo de modelo permite que se isole o fator predominante do teste a fim de se investigar aspectos mais específicos de sua estrutura interna. É estimado um fator geral (presente em todos os itens) e alguns subfatores (presentes em subconjuntos de itens), que corresponderiam a construtos conceitualmente mais específicos. Normalmente assume-se que o fator geral é ortogonal aos subfatores. Os modelos bifator permitem: a) investigar o particionamento da variância quando se acredita que há um fator mais geral e alguns subfatores; b) o controle da multidimensionalidade em testes essencialmente unidimensionais; c) avaliar se o fator geral é forte o suficiente para se justificar modelos unidimensionais; d) determinar a adequação do score geral, e se há algum ganho com a inclusão dos sub scores (RODRIGUEZ; REISE; HAVILAND, 2016).

Para a análise da coerência entre as competências de CN e o comportamento empírico dos itens, foram calculadas as correlações tetracóricas inter-item de todos os 45 itens. Depois disso, foram calculadas as médias dessas correlações, agrupadas por competência. O mesmo procedimento foi repetido agrupando-se as correlações por disciplina, para fins de comparação.

Foram analisados os microdados do ENEM de 2014 e 2015 com o software R, pacote Psych (REVELLE, 2018). A amostra incluiu apenas os concluintes do Ensino Médio, provenientes de escolas regulares, presentes nos dois dias. Todos os cadernos foram incluídos, com uma exceção.⁴ Para tanto, os itens foram reordenados com as informações fornecidas nos próprios microdados. A referência para a ordenação de itens foi o caderno azul para o sábado – Ciências Naturais (CN) e Ciências Humanas (CH) – e o caderno amarelo para o domingo – Matemática (MT) e Linguagens e Códigos (LC). Foi utilizada uma amostra aleatória de 300 mil alunos, em virtude de limitações computacionais.

3 ESTRUTURA INTERNA DO ENEM

As oito provas apresentaram coeficiente Alfa de Cronbach aceitável (maior do que 0,7), sendo os valores mais baixos nas provas de CN (0,76 em 2014 e 0,79 em 2015). Analisando o conjunto das quatro provas, o coeficiente alfa foi de 0,95, embora tenham sido encontrados alguns itens

⁴ O caderno rosa da prova de CH de 2014 foi retirado pois apresentou uma inconsistência, provavelmente relacionada à tabela que descreve a correspondência entre itens e cadernos.

negativamente correlacionados com o todo (4 em 2014 e 3 em 2015). Cabe notar que um coeficiente alfa maior do que 0,9 pode indicar número excessivo de itens.

A análise paralela dos 175 itens revelou, nos dois anos, um fator predominante e um segundo fator menor também proeminente, além de diversos outros potencialmente significativos. Analisando as quatro provas separadamente, as de MT e CN apresentaram maior número de componentes do que as outras duas (especialmente CN, com 7 componentes em 2014 e 9 em 2015).⁵ Tais resultados são compatíveis com a hipótese de Norris e Phillips (2003), relacionada à maior necessidade de conhecimentos específicos prévios para a alfabetização científica. A análise fatorial revelou que um fator seria o suficiente para contemplar a variância comum. Por outro lado, uma análise mais completa pode ser realizada com três fatores, tendo em vista que a análise com um fator captou 12% da variância total, enquanto a análise com três fatores captou 14%.

Foi estimado também um modelo bifator com três subfatores para cada ano, contemplando os 175 itens. O coeficiente ômega hierárquico estima a proporção da variância no score total que pode ser atribuída ao fator geral, tratando como erro de medida a variância “decorrente” dos subfatores (RODRIGUEZ; REISE; HAVILAND, 2016). Quando comparado ao ômega total, ele indica a adequação de modelos unidimensionais e possíveis ganhos com modelos multidimensionais. No ENEM 2014, o coeficiente ômega hierárquico foi de 0,77 e em 2015 de 0,72, sendo que no dois anos o coeficiente ômega total foi de 0,95. Com efeito, cerca de um quinto da variância nos scores únicos pode ser atribuído à multidimensionalidade.

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

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

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

⁵ Número de componentes com autovalor maior do que autovalor aleatório na análise paralela.

avaliadas pelas provas de MT e LC. Embora tais diferenças sejam pouco impactantes nas notas do ENEM, tal como calculadas hoje.

Tabela 1: coeficientes de congruência

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

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

Em relação à prova de Ciências Naturais, observa-se novamente certa proximidade com o sub fator F2. No entanto, este fator está igualmente presente nas provas de MT, dificultando a identificação de um fator específico para CN. O mesmo ocorre entre Ciências Humanas e F1.

A análise fatorial da prova de Ciências Naturais 2015 sugere que um único fator seria suficiente para representar a variação comum, explicando 10% da variância nos 45 itens. A análise paralela mostra a existência de um segundo fator proeminente e um total de 15 fatores (ou 9 componentes) com auto valor maior do que o esperado aleatoriamente. O coeficiente Alfa foi 0,8 (considerado adequado, supondo-se a unidimensionalidade). No modelo bifator, o ômega hierárquico foi de 0,64 (ômega total = 0,81) e em 2014 foi de 0,32 (0,78), sugerindo que um score único não representa adequadamente a estrutura interna da prova de CN.

4 EVIDÊNCIAS EMPÍRICAS DAS COMPETÊNCIAS DE CIÊNCIAS NATURAIS

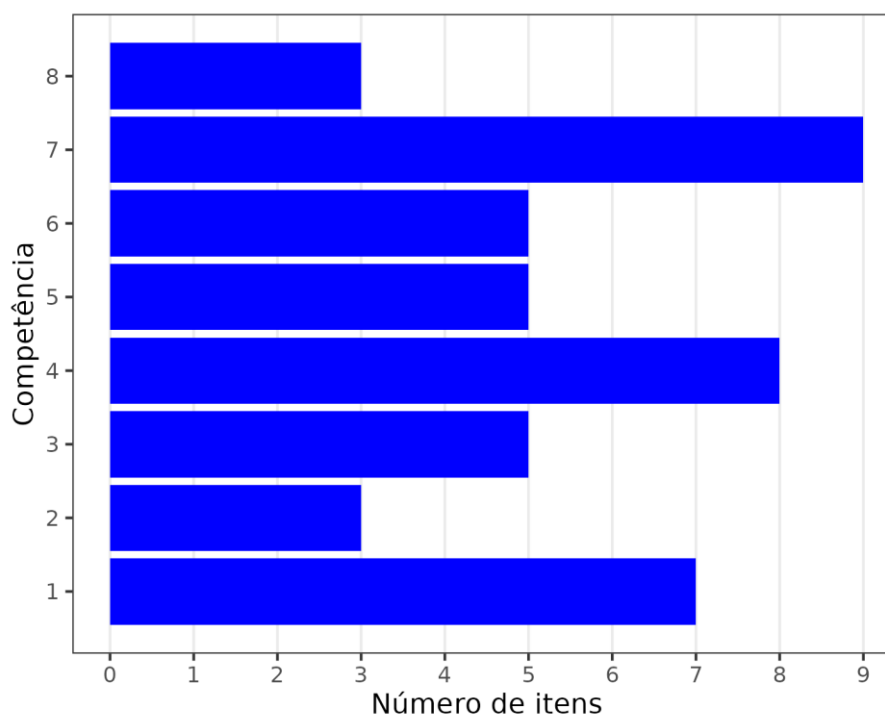
Analisamos a prova de CN de 2015 mais profundamente, com objetivo de verificar em que medida a estrutura teórica de habilidades e competências é coerente com o comportamento empírico dos itens. Espera-se que os acertos nos itens de uma certa competência estejam mais correlacionados entre si do que com os itens das outras competências, o que seria uma evidência de discriminação no nível das competências. A Matriz de Referência do ENEM (INEP, 2009) descreve 30 habilidades para a área de CN, estruturadas nas 8 competências a seguir:

1. Compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade.
2. Identificar a presença e aplicar as tecnologias associadas às ciências naturais em diferentes contextos.

3. Associar intervenções que resultam em degradação ou conservação ambiental a processos produtivos e sociais e a instrumentos ou ações científico-tecnológicos.
4. Compreender interações entre organismos e ambiente, em particular aquelas relacionadas à saúde humana, relacionando conhecimentos científicos, aspectos culturais e características individuais.
5. Entender métodos e procedimentos próprios das ciências naturais e aplicá-los em diferentes contextos.
6. Apropriar-se de conhecimentos da física para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.
7. Apropriar-se de conhecimentos da química para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.
8. Apropriar-se de conhecimentos da biologia para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

A prova de CN de 2015 apresenta ao menos um item para cada uma das 30 habilidades da matriz de referência. Evidentemente, como são 45 itens no total, não é possível haver 3 itens para cada habilidade, número mínimo recomendado para mensuração de construtos em geral. Por outro lado, tal condição é satisfeita no nível das competências, como se observa na Figura 1, que informa o número de itens para cada competência.

Figura 1: correlações inter-item médias entre as 8 competências de Ciências Naturais



Fonte: elaboração própria com dados do ENEM 2015.

Nota-se na Figura 1 certo desbalanceamento entre as competências, ou seja, há considerável variação no número de itens relativo a cada uma, o que acaba sobrevalorizando algumas competências em detrimento de outras, no cálculo da proficiência de CN. Cabe destacar ainda, nesse sentido, as competências de 6 a 8, que são específicas de cada disciplina: há 3 vezes mais itens de química (competência 7) do que de biologia (competência 8). Contudo, isso não é necessariamente um problema no que se refere ao equilíbrio entre as três disciplinas de CN, pois há itens de biologia nas outras competências, por exemplo. Segundo nossa leitura do conteúdo da prova, há 16 itens com conteúdos de física, 13 de química e 15 de biologia, o que pode ser considerado suficientemente balanceado. Mas no que se refere às 8 competências da matriz, a falta de balanceamento é nítida.

Para verificar a coerência entre a matriz teórica e os resultados empíricos, calculamos a correlação inter-item entre todos os 45 itens da prova, e depois calculamos a média dessas correlações agrupando-as por competência. Por exemplo, como mostra a Tabela 2, a média das correlações inter-item entre os 7 itens da competência 1 é de 0,19 enquanto a média das correlações inter-item entre esses 7 itens e os 3 itens da competência 2 é de 0,05. Para se ter uma referência geral, a média de todas as correlações inter-item dos 45 itens da prova de CN é de 0,11.

Tabela 2: correlações inter-item médias entre as 8 competências de Ciências Naturais

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8
Comp. 1	0,19	0,05	0,07	0,08	0,07	0,07	0,05	0,05
Comp. 2	0,05	0,36	0,06	0,07	0,07	0,07	0,04	0,04
Comp. 3	0,07	0,06	0,28	0,12	0,11	0,09	0,07	0,07
Comp. 4	0,08	0,07	0,12	0,23	0,12	0,09	0,08	0,08
Comp. 5	0,07	0,07	0,11	0,12	0,3	0,09	0,08	0,07
Comp. 6	0,07	0,07	0,09	0,09	0,09	0,25	0,06	0,05
Comp. 7	0,05	0,04	0,07	0,08	0,08	0,06	0,15	0,04
Comp. 8	0,05	0,04	0,07	0,08	0,07	0,05	0,04	0,37

Fonte: elaboração própria com dados do ENEM 2015.

A informação mais importante fornecida pela Tabela 2 é de que, em cada uma das 8 competências, as maiores correlações inter-item são aquelas entre os itens da própria competência (destaques em negrito). Tal resultado é uma evidência de validade da matriz de referência do ENEM, pois se um respondente acerta um dos itens de certa competência, há uma probabilidade maior dele acertar os outros itens dessa competência, em relação aos itens de outras competências. Em outras palavras, a estrutura teórica representada pelas 8 competências apresenta coerência com o comportamento empírico dos itens.

Além disso, vale destacar que as competências com maior consistência interna (maior média das correlações inter-item) são a 2 e a 8. Como se observa na Figura 1, essas são justamente as competências com menor número de itens. De forma análoga, as competências com menor consistência interna são a 1 e 7, que estão entre as competências com maior número de itens. Com

efeito, parece haver uma relação entre o número de itens e a consistência interna das competências, o que não é inesperado.

Outro ponto a destacar da Tabela 2 é que as correlações inter-item sugerem certa semelhança entre as competências 3, 4 e 5. A simples leitura dos descritores dessas competências não revela semelhanças claras entre as três, embora as competências 3 e 4 tratem da temática ambiental em alguma medida. Assim sendo, tal resultado precisaria ser melhor compreendido, caso confirmado em outras edições do ENEM.

Por fim, realizamos procedimento semelhante ao da Tabela 2 para as três disciplinas de Ciências Naturais, resultado na Tabela 3. As informações da Tabela 3 não têm função de validação da prova de CN, pois a matriz de referência é estruturada em competências e habilidades, não em disciplinas. Assim sendo, as informações da Tabela 3 podem servir mais como um complemento ou um referencial de comparação, a fim de compreender melhor o que está sendo avaliado pelos itens de CN.

Tabela 3: correlações inter-item médias entre as 3 disciplinas de Ciências Naturais

	Física	Química	Biologia
Física	0,12	0,06	0,08
Química	0,06	0,14	0,08
Biologia	0,08	0,08	0,17

Nota-se que, também na Tabela 3, há evidências de discriminação, pois as correlações são maiores entre os itens de uma mesma disciplina (destaques em negrito). Por outro lado, a magnitude das correlações é, via de regra, menor do que a encontrada quando se separa os itens por competência (Tabela 2). Com efeito, os resultados da Tabela 3 corroboram a hipótese de que a estrutura teórica de competências é coerente com o comportamento empírico dos itens. Em outras palavras, a organização dos itens por competências é mais relevante para os resultados do Enem do que a organização dos itens por disciplina, o que está de acordo com os princípios que norteiam o exame.

Cabe ressaltar ainda que, de forma geral, as correlações nas Tabelas 2 e 3 são relativamente pequenas, de baixa magnitude, se comparadas com outros contextos. No entanto, no contexto do ENEM, que é um exame cuja nota é calculada utilizando-se o modelo logístico da Teoria da Resposta ao Item, espera-se que cada item seja uma forma independente de medida para um mesmo construto, um princípio conhecido como independência local (PASQUALI, PRIMI, 2003). Em tal contexto, é esperado que as correlações entre os itens que medem o mesmo construto sejam positivas, porém baixas. Uma correlação muito alta entre dois itens poderia indicar que eles não satisfazem a condição de independência local, distorcendo assim a estimação da proficiência, ou seja, o cálculo da nota em cada uma das quatro provas do ENEM.

5 CONCLUSÕES

A estrutura interna do conjunto de 175 itens se mostrou unidimensional na prática, embora tenham sido identificados três subfatores, além do fator g. O score do fator g apresentou correlação de 0,97 com a média das 4 provas. Os subfatores apresentaram congruência sistemática, em 2014 e 2015, observando-se diferença clara entre os itens de Matemática e de Linguagens e Códigos. De forma geral, os resultados apontam a necessidade de se aprimorar a especificidade dos construtos relativos às provas de Ciências Naturais e Humanas.

Na prova de Ciências Naturais, o coeficiente ômega hierárquico foi de 0,32 em 2014, revelando sérias limitações da nota de CN, uma única nota, para representar a multidimensionalidade interna desta prova – resultado compatível com a hipótese de Norris e Phillips sobre a heterogeneidade interna da alfabetização científica. Por outro lado, o comportamento empírico dos itens se mostrou coerente com a estrutura teórica de competências descrita na Matriz de Referência do ENEM. Adicionalmente, o comportamento empírico dos itens de CN se mostrou mais coerente com as oito competências do que com as três disciplinas, física, química e biologia. Com efeito, esses resultados confirmam a ideia de que o ENEM é estruturado em torno de competências gerais de CN, superando em alguma medida a visão da alfabetização científica fragmentada em disciplinas.

REFERÊNCIAS

- AERA; APA; NCME. Standards for educational and psychological testing. Washington: American Educational Research Association, 2014.
- GOMES, Cristiano Mauro Assis; GOLINO, Hudson Fernandes; DE SOUZA PERES, Alexandre José. Fidedignidade dos escores do Exame Nacional do Ensino Médio (Enem). *Psico*, v. 51, n. 2, 2020.
- INEP. Matriz de referência do ENEM. Brasília: INEP. 2009.
- KOERBER, S.; SODIAN, B.; THOERMER, C.; NETT, U. Scientific Reasoning in Young Children: Preschoolers' Ability to Evaluate Covariation Evidence. *Swiss Journal of Psychology*, v. 64, n. 3, p. 141–152, set. 2005.
- NORRIS, S. P.; PHILLIPS, L. M. How Literacy in Its Fundamental Sense Is Central to Scientific Literacy. *Science Education*, v. 87, n. 2, p. 224–240, 2003.
- PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, v. 2, n. 2, p. 99-110, 2003.
- PRIMI, R. Inteligência: Avanços nos Modelos Teóricos e nos Instrumentos de Medida. *Avaliação Psicológica*, v. 1, p. 67–77, 2003.
- PRIMI, R.; SANTOS, A. a. A. Dos; VENDRAMINI, C. M.; TAXA, F.; MULLER, F. A.; LUKJANENKO, M. D. F.; SAMPAIO, I. S. Competências e habilidades cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, v. 17, n. 2, p. 151–159, 2001.
- REVELLE, W. psych: Procedures for Psychological, Psychometric, and Personality Research. Disponível em: <<http://cran.r-project.org/package=psych>>. Acesso em: 4 mar. 2024.
- RODRIGUEZ, A.; REISE, S. P.; HAVILAND, M. G. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, v. 21, n. 2, p. 137–150, 2016.
- SASSERON, Lúcia Helena; CARVALHO, Anna Maria Pessoa. Alfabetização científica: uma revisão bibliográfica. *Investigações em ensino de ciências*, v. 16, n. 1, p. 59-77, 2011.
- SCHMID, J.; LEIMAN, J. M. The development of hierarchical factor solutions. *Psychometrika*, v. 22, n. 1, p. 53–61, 1957.
- TRAVITZKI, R. Avaliação da qualidade do Enem 2009 e 2011 com técnicas psicométricas. *Estudos em Avaliação Educacional*, v. 28, n. 67, p. 256, 28 abr. 2017.