



Educação Física no Exame Nacional do Ensino Médio: análise via teoria clássica dos testes

Physical education in the National Exam of Upper Secondary Education:
analysis from classical test theory

Educación física en el Examen Nacional de la Secundaria Superior:
análisis por la teoría clásica de las pruebas

Volumen 20, Número 1
Enero - Abril
pp. 1-18

Leandro Araujo de Sousa
José Airton de Freitas Pontes Junior
Adriana Eufrásio Braga

Citar este documento según modelo APA

Sousa, Leandro Araujo de; Pontes Junior, José Airton de Freitas e Braga, Adriana Eufrásio. (2020). Educação Física no Exame Nacional do Ensino Médio: análise via teoria clássica dos testes. *Revista Actualidades Investigativas en Educación*, 20(1), 1-18. Doi. 10.15517/aie.v20i1.40126

Educação Física no Exame Nacional do Ensino Médio: análise via teoria clássica dos testes

Physical education in the National Exam of Upper Secondary Education: analysis from classical test theory

Educación física en el Examen Nacional de la Secundaria Superior: análisis por la teoría clásica de las pruebas

Leandro Araujo de Sousa¹

José Airton de Freitas Pontes Junior²

Adriana Eufrásio Braga³

Resumo: O Exame Nacional do Ensino Médio (ENEM) no Brasil, é uma avaliação em larga escala que tem o objetivo de medir o conhecimento de estudantes ao final da Educação Básica. Em 2009 a Educação Física passou a compor o Exame. Dessa forma, a pesquisa objetivou analisar os itens de Educação Física do exame dos anos de 2009 a 2014 a partir da teoria clássica. Para isso, a pesquisa tem um enfoque predominantemente quantitativo e exploratório. O estudo foi realizado com uma amostra aleatória de participantes do estado do Ceará, Brasil, que realizaram o exame. Foram analisados os seguintes parâmetros: validade, fidedignidade, dificuldade e discriminação. Os itens apresentaram bons valores de correlação e adequação da amostra. No entanto, apresentaram escores de comunalidade e cargas fatoriais inadequados para composição da prova. A Análise Fatorial Exploratória apresentou baixa explicação da variância considerando apenas um fator, mesmo Scree plot indicando a unidimensionalidade. Os valores de fidedignidade foram bons, não havendo influência dos itens de Educação Física. A dificuldade e discriminação apresentaram valores aceitáveis em quase todos os anos. No entanto, em 2014 a prova não apresentou unidimensionalidade. Neste ano, os itens apresentaram alta dificuldade e baixa discriminação. Dessa forma, alguns itens de Educação Física do exame não apresentaram parâmetros adequados. As provas de Linguagens e Códigos do exame apresentaram dificuldades de comprovação da unidimensionalidade. Tais fatores podem comprometer a validade da medida e consequentemente dos resultados desse exame, podendo interferir no ingresso de milhões de brasileiros no Ensino Superior.

Palavras-chave: avaliação padronizada, ensino médio, teoria clássica.

Abstract: The National Exam of Upper Secondary Education (ENEM) in Brazil, is a large scale evaluation that aims to measure students' knowledge at the end of Basic Education. Physical Education began to form of the Exam in 2009. Thus, the research aimed to analyze the Physical Education items of the exam from the years 2009 to 2014 from the classical theory. The research has a predominantly quantitative and exploratory approach. The study was performed with a random sample of participants from state of Ceará, Brazil, who were examined. The following parameters were analyzed: validity, reliability, difficulty and discrimination. The items presented good values of correlation and adequacy of the sample. However, they presented commonality scores and factorial loads that were not adequate for the test composition. The Exploratory Factor Analysis presented low explanation of the variance considering only one factor, even Scree plot indicating the unidimensionality. The reliability values were good, with no influence of Physical Education items. The difficulty and discrimination presented values acceptable in almost every year. However, the test did not present unidimensionality in 2014. This year, the items presented high difficulty and low discrimination. Thus, some Physical Education items of the exam did not present adequate parameters. The tests of Languages and Codes presented difficulties in proving unidimensionality. Such factors may compromise the validity of the measure and consequently of the results of this exam, and may interfere with the enrollment of millions of Brazilians in Higher Education.

Keywords: standardized assessment, secondary education, classical test theory.

¹ Docente do Instituto Federal do Ceará (IFCE), Brasil. Doutor em Educação pela Universidade Federal do Ceará (UFC), Brasil. Dirección electrónica: leandro.sousa@ifce.edu.br ORCID <https://orcid.org/0000-0002-0482-2699>

² Docente da Universidade Estadual do Ceará (UECE), do Programa de Pós-Graduação em Educação (PPGE), Brasil. Doutor em Educação pela Universidade Federal do Ceará (UFC), Brasil. Dirección electrónica: japontesjr@gmail.com ORCID <https://orcid.org/0000-0003-2045-2461>

³ Docente da Universidade Federal do Ceará (UFC), do Programa de Pós-Graduação em Educação (PPGE), Brasil. Doutor em educação pela Universidade Federal do Ceará (UFC), Brasil. Dirección electrónica: adrianaufc@yahoo.com.br ORCID <https://orcid.org/0000-0001-5163-209X>

Artículo recibido: 26 de junio, 2019

Enviado a corrección: 17 de setiembre, 2019

Aprobado: 14 de octubre, 2019

Resumen: El Examen Nacional de la Secundaria Superior (ENEM), en Brasil, es una evaluación a gran escala que tiene el objetivo de medir el conocimiento de estudiantes al final de la Educación Básica. En 2009, la educación física pasó a componer el examen. De esta forma, la investigación tuvo por objetivo analizar los ítems de Educación Física del Examen de los años 2009 a 2014 a partir de la teoría clásica. Para ello, la investigación presenta un enfoque predominantemente cuantitativo y exploratorio. El estudio se realizó con una muestra aleatoria de participantes del estado de Ceará, Brasil, que fueron examinados. Se analizaron los siguientes parámetros: validez, fiabilidad, dificultad y discriminación. Los ítems presentaron buenos valores de correlación y adecuación de la muestra. Sin embargo, indica comunalidad y cargas factoriales inadecuadas para la composición de la prueba. El Análisis Factorial Exploratorio presentó baja explicación de la varianza al considerar solo un factor, a pesar de que lo scree plot indica la unidimensionalidad. Los valores de fiabilidad fueron buenos, no habiendo influencia de los ítems de Educación Física. La dificultad y la discriminación presentaron valores aceptables en casi todos los años. No obstante, en 2014, la prueba no presentó unidimensionalidad. Este año, los ítems presentaron alta dificultad y baja discriminación. Así, algunos ítems de Educación Física del examen no muestran parámetros adecuados. Las pruebas de Lenguajes y Códigos del examen presentaron dificultades de comprobación de la unidimensionalidad. Tales factores pueden comprometer la validez de la medida y consecuentemente de los resultados de ese examen, y pueden interferir en el ingreso de millones de brasileños en la Enseñanza Superior.

Palabras clave: evaluación estandarizada, enseñanza media, teoría clásica.

1. Introdução

As problemáticas que permeiam a avaliação educacional, como a seleção de instrumentos e o uso dos resultados, têm sido amplamente discutidas por pesquisadores, estudiosos, professores e pessoas ligadas direta ou indiretamente à avaliação. No Brasil, muitas discussões têm sido empreendidas sobre as avaliações em larga escala a partir da década de 1980 (Gatti, 2013). Isso fez com que fossem desenvolvidos estudos sobre sistemas educacionais, escolas, rendimentos dos alunos, insumos e oportunidades educacionais (Freitas, 2013). A partir disso, apesar do desenvolvimento ocorrido nessa área, ainda há muito a se investigar.

Essas avaliações têm implementações mais consistentes na Educação Básica. No entanto, com concentração de esforços nas disciplinas de português e matemática. A exceção é o Exame Nacional do Ensino Médio (ENEM) e o Exame Nacional e Certificação de Jovens e Adultos (ENCCEJA) que avaliam todas as disciplinas do Ensino Médio, embora contemple mais itens das disciplinas de português e matemática. A Educação Física foi inserida na matriz de referência do ENEM em 2009. Desde então foi possível identificar pelo menos 18 questões relacionadas ao objeto de estudo da Educação Física (Fernandes, Rodrigues, e Nardon, 2013). Isso pode de alguma forma ter contribuído para a perspectiva da avaliação de aspectos cognitivos na disciplina.

Em Educação Física esse tema é ainda mais polêmico por causa da sua multiplicidade de objetivos ligados aos aspectos conceituais, procedimentais e atitudinais, exigindo uma variedade de instrumentos que contemplem todos esses fatores.

A avaliação de aspectos cognitivos em Educação Física ainda é algo relativamente novo. Todavia, a partir da elaboração dos Parâmetros Curriculares Nacionais (PCN's), a dimensão cognitiva do ensino de Educação Física ganhou destaque, uma vez que, o documento ressalta o desenvolvimento de competências e habilidades cognitivas específicas para esse componente curricular (Governo do Brasil, Ministério da Educação [MEC], 2000). Essa dimensão foi reforçada nas competências e habilidades para a Educação Física na Base Nacional Comum Curricular (Governo do Brasil, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [INEP], 2018). Essas medidas podem ter contribuído para que os conteúdos da Educação Física fossem incluídos na matriz de competências e habilidade avaliadas pelo ENEM a partir de 2009, desde então, itens relacionados a essa disciplina têm sido contemplados nessa avaliação.

Por ser recente, estudos que se debruçam na análise da Educação Física em avaliações de larga escala ainda são escassos. Encontram-se estudos que analisam as influências da inserção da Educação Física na matriz de referência do ENEM nos currículos e na prática de professores da Educação Básica (Beltrão, 2014); que discutem a “desvantagem” de estudantes do turno noturno, já que alguns são isentos da disciplina por determinação legal (Fernandes, Rodrigues e Nardon, 2013) e que analisam qualitativamente os itens de Educação Física desse exame (Souza, Diniz, Ditomaso, e Darido, 2012). No entanto, estudos que analisam especificamente os aspectos psicométricos dos itens de Educação Física do ENEM são insuficientes ou mesmo inexistentes.

Quando se considera estudos sobre a avaliação em larga escala em Educação Física no Brasil, poucas pesquisas são encontradas (Beltrão, 2014; Fernandes et al., 2013; Souza et al., 2012). Entretanto, esses estudos apresentam caráter amplo do assunto e sem aprofundamentos nas análises dos dados das avaliações. A partir disso, propomos uma análise psicométrica mais detalhada dos itens de Educação Física do Exame Nacional do Ensino Médio entre os anos de 2009 a 2014.

As características psicométricas, ou seja, os parâmetros de dificuldade e discriminação dos itens, assim como medidas de validade e fiabilidade das provas do ENEM são pouco exploradas nos estudos e pesquisas em avaliação educacional, e quando se trata dos itens de Educação Física as pesquisas são quase inexistentes, salvo os relatórios pedagógicos que traçam uma breve e superficial análise, avaliando o comportamento dos participantes em relação a cada item (INEP, 2013). Diante disso, surge a seguinte pergunta: Os itens de

Educação Física do ENEM apresentam características psicométricas adequadas para medir os conhecimentos dos candidatos nessa área?

A partir disso, o estudo possibilitará conhecer as características psicométricas dos itens de Educação Física do ENEM a partir das respostas dos candidatos ao exame, permitindo assim, conhecer seus parâmetros métricos, adequabilidade e qualidade desses itens em medir conhecimentos e conseqüentemente estimar o desempenho nessa área dos candidatos que se submetem ao exame. Tal análise se justifica desde que, nos últimos anos, o ENEM passou a ser utilizado como forma de seleção dos candidatos aos cursos superiores das Instituições de Ensino Superior (IES) públicas e privadas brasileiras. Com isso, torna-se necessário a utilização de um instrumento confiável para realização justa dessas seleções.

Nesse contexto, este estudo tem o objetivo de analisar características psicométricas dos itens de Educação Física do Exame Nacional do Ensino Médio (ENEM) de 2009 a 2014 via Teoria Clássica dos Testes. Especificamente, apresenta-se os objetivos de estimar os parâmetros de dificuldade e discriminação dos itens de Educação Física do ENEM e verificar a adequabilidade dos itens de Educação Física para a validade e fidedignidade da prova de Linguagens e Códigos do exame.

2. Breve marco teórico

Os testes e resultados obtidos nas avaliações em larga escala têm sido analisados através de técnicas psicométricas, pois como afirmar Sartes e Souza-Formigoni (2013), instrumentos e testes construídos com base nessa técnica tem sido uma forma de avaliar objetivamente os fenômenos psicológicos. Segundo Pasquali (2009, p. 67) "a psicometria procura explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas, tipicamente chamadas de itens". Apesar de se fundamentar em uma base epistemológica eminentemente quantitativa, assumindo pressupostos da teoria da mensuração, a psicometria é considerada como um ramo da psicologia, ou seja, das ciências empíricas (Pasquali, 2009).

Dessa forma, a psicometria faz uso de testes, em que pesquisadores das ciências psicossociais utilizam-se destes para estimar o comportamento psicológico do sujeito, prevendo de certa forma o erro presente nessa medida (Muñiz, 1998). Para tanto, indagando sobre as condições para se obter uma medida adequada, o autor ressalta a observância de três características: a fiabilidade, denominada como o grau de precisão na medição do teste;

a validade, ou seja, a garantia que as inferências realizadas a partir da medição são corretas e; a fundamentação teórica em que o teste está embasado.

Entre as técnicas utilizadas para analisar os resultados temos a Teoria Clássica dos Testes (TCT), apresentados através de scores padronizados, utilizando-os como processo de seleção das pessoas (Valle, 2000). A TCT se preocupa com o resultado final através do somatório de itens de um teste, expresso no escore total, tendo interesse em produzir testes válidos e com qualidade (Pasquali, 2009). No entanto, a TCT não permite analisar os dados entre diferentes, assim é impossível comparar pessoas que não foram submetidas à mesma avaliação. Isso ocorre porque os resultados são dependentes dos itens, sendo que a análise é realizada em função do instrumento como um todo (Andrade, Tavares e Valle, 2000).

Para isso, foi desenvolvida a Teoria da Resposta ao Item (TRI), que “é um conjunto de modelos matemáticos que considera o item como unidade básica de análise” (Andrade, Laros e Gouveia, 2010). Dessa forma, a TRI não está interessada no escore total de um teste, mas sim em cada item de um teste, tendo interesse em produzir itens válidos e com qualidade (Pasquali, 2009). A vantagem da TRI, segundo Andrade et al., (2000), é que ela permite comparar populações diferentes, desde que o instrumento tenha alguns itens em comum, e mesmo indivíduos da mesma população ainda que submetidos a provas distintas. Segundo os mesmos autores alguns problemas em educação podem ser solucionados, como por exemplo, permite o acompanhamento de uma série ao longo dos anos, bem como a comparação de desempenho entre escolas.

No entanto, embora a TRI ofereça soluções para a dependência que os parâmetros dos itens têm da amostra, os parâmetros de TCT ainda são utilizados para a análise inicial da qualidade psicométrica dos itens. A partir disso, este estudo realiza uma análise dos parâmetros de teoria clássica dos itens com o objetivo de analisar os itens de Educação Física.

3. Método

3.1 Enfoque metodológico

A presente pesquisa se enquadra no viés quantitativo, pautada nos pressupostos da Teoria da Mensuração (Pasquali, 2009). Caracteriza-se como descritiva e exploratória, uma vez que, procura descrever um instrumento de avaliação (ENEM) através da exploração de sua estrutura estatística. Foram analisados os itens da prova de Linguagens e Códigos do ENEM, e de forma mais específica os parâmetros dos itens de Educação Física do Exame.

3.2 População e amostra

Os microdados dos resultados das provas de cada candidato do ENEM estão disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais “Anísio Teixeira” (INEP) em tabelas do SPSS (*Statistical Package for Social Sciences*) e são de livre acesso ao público. Neste estudo utilizaram-se os microdados disponíveis a partir de 2009, quando ocorreu mudança na metodologia de análise do desempenho dos candidatos.

A população é constituída pelos candidatos cearenses participantes do ENEM entre os anos de 2009 e 2014. Para este estudo, selecionaram-se os candidatos por amostragem aleatória simples. O tamanho da amostra foi estimado a partir da seguinte equação:

$$n = \left(\frac{Z_{\alpha/2}^2 \cdot 0,25}{E^2} \right)$$

Em que,

n = é o número de indivíduos na amostra;

$Z_{\alpha/2}$ = é o valor crítico que corresponde ao grau de confiança desejado;

α = nível de significância;

E = é a margem de erro ou Erro Máximo de Estimativa ($X^- - \mu$). Foi utilizado um erro de 5 % (0.05).

A partir disso foi obtido um tamanho amostral de 385 candidatos para todos os anos. Para estimar a normalidade dos dados utilizou-se o teste *Kolmogorov-Smirnov*. Foram considerados normais os dados que apresentaram valor de $p > 0.05$. Adotou-se como variável principal para o teste de normalidade a nota em Linguagens e Códigos, área que os itens de Educação Física estão contemplados. Com isso, as amostras apresentaram normalidade para a prova do ano de 2009 ($D(385)= 0.037$, $p = 0.200$), 2010 ($D(385)= 0.045$, $p = 0.063$), 2011 ($D(385)= 0.045$, $p = 0.063$), 2012 ($D(385)= 0.037$, $p = 0.200$), 2013 ($D(385)= 0.045$, $p = 0.057$) e 2014 ($D(385)= 0.036$, $p = 0.200$). A Tabela 1 apresenta as características das amostras de cada ano.

Tabela 1
Caracterização da amostra do Exame Nacional do Ensino Médio (ENEM), Brasil, 2009 a 2014.

Variáveis		2009		2010		2011		2012		2013		2014	
		n	%	n	%	n	%	n	%	n	%	n	%
Sexo	Feminino	236	61.3	238	61.8	216	56.1	220	57.1	215	55.8	210	54.5
	Masculino	149	38.7	147	38.2	169	43.9	165	42.9	170	44.2	175	45.5
Tipo de instituição	Pública	310	80.5	327	84.9	338	87.8	349	90.6	339	88.1	360	93.5
	Privada	75	19.5	58	15.1	47	12.2	36	9.4	46	11.9	25	6.5
Dependência Administrativa	Municipal	3	0.8	1	0.3	-	-	1	0.3	-	-	3	0.8
	Estadual	301	78.2	324	84.2	336	87.3	346	89.9	337	87.5	351	91.2
	Federal	5	1.3	2	0.5	2	0.5	2	0.5	2	0.5	-	-
	Privada	76	19.7	58	15.1	47	12.2	36	9.4	46	11.9	31	8.0
Localização	Rural	6	1.6	3	0.8	11	2.9	20	5.2	112	3.1	9	2.3
	Urbana	379	98.4	382	99.2	374	97.1	365	94.8	373	96.9	381	97.7
Necessidades especiais	Sim	1	0.3	1	0.3	2	0.5	4	1.0	6	1.6	4	1.0
	Não	384	99.7	384	99.7	383	99.5	381	99.0	379	98.4	381	99.0
Idade (média e desvio padrão)		19.03 (2.12)		18.99 (4.80)		18.60 (3.96)		18.66 (3.94)		18.63 (4.15)		18.91 (4.49)	

Fonte: Elaboração própria, 2018.

Foram excluídos os candidatos que não indicaram o sexo, dependência administrativa da escola, localização da escola (zona rural ou urbana) os que não estiveram presentes na prova de Linguagens e Códigos. Também se optou por selecionar os candidatos que responderam o caderno de prova de cor azul. Esse procedimento foi necessário para a obtenção de uma amostra de participantes com boa representatividade da realidade educacional brasileira.

3.3 Procedimentos

A elaboração e aplicação das provas do ENEM é realizada pelo INEP. O Exame é constituído de quatro provas referentes as áreas de Linguagens e Códigos, Matemática, Ciências da Natureza e Ciências Humanas. Cada prova contém 45 itens objetivos de múltipla escolha com cinco alternativas, em que apenas uma é correta. Foram objetos desse estudo apenas os itens referentes ao conteúdo de educação física inseridos na prova de Linguagens e Códigos. As provas foram aplicadas em formato impresso em dois dias consecutivos (sábado e domingo).

3.4 Análise dos dados

Para a análise dos dados, de início foi utilizado os recursos da Análise Fatorial Exploratória pelo método da Fatoração dos Componentes Principais para verificar a unidimensionalidade da prova de Linguagens e Códigos, a qual contém os itens de Educação Física. Para tanto, analisou-se primeiro a adequação da amostra de variáveis (itens) através do teste de *Kaiser-Meyer-Olkin* (KMO), esfericidade de *Bartlett* (BTS). Considerou-se como adequado os valores de $KMO \geq 0.70$ e $BTS \leq 0.05$. Além disso, foi analisado a adequação de cada item através da Correlação de Pearson (r), sendo considerado adequado o item que apresentou $r \geq 0.50$. Em seguida solicitou-se a variância acumulada considerando um fator, uma vez que a prova é elaborada de forma a constituir uma dimensão. Também se utilizou como parâmetro de análise da estrutura fatorial da prova o gráfico *Scree Plot*. Na análise da adequação de cada item ao fator considerou-se itens com comunalidade e cargas fatoriais acima de 0.4.

Também foi realizado o cálculo de consistência interna (α de *Cronbach*) para a análise da fidedignidade da prova e o T^2 de *Hotelling* (T^2) para a existência de efeito halo (Field, 2009; Hair, Black, Babin, Anderson, e Tatham, 2005; Pasquali, 2009). Foram considerados adequados os valores $\alpha \geq 0.7$ e T^2 com $p \leq 0.05$.

Em seguida foram analisados os parâmetros de dificuldade e discriminação dos itens, sendo este último realizado através da correlação bisserial por ponto (r_{bp}), uma vez que os itens foram dicotomizados em certo e errado (Pasquali, 2009). Os valores de dificuldade foram classificados da seguinte forma: muito fácil de 0 a menos de 0.21; fácil de 0.21 a menos de 0.41; médio de 0.41 a menos de 0.61; difícil de 0.61 a menos de 0.81 e; muito difícil de 0.81 a 1.0. Para o segundo parâmetro, foram considerados discriminativos os itens com $r_{bp} \geq 0.20$. As seguintes análises foram conduzidas com o *software SPSS*, versão 20.0.

4. Resultados

4.1 Análise da prova: Análise Fatorial Exploratória, Consistência Interna e Sensibilidade

Inicialmente foram analisados os valores de correlação dos itens de Linguagens e Códigos, com destaque para os itens de Educação Física através de uma matriz de anti-imagem para verificar sua adequabilidade para compor a prova. Posteriormente verificados os valores de comunalidade e as cargas fatoriais dos itens considerando que o instrumento apresenta um fator. A correlação apresentou adequabilidade de todos os itens de Educação

Física, com valores de r variando entre 0.54 a 0.89. Embora, alguns itens apresentaram valores abaixo de 0.50, o que inviabilizaria a continuação da análise fatorial, sendo necessária a exclusão de alguns itens para se obter adequabilidade do conjunto de itens para a análise fatorial. As comunalidades dos itens foram baixas, variando entre 0.04 e 0.30 e cargas fatoriais entre 0.20 e 0.54, com exceção do item 97 de 2014. Os valores referentes aos itens de Educação Física estão dispostos na Tabela 2.

Tabela 2
Valores de adequação dos itens, Exame Nacional do Ensino Médio (ENEM), Brasil, 2009 a 2014.

Ano	Item	r^*	Comunalidade	Cargas Fatoriais
2009	103	0.76	0.13	0.35
	134	0.54	0.04	0.20
2010	106	0.81	0.12	0.35
	110	0.88	0.30	0.54
	120	0.69	0.07	0.26
	96	0.89	0.22	0.47
2011	105	0.87	0.30	0.39
	108	0.79	0.08	0.28
	133	0.89	0.23	0.47
	96	0.73	0.11	0.33
2012	100	0.61	0.04	0.20
	115	0.74	0.11	0.34
	97	0.72	0.07	0.26
2013	108	0.83	0.21	0.46
	111	0.78	0.11	0.32
	97	0.54	0.01	0.08
2014	98	0.72	0.18	0.42

Nota: r = Correlação na matriz de anti-imagem.

Fonte: Elaboração própria, 2018.

Na Tabela 3 estão os valores referentes à adequação da prova de Linguagens e Códigos para a realização da análise fatorial. São apresentados os valores de KMO, testes de esfericidade de Bartlett e a variância explicada com um fator. O teste de KMO apresentaram valores adequados, com exceção das provas de 2012 e 2014, mas permanecendo em níveis aceitáveis. O teste de esfericidade apresentou significância em todos os anos. Esses valores indicam a adequação da amostra de itens para a realização da análise fatorial. Ao realizar a análise solicitando a extração de um fator, obteve-se uma variância explicada muito baixa (7.80 a 14.71).

Tabela 3
Valores de adequação do teste, Exame Nacional do Ensino Médio (ENEM), Brasil, 2009 a 2014.

Ano	KMO	Teste de esfericidade de Barlett's (X^2)	Variância explicada (%)
2009	0.77	2116.96*	11.94
2010	0.81	2431.65*	13.48
2011	0.84	2568.24*	14.71
2012	0.67	1654.79*	8.79
2013	0.76	1917.90*	10.96
2014	0.66	1859.68*	7.80

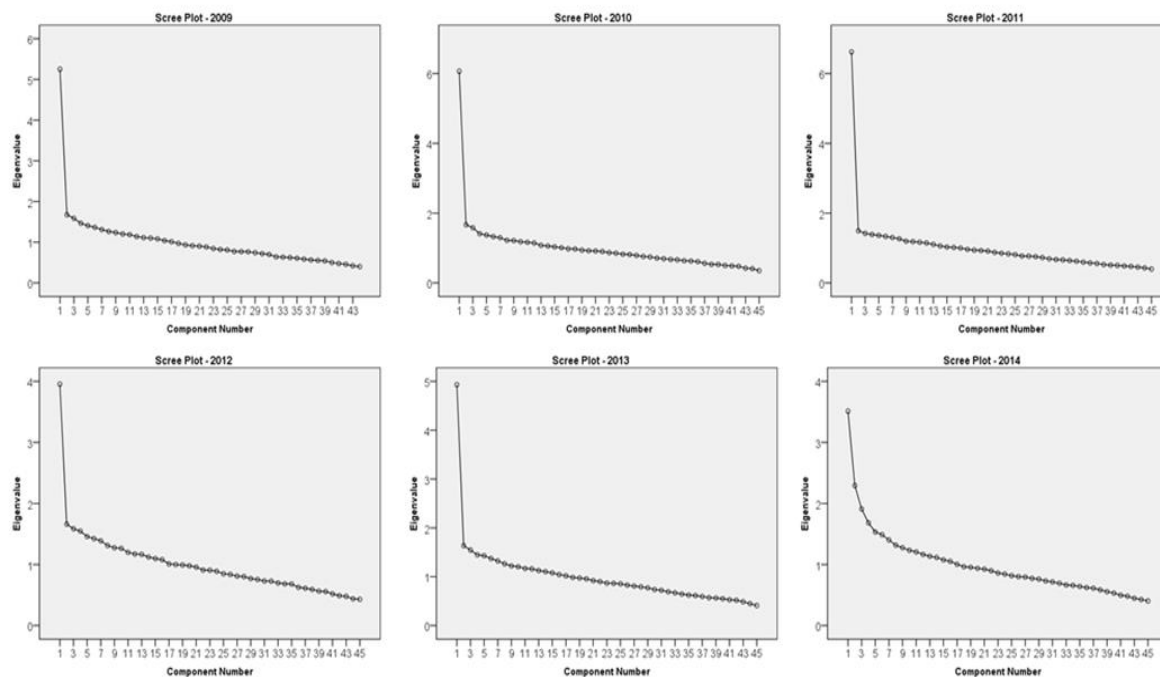
Nota: KMO = Coeficiente Kaiser-Mayer-Olkin. X^2 = Valor qui-quadrado.

* $p \leq 0.001$ significância do teste.

Fonte: Elaboração própria, 2018.

A análise gráfica indica que os testes são unifatoriais, apesar da baixa variância explicada do fator das provas. A exceção foi a prova de 2014 que apresentou uma estrutura fatorial insatisfatória. A Figura 1 mostra o gráfico *scree plot* que apresenta a relação dos componentes e os autovalores.

Figura 1
Scree plot das provas do Exame Nacional do Ensino Médio (ENEM), Brasil, 2009 a 2014.



Fonte: Elaboração própria, 2018.

Na Tabela 4 são apresentados os valores de confiabilidade da prova. A sensibilidade das provas apresentou valores de 0.85 a 0.99 e erro de medida entre 2.90 e 3.19. O teste T^2 de *Hottelling* apresentou valores satisfatórios, indicando a inexistência de efeito de halo. Com relação a precisão da prova, apenas os anos de 2009 a 2011 tiveram valores adequados. Ressalta-se a baixa precisão ($\alpha = 0.53$) da prova de 2014. De modo geral, os itens não influenciam a precisão caso sejam excluídos.

Tabela 4
Valores de confiabilidade das provas do Exame Nacional do Ensino Médio (ENEM), Brasil, 2009 a 2014.

Ano	Sensibilidade	T^2 da prova	α	Item	Correlação Item-Total	α se o item for excluído
2009	0.98	$T^2 = 4848.22$ $F = 100.45$ $p \leq 0.05$	0.80	103	0.29	0.79
				134	0.15	0.80
2010	0.98	$T^2 = 1443.04$ $F = 29.12$ $p \leq 0.05$	0.82	106	0.28	0.82
				110	0.45	0.82
				120	0.22	0.82
2011	0.99	$T^2 = 1034.75$ $F = 20.88$ $p \leq 0.05$	0.85	96	0.40	0.85
				105	0.33	0.85
				108	0.24	0.85
				133	0.41	0.85
2012	0.94	$T^2 = 1101.70$ $F = 22.23$ $p \leq 0.05$	0.72	96	0.24	0.72
				100	0.17	0.72
				115	0.24	0.71
2013	0.97	$T^2 = 1166.33$ $F = 23.54$ $p \leq 0.05$	0.79	97	0.21	0.79
				108	0.36	0.78
				111	0.25	0.79
2014	0.85	$T^2 = 2405.62$ $F = 48.63$ $p \leq 0.05$	0.53	97	0.03	0.53
				98	0.21	0.51

Nota: T^2 = Coeficiente T^2 de Hotelling. α = Coeficiente Alfa de Cronbach.

Fonte: Elaboração própria, 2018.

4.2 Análise dos Itens: dificuldade e discriminação

O índice de dificuldade e discriminação dos itens estão apresentados na Tabela 5. A dificuldade dos itens variou entre 0.28 e 0.88. Quanto a discriminação, a maioria apresenta discriminação acima de 0.20, com exceção do item 100 do ano de 2012 e dos dois itens de 2014. Ambos os itens deste ano apresentaram dificuldade alta. Além disso, o item de 2014 teve discriminação inversa.

Tabela 5
Dificuldade e discriminação dos itens das provas do Exame Nacional do Ensino Médio (ENEM),
Brasil, 2009 a 2014.

Ano de Aplicação	Item	d	r_{bp}
2009	103	0.28	0.37
	134	0.66	0.20
2010	106	0.67	0.34
	110	0.32	0.58
	120	0.54	0.28
2011	96	0.57	0.45
	105	0.49	0.38
	108	0.66	0.27
	133	0.41	0.47
2012	96	0.53	0.36
	100	0.72	0.17
	115	0.64	0.35
2013	97	0.84	0.24
	108	0.40	0.49
	111	0.39	0.34
2014	97	0.88	0.00
	98	0.78	-0.06

Nota: d = Dificuldade do item. r_{bp} = Correlação ponto-biserial.

Fonte: Elaboração própria, 2018.

5. Discussão

5.1 Discussão da análise da prova

A análise estatística da prova objetivou conhecer as características métricas do conjunto de itens de Linguagens e Códigos. As análises iniciais de adequação da amostra de itens de Educação Física apresentaram valores parcialmente insatisfatórios (<0.50), ou seja, o conjunto de itens não apresentou adequabilidade para a realização da análise fatorial exploratória. A correlação entre os itens mostra o quanto cada item está adequado para compor a prova. Na análise os itens apresentaram valores adequados (Field, 2009), ou seja, nenhum item de Educação Física se mostrou inapropriado, inicialmente, para fazer parte da prova. Considerando que o ENEM é uma avaliação que tem grandes repercussões sociais, sendo necessários resultados confiáveis, essa primeira análise mostrou deficiência no instrumento.

Mesmo com alguns itens com valores de correlação abaixo do aceitável, o teste de KMO apresentou valores aceitáveis, acima de 0.7, embora considere-se ótimo os valores acima de 0.8 e com significância no teste de esfericidade (Field, 2009, Pontes Junior,

Soares, Almeida e Trompieri Filho, 2014). Essas medidas são importantes, pois indicam a adequação do conjunto de itens para compor o instrumento e prosseguir com análise fatorial.

Após analisar a adequação da amostra de itens, procedeu-se com a análise fatorial. Ressalta-se, que nesse momento optou-se por solicitar análise com a extração de um fator, uma vez que o Inep estima a proficiência dos candidatos via TRI, a qual tem como pressuposto a unidimensionalidade dos itens. O objetivo foi identificar se os itens atendiam a esse critério. Nessa análise, obteve-se variância explicada muito baixa considerando um fator, todos abaixo de 15%. Pontes et al. (2014) destaca a necessidade de os fatores do instrumento explicarem pelo menos 50% da variância. Hair et al. (2005) indica que muitos autores apresentam a necessidade de pelo menos 60% de explicação da variância. No caso da prova de Linguagens e Códigos do ENEM, levando em conta que a organizadora analisa os resultados pela técnica da TRI, portando assumindo a unidimensionalidade dos itens, a variância dos itens explicada pelo modelo unidimensional não foi maior que 20%, valor considerado abaixo do esperado (Hair et al., 2005). Isso se torna um problema porque o teste assume que o desempenho de estudantes é representado por uma variável latente.

Pasquali (2009) ressalta a dificuldade de se obter a unidimensionalidade, uma vez que, os aspectos cognitivos dos seres humanos são multifacetados e multideterminados. Tavares (2013) questiona o fato de avaliações educacionais pretenderem medir um fator unidimensional, uma vez que, é consensual entre educadores pesquisadores que o ser humano é determinado por vários fatores. Mesmo assim, Pasquali (2009) propõe ser suficiente a existência de um fator dominante para garantir esse pressuposto. Em modo complementar, Vitória, Almeida e Primi (2006) afirmam, com base em pesquisas, que a dimensionalidade (os diferentes graus) pouco afeta os parâmetros dos itens e os resultados de testes.

Além disso, ao extrair um fator, os itens apresentam variâncias comuns muito baixas, não passando de 0.30, enquanto que o mínimo aceitável é 0.40, acontecendo o mesmo com as cargas fatoriais dos itens, em que muitos não atingiram o valor mínimo (Pontes Junior et al., 2014). Apesar dos valores inadequados, a análise gráfica através do *scree plot* mostra a unidimensionalidade dos dados ao observar o ponto de inflexão, com exceção dos dados de 2014 que apresentaram uma estrutura inadequada.

No que se refere à precisão da prova utilizou-se o coeficiente *Alpha de Cronbach*, que é uma das técnicas mais utilizadas para a análise da confiabilidade de um instrumento (Cunha, Almeida Neto e Stackfleth, 2016). Com exceção das provas de Linguagens e

Códigos de 2009 e 2011 com valores α acima de 0.80, considerados excelentes, o demais tiveram valores inferiores, embora aceitáveis. Cabe destacar a baixa precisão da prova de 2014, que teve o coeficiente α de 0.53. De modo geral, os itens de Educação Física não influenciaram na precisão do instrumento, o que é considerado bom.

A precisão ou fidedignidade de um instrumento de medida é central na sua validação. Uma boa precisão é uma garantia da confiabilidade do teste. Quando se trata de testes para fins de seleção, sua importância aumenta, já que pode influenciar nos resultados.

A fidedignidade ou precisão pode ser afetada por uma série de fatores, como afirma Vianna (1976). Segundo o autor, a precisão do teste aumenta quanto maior for o número de itens, desde que com índice de correlação suficiente com os demais itens do teste. Nesse quesito, o ENEM não poderia apresentar problemas, uma vez que cada área do exame, no caso deste estudo a área de Linguagens e Códigos, é constituída de 45 itens, o que é considerado um grande número. Além do mais, o exame é alvo de muitas críticas no sentido de diminuir o número de itens e assim diminuir o tempo de aplicação da prova.

Outro fator que pode influenciar na precisão da prova é a amplitude de dificuldade dos itens (Vianna, 1976). Quanto menor a amplitude maior a fidedignidade. Considerando apenas os itens de Educação Física de 2014, observa-se que os itens apresentam baixa dificuldade, fato também observado ao considerar os demais itens. Isso pode ter contribuído para o baixo índice alpha no exame deste ano.

5.2 Discussão da análise dos itens

A análise dos parâmetros de dificuldade e discriminação dos itens a partir da TCT objetivou observar as características estatísticas de adequação dos itens ao realizar medidas. A dificuldade apresentou índices adequados, ou seja, entre 0.20 e 0.80, com exceção do item 97 de 2014. Quanto à discriminação, a maioria permaneceu acima de 0.20. No entanto, os itens de Educação Física da prova de 2014 apresentaram discriminação muito baixa, ou seja, são itens ruins para avaliar, uma vez que não conseguem diferenciar os sujeitos de baixa habilidade dos de alta habilidade.

O índice de dificuldade é um parâmetro simples de um teste. Neste caso, como coloca Vianna (1976), poderia ser chamado de índice de facilidade, uma vez que se utiliza da proporção de acertos. Dessa forma, quanto maior a proporção mais fácil é o item. No caso dos itens de Educação Física analisados, a maioria apresentou índices adequados.

Vale ressaltar que esse parâmetro depende do particular conjunto de sujeitos que constituem a amostra avaliada (Laros, 2009). Dessa forma, o autor resalta que os sujeitos habilidosos a dificuldade vai aumenta. Do mesmo modo, a dificuldade do item será baixa se os sujeitos não forem tão habilidosos. Entretanto, destaca-se que esse problema desaparece se a amostra for representativa da população, em que qualquer amostra aleatória apresentará os mesmos índices para os itens (Laros, 2009). No caso deste estudo em particular, acredita-se ter superado esse problema, considerando que o pressuposto da normalidade foi atendido para a amostra de todos os anos.

Por outro lado, quanto à discriminação, os itens de 2014 apresentaram baixo índice. Esses mesmos itens apresentaram dificuldade elevada. Alguns autores (Pasquali, 2009; Vianna, 1976) ressaltam que itens com dificuldade muito alta ou muito baixa apresentam discriminação inadequada, pois não conseguem distinguir sujeitos de habilidades distintas.

Nesse estudo, utilizou-se da correlação bisserial por ponto, uma vez que, segundo Silveira (1983) é o melhor indicador da discriminação de um item. Vários estudos que se utilizam da TCT para analisar testes utilizam-se desse parâmetro.

No entanto, Laros (2009) considera inadequado para o cálculo da discriminação por apresentar problemas teóricos. Ressalta ser incoerente avaliar um parâmetro de um item através da correlação do item com o escore total do teste, uma vez que os outros itens ainda não foram testados. Destaca ainda a necessidade de os itens do teste apresentar unidimensionalidade para que a discriminação seja consistente. Além do mais, indica que o parâmetro é falho quando se tem itens muito fáceis ou muito difíceis.

6. Conclusões

A prova de Linguagens e Códigos apresenta problemas de dimensionalidade, considerada pressuposto fundamental para validade dos testes nos modelos de resposta ao item utilizado no exame. Apesar de a análise gráfica indicar o cumprimento desse indicador, baixa explicação foi demonstrada. Isso compromete a medida realizada pelo instrumento. Quanto à precisão ou fidedignidade da prova apenas em 2014 apresentou baixo índice. Nesses aspectos os itens de Educação Física permaneceram adequados.

No geral, os itens da área apresentaram valores adequados de dificuldade e discriminação a partir da TCT, com exceção da prova de 2014, em que apresentaram valores inadequados, o que pode ter influenciado nos valores de fidedignidade e unidimensionalidade. Esses índices indicam baixa qualidade dos itens de Educação Física.

As análises empreendidas são dificultadas pela baixa representatividade dos itens de Educação Física, já que poucos itens são contemplados no exame. Esse fato compromete também a avaliação dos conteúdos da disciplina, sendo difícil obter um diagnóstico da aprendizagem nessa área com um número limitado de questões.

Ademais, ressaltamos que os problemas apresentados no exame podem comprometer a validade dos resultados, principalmente a não adequação ao pressuposto de unidimensionalidade dos itens que compõe o exame. Esse pressuposto é essencial para que os resultados sejam confiáveis e dessa forma possam oferecer uma medida justa do rendimento dos participantes na prova.

Diante disso, levando em consideração que o ENEM passou a ser analisado a partir da TRI em substituição a TCT, levanta-se o seguinte questionamento: A TRI trouxe avanços nas medidas educacionais em relação a TCT? Dessa forma, indica-se a realização de pesquisas que comparem os parâmetros das provas e dos itens dos exames educacionais, de forma a possibilitar informações que forneçam indicadores que proporcionem a comparação das técnicas de análise utilizadas pelas duas teorias, de modo a testar os supostos avanços pretendidos pela TRI.

Referências

- Andrade, Dalton Francisco de; Tavares, Héilton Ribeiro e Valle, Raquel da Cunha. (2000). *Teoria de Resposta ao Item: Conceitos e Aplicações*. ABE – Associação Brasileira de Estatística. Recuperado de https://docs.ufpr.br/~aanjos/CE095/LivroTRI_DALTON.pdf
- Andrade, Josemberg Moura; Laros, Jacob Arie e Gouveia, Valdiney Veloso. (2010). O uso da teoria da resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, 9(3), 421-435. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300009
- Beltrão, José Arlén. (2014). A educação física na escola do vestibular: as possíveis implicações do ENEM. *Movimento*, 20(2), 819-840. doi: <https://doi.org/10.22456/1982-8918.41801>
- Cunha, Cristiane Martins; Almeida Neto, Omar Pereira e Stackfleth, Renata Stackfleth. (2016). Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Rev. Aten. Saúde*, 14(49), 98-103. doi: <https://doi.org/10.13037/ras.vol14n49.3671>
- Fernandes, Anael, Rodrigues, Heitor Andrade, e Nardon, Tiago Aparecido. (2013). A inserção dos conteúdos de educação física no ENEM: entre a valorização do componente curricular e as contradições da democracia. *Motrivivência*, (40), 13-24. doi: <https://doi.org/10.5007/2175-8042.2013v25n40p13>

- Field, Andy. (2009). *Descobrimdo estatística usando o SPSS*. (L. Vialli, Trad.). Porto Alegre: ArtMed.
- Freitas, Dirce Nei Teixeira. (2013). Avaliação da educação básica no Brasil: origens e pressupostos. Adriana Bauer, Bernadette Angelina Gatti, e Marialva Tavares (Orgs.), *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos* (pp. 147-176). Florianópolis: Insular.
- Gatti, Bernadette Angelina. (2013). Possibilidades e fundamentos de avaliação em larga-escala: primórdios e perspectivas contemporâneas. In Adriana Bauer, Bernadette Angelina Gatti, e Marialva Tavares (Orgs.), *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos* (pp. 47-69) Florianópolis: Insular.
- Governo do Brasil, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (2013). *Exame Nacional do Ensino Médio (Enem): relatório pedagógico 2009-2010*. Brasília, Brasil: O Instituto.
- Governo do Brasil, Ministério da Educação. (2000). *Parâmetros curriculares nacionais: ensino médio*. Brasília, Brasil: Ministério da Educação.
- Hair, Joseph F., Black, William C., Babin, Barry J., Anderson, Rolph E., and Tatham, Ronald L. (2005). *Análise multivariada de dados* (5a ed.). Porto Alegre: Bookman.
- Laros, Jacob Arie. (2009). Análise gráfica de itens. In: Pasquali, Luis (Org.). *Psicometria: teoria dos testes na psicologia e na educação* (Cap. 5; 3a ed.). Petrópolis: Vozes.
- Muñiz, José. La medición de lo psicológico. (1998). *Psicothema*, 10(1), 1-21. Recuperado de <http://www.psicothema.com/psicothema.asp?id=138>
- Pasquali, Luis. (2009). *Psicometria: teoria dos testes na psicologia e na educação* (3a. ed.) Petrópolis: Vozes.
- Pontes Junior, José Airton de Freitas; Soares, Edson Silva; Almeida, Leandro Silva e Trompieri Filho, Nicolino. (2014). Análise fatorial exploratória e alpha de Cronbach: elementos iniciais na validação de instrumentos de avaliação educacional. *Educação & Linguagem*, 1(1), 63-75. Recuperado de <https://www.fvj.br/revista/wp-content/uploads/2014/12/5Artigo1.pdf>
- Sartes, Laisa Marcorela Andreoli e Souza-Formigoni, Maria Lúcia Oliveira. (2013). Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, 26(2), 241-250. doi: <http://dx.doi.org/10.1590/S0102-79722013000200004>
- Silveira, Fernando Lang. (1983). Considerações sobre o índice de discriminação de itens em testes educacionais. *Educação & Seleção*, (07). Recuperado de <http://publicacoes.fcc.org.br/ojs/index.php/edusel/article/view/2546>

- Souza Junior, Osmar Moreira; Diniz, Irla Karla Santos; Ditomaso, Aline e Darido, Suraya Cristina. (2012, outubro). Educação física no ENEM: análise das questões à luz dos PCN's. *Anais do II Congresso Internacional de Educação Física, Esporte e Lazer*. São Carlos, Brasil. Recuperado de <http://motricidades.org/conference/index.php/cpqmh/5cpqmh/paper/viewFile/230/140>
- Tavares, Cristina Zukowsky. (2013). Teoria de resposta ao item: uma análise crítica dos pressupostos epistemológicos. *Est. Avali. Educ.*, 24(54), 56-76. doi: <http://dx.doi.org/10.18222/eae245420131902>
- Valle, Raquel da Cunha. (2000). Teoria de resposta ao item. *Est. Aval. Educ.*, (21), 07-92. doi: <http://dx.doi.org/10.18222/eae02120002225>
- Vianna, Heraldo Marelim. (1976). *Testes em educação*. São Paulo: IBRASA.
- Vitoria, Florbela; Almeida, Leandro Silva e Primi, Ricardo. (2006). Unidimensionalidade em testes psicológicos: conceito, estratégias e dificuldades na sua avaliação. *Psic*, 7(1), 01-07. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1676-73142006000100002

Revista indizada en



Distribuida en las bases de datos:

