

UMA PROPOSTA DE ANÁLISE DE ITENS DA PROVA PREPARATÓRIA PARA O ENADE APLICADA AOS DISCENTES DE ENGENHARIA CIVIL DO IFMG – CAMPUS AVANÇADO PIUMHI¹

Milene Cunha Costa²
Samuel Henrique Oliveira Lima
Denilson Junio Marques Soares

RESUMO

Considerando a necessidade cada vez maior de se produzir avaliações consistentes, capazes de fornecer informações válidas sobre seus respondentes, este artigo tem o objetivo de analisar dois provões aplicados aos discentes do curso de Bacharelado em Engenharia Civil do Instituto Federal de Minas Gerais – *Campus* Avançado Piumhi e comparar seus resultados. Trata-se de uma avaliação aplicada semestralmente aos estudantes, visando a oferecer-lhes um preparo para o Exame Nacional de Desempenho de Estudantes (ENADE). Para isto, foram considerados três índices psicométricos provenientes da Teoria Clássica dos Testes: dificuldade, discriminação e confiabilidade. Também foram realizados Testes de hipóteses para verificar se estes provões apresentaram resultados equiparáveis. Estas análises foram realizadas com o auxílio do software estatístico R. Os resultados demonstram que os valores de referência, dispostos na literatura, não foram, em sua maioria, atingidos, apontando para a necessidade de uma revisão sistemática no processo de elaboração dessas provas e de um debate sobre os diferentes resultados obtidos por estudantes matriculados em períodos distintos no curso.

Palavras-chave: Teoria Clássica dos Testes (TCT). Engenharia Civil. *Software* R.

A PROPOSAL FOR ANALYSING THE ITEMS OF THE ENADE PREPARATORY ASSESSMENT APPLIED TO CIVIL ENGINEERING STUDENTS OF INSTITUTO FEDERAL MINAS GERAIS - IFMG - CAMPUS AVANÇADO PIUMHI

ABSTRACT

Considering the increasing need to produce consistent assessments able to provide valid information about your respondents, this article has the objective of analyzing two tests applied to students of the Bachelor's Degree in Civil Engineering course of the IFMG – *Campus* Avançado Piumhi (Federal Institute of Minas Gerais - Piumhi Advanced Campus) and compare your results. This is an evaluation applied to the students every semester, aiming to offer them a preparation for The National Assessment of Students Performance (ENADE). For this, it were considered three psychometric indexes from the Classical Theory of Tests: difficulty, discrimination and reliability. Statistical hypothesis testing were also performed

¹**Como citar este artigo:** COSTA, M. C.; LIMA, S. H. O.; SOARES D. J. M. Uma proposta de análise de itens da prova preparatória para o Enade aplicada aos discentes de engenharia civil do IFMG – *Campus* Avançado Piumhi. *ForScience*, Formiga, v. 8, n. 1, e00706, jan./jun. 2020. DOI: 10.29069/forscience.2020v8n1.e706.

² **Autor para correspondência:** Milene Cunha Costa, e-mail: milencunha.eng@gmail.com

to verify whether these tests had comparable results. These analyzes were carried out with the aid of statistical software R. The results show that the reference values, presented in the literature, were mostly not reached, pointing toward the need for a systematic review in the elaboration of these tests, besides a debate on the different results obtained by students enrolled in different periods of the course.

Keywords: Classical Theory of Tests (CTT). Civil Engineering. *Software R*.

1 INTRODUÇÃO

Conforme a Resolução nº 01 de 17 de junho de 2010 da Comissão Nacional de Avaliação da Educação Superior (CONAES), “o Núcleo Docente Estruturante (NDE) de um curso de graduação constitui-se de um grupo de docentes, com atribuições acadêmicas de acompanhamento, atuante nos processos de concepção, consolidação e contínua atualização do projeto pedagógico do curso” (BRASIL, 2010).

Fundamentado nesses princípios, em março de 2017, em reunião desenvolvida pelo NDE do Curso de Bacharelado em Engenharia Civil do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) - *Campus* Avançado Piumhi, foi sugerido por um dos integrantes, que fossem aplicados provões, uma vez por semestre, que abrangessem todas as disciplinas pertinentes à grade curricular do curso, com finalidade de reduzir a retenção, fazer avaliação interdisciplinar, diminuir a evasão e, por consequência, preparar o discente para o Exame Nacional de Desempenho de Estudantes (ENADE).

A proposta foi aprovada em ata, ficando ainda acordado que o provão seria “composto por 30 (trinta) questões de múltipla escolha, distribuídas por área” (IFMG, 2017, p.2). Também foi sugerido que fosse atribuída uma bonificação à nota final do estudante em cada disciplina matriculada no semestre corrente. Assim, o estudante poderia aumentar sua nota final em até 5%, de acordo com o seu desempenho no provão. Por exemplo, o estudante que obtivesse 18 acertos, teria suas notas finais aumentadas em 3% $\left(\frac{18}{30} \times 5\% = 3\%\right)$. Com duração de quatro horas, a primeira aplicação da respectiva avaliação ocorreu em julho de 2017 e, atualmente, encontra-se em sua quarta edição.

Considerando a importância do provão para o processo de ensino-aprendizagem no âmbito do IFMG - *Campus* Avançado Piumhi e a necessidade de aprimorar a dinâmica da avaliação dos resultados desses provões, partindo do pressuposto que eles podem fornecer características consistentes sobre os estudantes que os realizam, este artigo tem o objetivo de

oferecer uma análise estatística das duas últimas edições do provão. Para isto, utiliza-se de uma vertente clássica da Psicometria: a Teoria Clássica dos Testes (TCT).

Dessa forma, discute-se sobre o processo de elaboração desses provões, analisando algumas características psicométricas das questões que o constituem, além da confiabilidade de seus resultados. Também será realizada uma análise comparativa entre eles. Estas análises poderão ser utilizadas, por exemplo, na seleção das melhores questões para a futura composição de um banco de questões, que poderá auxiliar os docentes do curso, produzindo informações mais consistentes acerca dos seus respondentes.

Outros trabalhos dessa natureza podem ser encontrados em Soares (2018) e Vilarinho (2015). O primeiro analisa, com o auxílio da TCT, uma avaliação de matemática básica elaborada pelos professores do departamento de matemática da Universidade Federal de Viçosa (UFV), a fim de proporcionar uma intervenção pedagógica capaz de reverter a reprovação nas Disciplinas de Cálculo. O segundo analisa os resultados da primeira fase da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) de 2014, em uma escola de uma cidade satélite do Distrito Federal.

2 REFERENCIAL TEÓRICO

2.1 Teoria Clássica dos Testes (TCT)

A TCT é proveniente da Psicometria, área que se dedica ao estudo e elaboração de testes de avaliação psicológica e ao desenvolvimento e aplicação de conhecimentos estatísticos e outros processos matemáticos na área da psicologia.

Segundo Pasquali (2013), de um modo geral, a Psicometria visa explicar o comportamento dos indivíduos perante as respostas em um determinado conjunto de questões, comumente chamadas de itens. Uma das vertentes mais utilizadas para análise de estrutura de testes é a TCT.

Na TCT, consolidada no início da década de 1900, a análise está centrada na seleção dos melhores itens de um teste, avaliando duas características do mesmo: a dificuldade e a discriminação.

2.1.1 Índice de dificuldade

Segundo Borgatto e Andrade (2012), o índice de dificuldade de um item é determinado por sua porcentagem de acerto, sendo que, quanto menor esta porcentagem, maior será o grau de dificuldade. Matematicamente,

$$D_i = \frac{A}{N}$$

em que D_i representa o nível de dificuldade, A representa a quantidade de respondentes que acertaram o item e N representa a quantidade total de respondentes.

Para que um item tenha um nível de dificuldade considerado ideal, Pasquali (2013) recomenda uma distribuição simétrica entre os níveis de dificuldades de itens no teste. A Tabela 1 apresenta como estes itens podem ser classificados.

Tabela 1 - Classificação do item, em relação ao índice de dificuldade

Classificação	Índice de dificuldade
Fácil	$D_i > 0,7$
Moderado	$0,3 < D_i \leq 0,7$
Difícil	$D_i \leq 0,3$

Fonte: Pasquali (2013).

2.1.2 Índice de discriminação

De acordo com Borgatto e Andrade (2012) o índice de discriminação (ID) verifica para determinado item, as porcentagens de acertos dos grupos de indivíduos com melhor (grupo superior) e com pior (grupo inferior) desempenho. Um item de boa qualidade deve apresentar diferenças significativas entre as médias de acertos entre estes dois grupos.

Um procedimento sugerido por Kelley (1939), conhecido como “regra 27”, consiste em dividir os indivíduos em dois grupos, sendo o primeiro, formado pelos 27% com piores notas e por um segundo, pelos 27% com melhores notas, conforme a Figura 1.

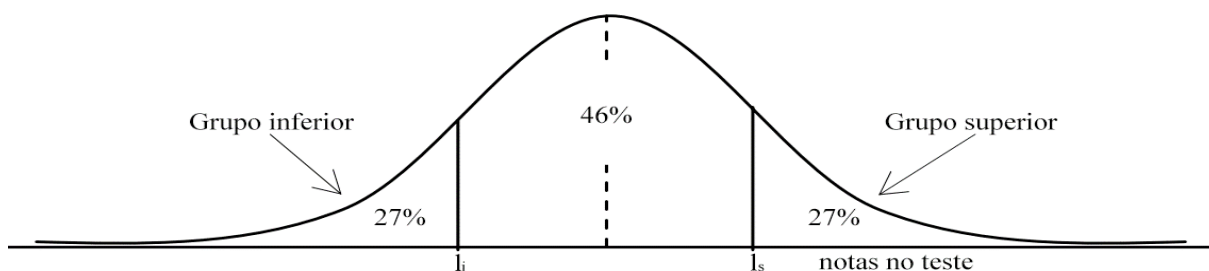


Figura 1- Método de Kelley para a seleção de grupos
Fonte: Kelley (1939).

Segundo Pasquali (2013), o índice de discriminação (ID) de um item é calculado por meio da diferença entre os índices de dificuldades calculados separadamente para os grupos superior e inferior, respectivamente.

Existem outras maneiras de verificar se um item é discriminativo ou não, como por exemplo, o coeficiente de correlação ponto-biserial, que trata de uma reformulação do coeficiente de correlação de Pearson, para o caso em que uma das variáveis analisadas é dicotômica, porém, neste trabalho utilizou-se apenas o ID.

Segundo Arias, Lloreda e Lloreda (2006), o parâmetro para a classificação do item do teste, de acordo com o índice de discriminação, é descrito pela Tabela 2.

Tabela 2 - Classificação do item, em relação ao índice de discriminação

Valores	Classificação
$ID \leq 0,20$	Ineficiente. Sugere-se eliminá-lo ou revisá-lo totalmente.
$0,20 < ID \leq 0,30$	Necessita ser revisado.
$0,30 < ID \leq 0,40$	Aceitável, não sendo necessária uma revisão.
$ID > 0,40$	Satisfatório. Deve permanecer no teste.

Fonte: Arias, Lloreda e Lloreda (2006).

2.2 Coeficiente de alfa de Cronbach

Para mensurar a precisão de um teste, considerando a TCT, é usual utilizar o coeficiente alpha de Cronbach. Um teste preciso, quando aplicado a um mesmo grupo de respondentes em ocasiões diferentes, produzirá os mesmos resultados. Dessa forma, a precisão de um teste garante que os resultados obtidos são confiáveis.

Segundo Cortina (1993), o coeficiente alpha de Cronbach foi descrito em 1951 por Lee J. Cronbach, sendo definido pela seguinte equação:

$$\alpha = \left(\frac{n}{n-1} \right) \cdot \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right)$$

em que n é o número total de itens (questões) do teste, σ_i^2 é a variância relacionada a cada questão do teste, e o σ_t^2 é a variância das notas finais obtidas pelos respondentes.

Como regra prática, valores acima de 0,6 (MALHOTRA, 2013) ou 0,7 (HAIR JÚNIOR *et al.*, 2009) são considerados como satisfatórios para a confiabilidade de um teste.

2.3 Teste de Hipóteses

Um teste de hipóteses é um procedimento estatístico que permite aceitar ou rejeitar uma determinada hipótese, utilizando os dados amostrados. Para a realização de um teste de hipótese é necessário: enunciar a hipótese de nulidade (H_0) e uma hipótese alternativa (H_a); fixar um nível de significância (α) a ser adotado; determinar a região crítica do teste, que representa a faixa de valores que levam à rejeição de H_0 , com o auxílio de tabelas estatísticas apropriadas; calcular a estatística do teste e concluir. Caso esta estatística pertença à região crítica, rejeita-se H_0 , caso contrário, não rejeita-se H_0 .

Os testes de hipóteses podem ser paramétricos, quando utilizam os parâmetros da distribuição (ou uma estimativa deles) para o cálculo de sua estatística ou, não paramétricos, quando utilizam postos atribuídos aos dados ordenados. De acordo com Fávero e Fávero (2015), para a aplicação dos testes paramétricos, as observações devem ser oriundas de populações com uma determinada distribuição, que geralmente é a distribuição normal.

Existem alguns testes para verificar o pressuposto de normalidade das variáveis. Entretanto, segundo Leotti, Coster e Riboldi (2012) um dos que apresentam maiores porcentagens de acerto é o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965), cuja estatística W é obtida pela razão entre o quadrado da combinação linear apropriada dos valores ordenados da amostra pela variância simétrica estimada.

2.3.1 Teste para a média com variância desconhecida

O Teste t é utilizado para testar a média de uma população normalmente distribuída, com variância desconhecida conhecida ($H_0: \mu = \bar{X}$). A estatística desse teste é dada por:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

em que n é tamanho amostral, μ a média populacional, \bar{X} a média e s o desvio-padrão da amostra, respectivamente.

2.3.2 Teste t para a comparação de médias

Para comparação de duas médias oriundas de amostras independentes ($H_0: \mu_1 = \mu_2$), frequentemente utiliza-se o t para a comparação de médias. Trata-se de um teste paramétrico cuja estatística é dada por:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

em que \bar{X}_1 e \bar{X}_2 representam as médias, s_1^2 e s_2^2 as variâncias e n_1 e n_2 os tamanhos das amostras analisadas.

2.3.3 Teste não-paramétrico de Mann-Whitney

Para populações não-normais, em comparações de duas amostras independentes pode-se utilizar o teste de Mann-Whitney (também conhecido como teste de Wilcoxon). De acordo com Bussab e Morettin (2010, p. 372), este teste é baseado nos postos dos valores obtidos combinando-se as amostras, sendo a estatística do teste o valor da soma dos postos associados aos valores amostrados de determinada população. Quanto maior o valor dessa soma, maior a tendência de que os valores de uma população sejam maior do que de outra.

2.3 Análises estatísticas utilizando o valor-p.

Os softwares que realizam análises estatísticas, como o Minitab, o R e o SPSS, retornam um valor para a probabilidade de se obter uma estatística de teste maior ou igual a observada pela amostra, sob a hipótese de nulidade. Esta probabilidade é conhecida como valor-p.

Na prática, valores-p pequenos referem-se à improbabilidade de se obter um valor da estatística de teste como o observado, acarretando na rejeição da hipótese nula. Adotando-se α como nível de significância, temos a seguinte regra prática:

$$\begin{aligned} \text{valor} - p \leq \alpha &\Rightarrow \text{Rejeita} - \text{se } H_0 \\ \text{valor} - p > \alpha &\Rightarrow \text{Não Rejeita} - \text{se } H_0 \end{aligned}$$

A facilidade de interpretação do valor-p auxilia na conclusão dos pesquisadores se suas hipóteses estão ou não corretas.

3 MATERIAL E MÉTODOS

Com a finalidade de investigar e analisar as características das questões que compõem os provões, primeiro foi realizado uma revisão de literatura sólida, baseada em livros e artigos científicos que tratam do assunto. Em seguida, partiu-se para as análises dos provões aplicados no primeiro (P1) e segundo (P2) semestres de 2018. Estes provões contaram com a participação de 122 e 108 discentes, respectivamente. Todos os provões foram considerados nas análises.

Cada provão era composto por três questões de Português, três de Matemática, duas de Física, duas de Sociologia, duas de Informática, duas de Química e dezesseis questões de conhecimentos específicos do curso: Construção Civil, Geotecnia, Estruturas, Hidráulica e Transportes. Estas questões são elaboradas pelo corpo docente do IFMG - *Campus Piumhi* e enviadas à coordenação do curso, que faz uma análise preliminar e as organiza nos moldes do ENADE. Cada questão contém 5 alternativas de resposta, em que apenas uma é a correta.

A coleta dos dados analisados ocorreu de maneira sigilosa, sem a presença dos nomes ou códigos que remetem aos respondentes, respeitando os padrões éticos pré-estabelecidos. Estes dados foram tabulados e transcritos em uma planilha do *software* Excel®, conforme ilustram as Figuras 2 e 3.

	A	B	C	D	E	F	...	AF
1		Questão 1	Questão 2	Questão 3	Questão 4	Questão 5	...	Questão 30
2	Gabarito	D	D	C	C	E	...	A
3	Discente 1	B	D	D	E	A	...	C
4	Discente 2	C	D	C	E	E	...	E
5	Discente 3	C	C	D	E	C	...	B
6	Discente 4	D	D	B	C	E	...	A
7	Discente 5	D	E	C	A	E	...	B
8	Discente 6	C	D	C	C	E	...	C
9	Discente 7	C	A	E	C	B	...	B
10	Discente 8	D	B	C	C	E	...	A
124	...	C	E	B	C	E	...	A
125	Discente 122	D	B	A	C	E	...	C

Figura 2 - Modelo do banco de dados na prova 1
 Fonte: Elaborado pelos autores (2019)

	A	B	C	D	E	F	...	AF
1		Questão 1	Questão 2	Questão 3	Questão 4	Questão 5	...	Questão 30
2	Gabarito	C	D	A	E	C	...	A
3	Discente 1	B	C	A	E	C	...	A
4	Discente 2	C	D	A	E	C	...	D
5	Discente 3	C	D	A	E	C	...	C
6	Discente 4	D	D	A	D	E	...	E
7	Discente 5	C	D	A	E	C	...	E
8	Discente 6	C	E	A	E	E	...	D
9	Discente 7	C	D	A	C	B	...	E
10	Discente 8	A	D	A	D	C	...	A
110	...	C	D	A	E	E	...	D
111	Discente 108	C	C	A	E	C	...	A

Figura 3 - Modelo do banco de dados na prova 2
 Fonte: Elaborado pelos autores (2019)

Em seguida, os dados foram convertidos ao formato de texto (.txt) e então exportados para o software R (R Core Team, 2019), utilizado por ser um software livre e de código aberto de fácil manuseio. A partir daí, obtiveram-se os resultados que foram determinantes para a análise em questão.

Os pacotes do software utilizados nas análises foram o *ltm* (RIZOPOULOS, 2006) e o *psych* (REVELLE, 2014). Estes pacotes contêm funções para o cálculo de parâmetros psicométricos utilizados: coeficiente alpha de Cronbach, índice de dificuldade e índice de discriminação. Também foram obtidas as principais estatísticas descritivas para as notas finais do provão (média, desvio-padrão, quartis, mínimo e máximo) e foram gerados gráficos para uma melhor visualização do comportamento das distribuições destes escores.

Através dos resultados obtidos pela análise via TCT, partiu-se para a realização de testes de hipóteses com o intuito de verificar se ambos os provões apresentaram resultados confiáveis e equiparáveis. Para isto, realizou-se o teste t para uma média e os testes t e Mann-Whitney para comparação de médias, verificando, quando necessário, seus pressupostos. O pacote do software R utilizado para a realização dos testes de hipótese foi o *TeachingDemos* (SNOW, 2016).

4 RESULTADOS E DISCUSSÃO

A análise descritiva permitiu observar que a prova P1 apresentou nota mínima de 2, e máxima de 22, em um total de 30 pontos. A média obtida foi de 12,98; o desvio-padrão de 4,51 pontos e o coeficiente de variação de 0,35. A prova P2 apresentou nota mínima de 6, e máxima análoga à prova anterior. Além disso, a nota média foi 14,13; o desvio-padrão 3,45 pontos e o coeficiente de variação foi 0,24 pontos, ou seja, os resultados para esta prova foram mais homogêneos, como ilustram os *boxplots* abaixo (Figura 4).

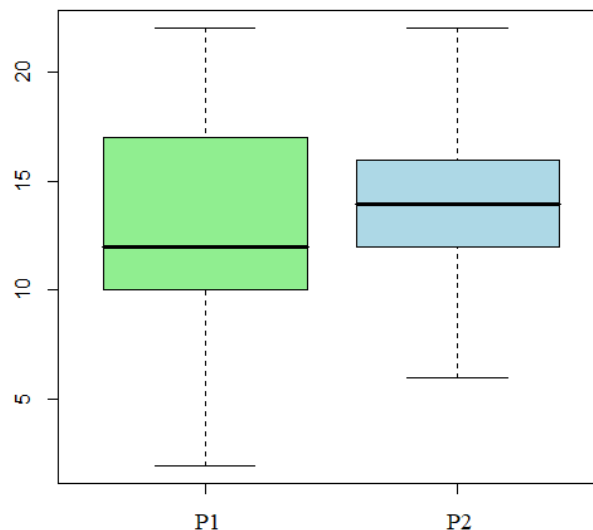


Figura 4 – Boxplots para a Variável Nota
Fonte: Elaborado pelos autores (2019)

As Figuras 5 e 6 representam as distribuições de frequências dos números de acertos da P1 e P2, respectivamente. Observe que os valores nas extremidades, em ambos os gráficos, aparecem com menor frequência e que há indícios, subjetivos, de que estas variáveis possuam um comportamento normal.

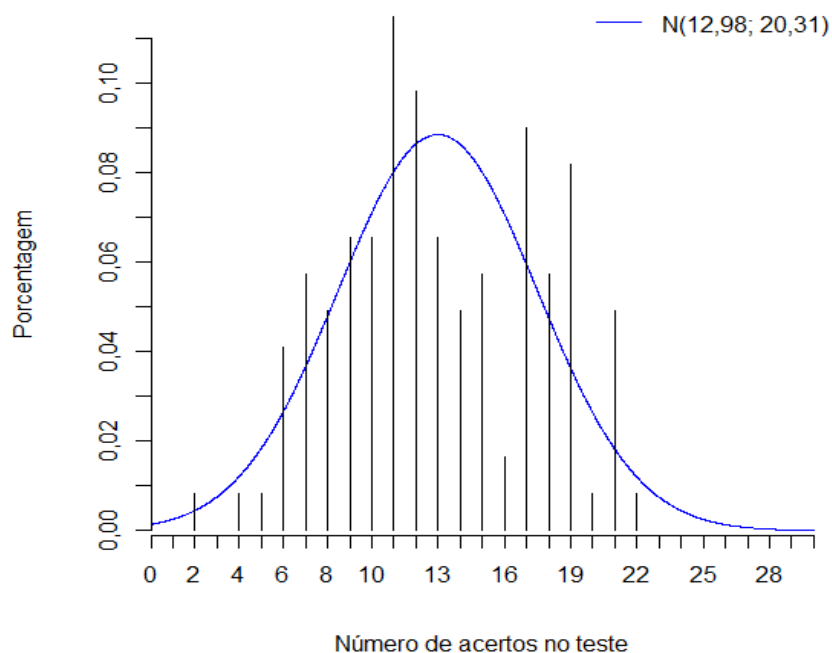


Figura 5 – Distribuição de frequência dos números de acertos dos discentes na P1
Fonte: Elaborado pelos autores (2019)

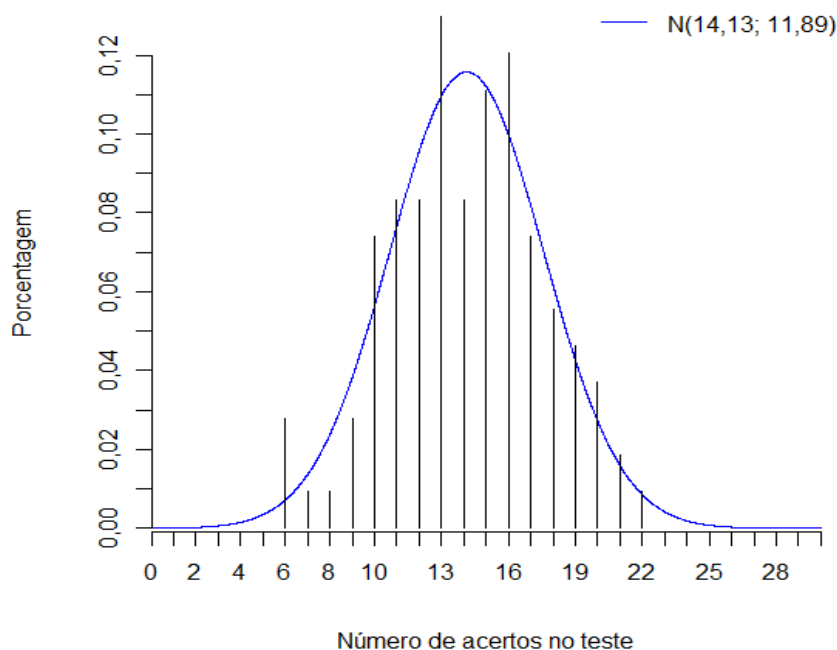


Figura 6 - Distribuição de frequência dos números de acertos dos discentes na P2
Fonte: Elaborado pelos autores (2019)

A partir do desempenho geral dos estudantes, a análise do comportamento dos itens que compõem as provas em questão, segundo os parâmetros definidos pela Teoria Clássica dos Testes (TCT), pôde ser estruturada nas Tabelas 3 e 4.

Tabela 3 – Análise geral da prova 1

Item	Coefficiente alfa de Cronbach (α)	Índice de dificuldade (D_i)	Índice de discriminação (ID)
1	0,7014	0,3852	0,4035
2	0,6945	0,6230	0,6579
3	0,6937	0,4426	0,4649
4	0,6879	0,6803	0,0263
5	0,6973	0,5328	0,2193
6	0,6901	0,3852	0,0702
7	0,7052	0,6885	0,7281
8	0,6880	0,1885	0,2017
9	0,6923	0,2213	0,2368
10	0,6978	0,3770	0,4035
11	0,6951	0,4918	0,1754
12	0,7074	0,3689	0,3596
13	0,7131	0,2131	0,2107
14	0,6839	0,5902	0,6228
15	0,6767	0,6885	0,7368
16	0,6915	0,5328	0,2281
17	0,6940	0,3197	0,3421
18	0,7031	0,1230	0,1140
19	0,6796	0,6066	0,2982
20	0,7025	0,1393	0,1403
21	0,6855	0,5246	0,5614
22	0,6804	0,6148	0,2895
23	0,6663	0,2869	0,3070
24	0,6704	0,5656	0,6053
25	0,6884	0,3443	0,0088
26	0,6975	0,2459	0,2544
27	0,6809	0,5492	0,5789
28	0,6735	0,3852	0,4035
29	0,6876	0,4262	0,0965
30	0,6778	0,4426	0,1228

Fonte: Elaborado pelos autores (2019).

Tabela 4 - Análise geral da prova 2

Item	Coefficiente alfa de Cronbach(α)	Índice de dificuldade (D_i)	Índice de discriminação (ID)
1	0,5206	0,8519	0,1327
2	0,4950	0,7778	0,3899
3	0,5294	0,9630	0,0332
4	0,5186	0,6944	0,2873
5	0,5262	0,8241	0,1953
6	0,4893	0,6019	0,5520
7	0,5075	0,6852	0,3756
8	0,5088	0,7500	0,3680
9	0,5380	0,3426	0,0686
10	0,5241	0,3796	0,3032
11	0,5388	0,0926	0,0437
12	0,5351	0,3056	0,2006
13	0,5248	0,6019	0,2474
14	0,5474	0,4722	0,1086
15	0,5062	0,4444	0,4864
16	0,5066	0,6296	0,3537
17	0,5286	0,1759	0,1938
18	0,5119	0,3611	0,3839
19	0,5154	0,3611	0,3032
20	0,5037	0,1296	0,3077
21	0,5438	0,3056	0,1199
22	0,5279	0,3704	0,2044
23	0,5030	0,5278	0,5083
24	0,5432	0,0833	0,0151
25	0,5399	0,2593	0,2338
26	0,5012	0,6574	0,4600
27	0,5246	0,1944	0,1938
28	0,5567	0,4167	0,1086
29	0,5414	0,1759	0,0799
30	0,5229	0,6944	0,2836

Fonte: Elaborado pelos autores (2019).

Por meio da classificação do índice de dificuldade recomendada por Pasquali (2013), pôde-se notar para a prova 1, que não houve itens classificados como fáceis, porém, os itens 4, 7 e 15, sendo o primeiro pertinente à disciplina de Sociologia, e os demais pertinentes às disciplinas de Matemática, embora considerados como moderados, foram os itens que mais se aproximaram desta classificação. Por outro lado, o item 18, pertencente ao conteúdo da disciplina de Hidrologia, foi classificado como o mais difícil.

Para a prova 2, pode-se observar que o item 3, pertinente à disciplina de Português, foi classificado como o mais fácil. Em contrapartida, o item classificado como o mais difícil, foi o 24, sendo esse referente ao conteúdo da disciplina de Teoria das Estruturas. Uma análise global permite inferir que os itens do núcleo básico do curso geralmente tendem a ser mais fáceis do que os itens do núcleo específico. A Figura 7, cujo eixo das abscissas representa a classificação para o índice de dificuldade e o eixo das ordenadas o quantitativo de itens, ilustra a situação descrita para ambos os testes.

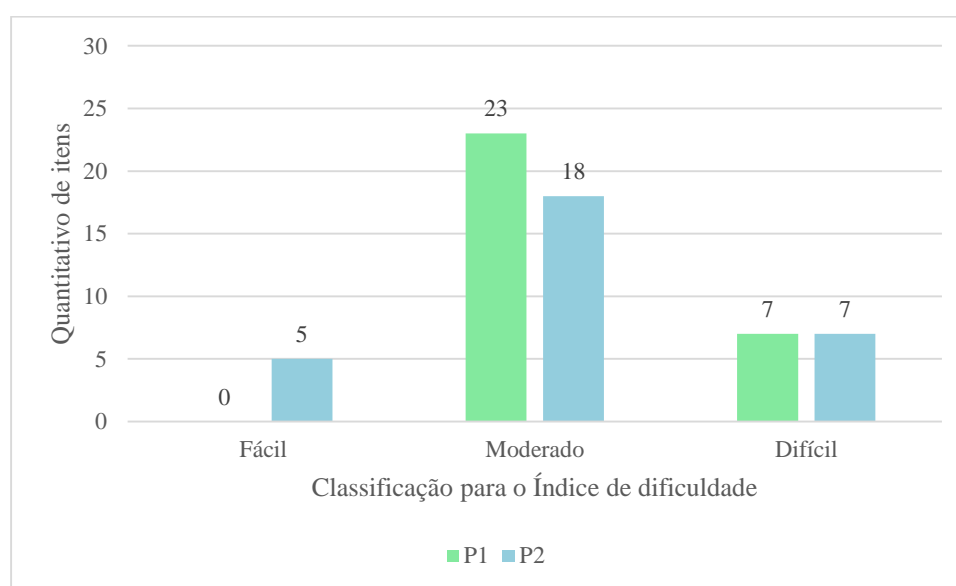


Figura 7 - Quantificação dos itens quanto à classificação do índice de dificuldade
Fonte: Elaborado pelos autores (2018)

Quanto ao índice de discriminação, considerando a classificação sugerida por Arias, Lloreda e Lloreda (2006) pôde-se verificar, para a primeira prova, que os itens 4, 6, 11, 18, 20, 25, 29 e 30 foram ineficientes. Os itens 5, 8, 9, 13, 16, 19, 22 e 26, necessitam ser revisados. Os itens 12, 17, 23, foram aceitáveis, não sendo necessária uma revisão e, por fim, os itens 1, 2, 3, 7, 10, 14, 15, 21, 24, 27 e 28, foram considerados satisfatórios, podendo permanecer na prova.

Para a prova 2, foram classificados como ineficientes os itens 1, 3, 5, 9, 11, 14, 17, 21, 24, 27, 28 e 29. Os itens 4, 12, 13, 22, 25 e 30 necessitam ser revisados. Os itens 2, 7, 8, 10, 16, 18, 19, e 20, são aceitáveis, não sendo necessária uma revisão e, por fim, os itens 6, 15, 23 e 26, são considerados satisfatórios e devem permanecer na prova. A Figura 8, cujo eixo das abscissas representa a classificação, conforme o índice de discriminação e o eixo das ordenadas o quantitativo de itens, ilustra a situação descrita para ambos os testes.

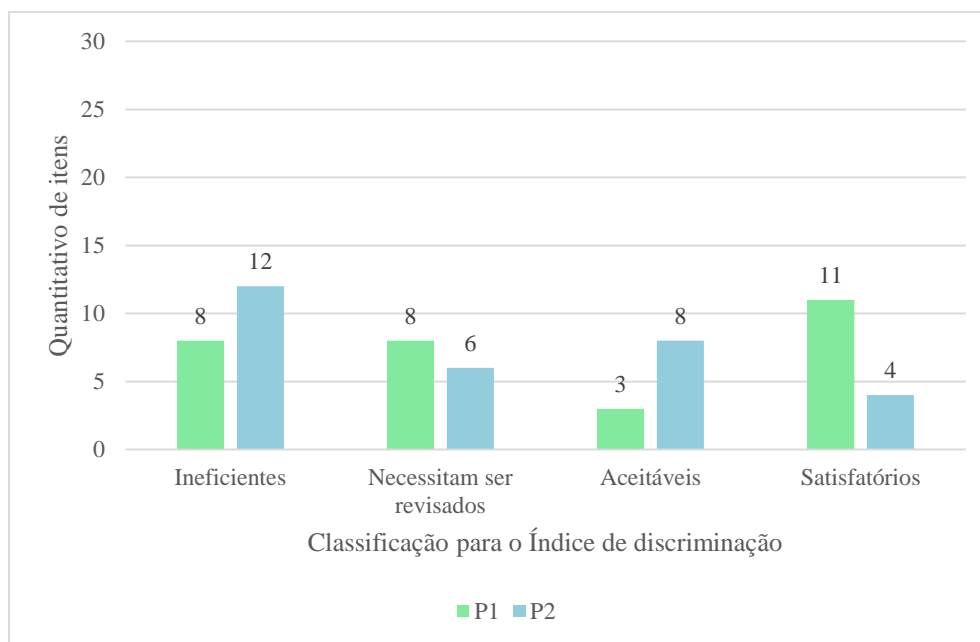


Figura 8 - Quantificação dos itens quanto classificação do índice de discriminação
 Fonte: Elaborado pelos autores (2018)

O coeficiente alpha de Cronbach médio, calculado para a P1 foi de 0,698, atingindo valores satisfatórios para a confiabilidade do teste. Para a P2, este coeficiente médio foi de 0,532. Assumindo o valor mínimo proposto por Malhotra (2013), realizou-se um teste t para verificar se este valor médio pode ser considerado igual a 0,6. Antes, porém, verificou-se que os valores seguiam uma distribuição normal, a partir da aplicação do teste de Shapiro-Wilk, cuja estatística calculada foi 0,982 (valor-p = 0,854).

A estatística t calculada foi de -25,673 (valor-p \approx 0). Assim, não houve indícios estatísticos para considerá-la igual a 0,6. Dessa forma, a P2 apresentou problemas de confiabilidade, sugerindo que enquanto instrumento de mensuração de habilidades, ela não é eficiente.

Também realizou-se o teste de Shapiro-Wilk para verificar se os dados referentes aos índices de dificuldade e discriminação seguem uma distribuição normal. Para o índice de dificuldade, verificou-se a normalidade para ambas as provas. As estatísticas obtidas foram

$W = 0,96$ (valor-p = 0,313) e $W = 0,963$ (valor-p = 0,373), para as provas P1 e P2, respectivamente. Quanto ao índice de discriminação, não houve indícios de normalidade. As estatísticas obtidas foram $W = 0,25$ e $W = 0,266$, respectivamente (valor-p ≈ 0 , em ambas).

Em seguida, buscou-se verificar se esses índices apresentaram diferenças significativas, quando comparadas às suas médias. Para a dificuldade, realizou-se o teste t para a comparação de médias, obtendo-se estatística 0,705 (valor-p = 0,484). Assim, pode-se concluir que o índice de dificuldade apresentou médias estatisticamente iguais para P1 e P2.

Para o índice de discriminação, recorreu-se ao teste não-paramétrico de Mann-Whitney, que compara as medianas, ao invés da média. A estatística do teste foi $W=645$ (valor-p = 0,02), apontando para a diferença entre os valores obtidos.

Em uma análise mais qualitativa, pôde-se observar que os itens considerados como fáceis geralmente estão associados às disciplinas ofertadas pelo núcleo básico do curso. Em contrapartida, os itens classificados como difíceis são de cunho mais específico, o que pode responder pela enorme disparidade observada quando comparadas as aplicações.

5 CONCLUSÃO

Este trabalho propôs-se a verificar o comportamento de dois provões preparatórios para o ENADE, aplicados no ano de 2018 no IFMG – *Campus* Avançado Piumhi. Para isto, baseou-se na análise da confiabilidade dos resultados do teste e nos índices de dificuldade e de discriminação dos itens, parâmetros psicométricos que constituem a Teoria Clássica os Testes.

Estas análises permitiram observar que, embora ocorressem em períodos consecutivos, os provões analisados apresentaram características estruturais distintas, tornando o processo de equiparação de resultados, duramente questionado na literatura, ainda mais complexo.

Embora alguns índices não tenham sido satisfatórios, verificou-se que os objetivos propostos neste artigo foram alcançados, visto que análises como a realizada são de grande importância para os elaboradores das provas, pois tornam possível identificar falhas e acertos, tornando o processo de mensuração de conhecimentos mais significativo e contribuindo para o processo ensino-aprendizagem das disciplinas a eles relacionadas.

Nesse sentido, propõe-se uma revisão sistemática no processo de elaboração e estruturação dessas provas, de forma que sejam constituídas com conteúdo condizente aos

períodos de cada respondente. Por exemplo, um discente matriculado no 5º período, responderá uma prova composta por assuntos que já tenham sido abordados nas disciplinas até o seu respectivo período.

Espera-se que este trabalho sirva como instrumento de difusão dos conhecimentos aqui abordados e que os resultados obtidos auxiliem a gestão do *campus* na administração dos provões preparatórios para o ENADE, tornando os objetivos aspirados pelo NDE, alcançáveis.

REFERÊNCIAS

- ARIAS, M. R. M.; LLOREDA, M. V. H.; LLOREDA, M. J. H. **Psicometria**. [s.l.]: Alianza, 2006. 488 p.
- BORGATTO, A. F.; ANDRADE, D. F. Análise clássica de testes com diferentes graus de dificuldade. **Estudos em Avaliação Educacional**, São Paulo, v. 23, n. 52, p. 146-156, 2012. Disponível em: <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/1934>. Acesso em: 22 nov. 2018.
- BRASIL. Ministério da Educação. Comissão Nacional de Avaliação da Educação Superior **Resolução nº 01 de 17 de junho de 2010**. Normatiza o Núcleo Docente Estruturante e dá outras providências. Brasília: CONAES, 2010. Disponível em: <https://www.ifmg.edu.br/piumhi/cursos-1/superior/nde/docs/resolucoes/Resoluon01de17dejunhode2010NormatizaoNDECONAES.pdf>. Acesso em: 13 nov. 2018.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva. 557 p., 2010.
- CORTINA, J. M. What is coefficient alpha? an examination of theory and applications. **Journal of Applied Psychology**, [s.l.], v. 78, n. 1, p. 98-104, 1993. Disponível em: <https://psycnet.apa.org/record/1993-19965-001>. Acesso em: 20 nov. 2018.
- FÁVERO, L.; FÁVERO, P. **Análise de dados: modelos de regressão com Excel, Stata e SPSS**. Rio de Janeiro: Elsevier, 2015.
- HAIR JÚNIOR, J. F.; BLACK, W. C.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.
- INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS. *Campus* Avançado Piumhi. **Ata de reunião do Núcleo Docente Estruturante (NDE)**. Piumhi: [IFMG], 2017. Disponível em: <https://www.ifmg.edu.br/piumhi/cursos-1/superior/nde/docs/atas/2017/AtareuniaoNDE290317.pdf>. Acesso em: 13 nov. 2018.
- KELLEY, T. L. The selection of upper and lower groups for the validation of test items. **Journal of educational psychology**, Warwick & York, v. 30, n. 1, p. 17-24, 1939. Disponível em: <https://psycnet.apa.org/record/1939-03313-001>. Acesso em: 20 nov. 2018.

LEOTTI, V. B.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista HCPA**. Porto Alegre, v. 32, n. 2, p. 227-234, 2012. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/29874>. Acesso em: 29 maio 2019.

MALHOTRA, N. **Review of marketing research**. Emerald Group Publishing Limited. Bingley. 2013. Disponível em: [https://www.emerald.com/insight/content/doi/10.1108/S1548-6435\(2007\)0000003004/full/html?skipTracking=true](https://www.emerald.com/insight/content/doi/10.1108/S1548-6435(2007)0000003004/full/html?skipTracking=true). Acesso em: 28 maio 2019.

PASQUALI, L. **Psicometria**: teoria dos testes na psicologia e na educação. Petrópolis: Vozes, 2013. 397 p.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2019. Disponível em: www.R-project.org/. Acesso em: 15 nov. 2018.

REVELLE, W. **Psych**: procedures for personality and psychological research. North western University, Evanston. Illinois, USA, 2014. Disponível em: <https://cran.r-project.org/web/packages/psych/index.html>. Acesso em: 28 maio 2019.

RIZOPOULOS, D. ltm: an R package for latent variable modeling and item response theory analyses. **Journal of statistical software**, [s.l.], v. 17, n. 5, p. 1-25, 2006. Disponível em: <https://core.ac.uk/download/pdf/6305163.pdf>. Acesso em: 28 nov. 2019.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3/4, p. 591-611, 1965.

SNOW, G. TeachingDemos: Demonstrations for teaching and learning. **R package version**, v. 2, 2016.

SOARES, D. J. M. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica**. 2018. 121 f. Dissertação (Mestrado em Estatística) -Universidade Federal de Viçosa, Viçosa, 2018. Disponível em: <https://www.locus.ufv.br/bitstream/handle/123456789/18404/texto%20completo.pdf?sequence=1&isAllowed=y>. Acesso em: 15 nov. 2018.

VILARINHO, A. P. L. **Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP**. Dissertação (Mestrado Profissional em Matemática). Universidade de Brasília, Brasília, 2015. Disponível em: <https://core.ac.uk/download/pdf/33552126.pdf>. Acesso em: 15 nov. 2018.

DADOS DOS AUTORES

Milene Cunha Costa

E-mail: milenecunha.eng@gmail.com

Currículo lattes: <http://lattes.cnpq.br/1442080081457206>

Graduanda em Engenharia Civil pelo Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, *Campus Piumhi*. Técnica em Eletrotécnica pela Escola Técnica Estadual de Furnas (ETEF).

Samuel Henrique Oliveira Lima

E-mail: samuelhol.eng@gmail.com

Currículo lattes: <http://lattes.cnpq.br/0003218676993442>

Graduando em Engenharia Civil pelo Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, *Campus Piumhi*.

Denilson Junio Marques Soares

E-mail: denilson.marques@ifmg.edu.br

Currículo lattes: <http://lattes.cnpq.br/1813099481556666>

Doutorando do Programa de Pós-Graduação em Educação (PPGE) da Universidade Federal do Espírito Santo (UFES). Mestre em Estatística Aplicada e Biometria e Licenciado em Matemática pela Universidade Federal de Viçosa (UFV). Professor do Instituto Federal de Minas Gerais, *Campus Piumhi*.