

<https://doi.org/10.18222/ea.v34.7951>

ANÁLISE DOS ITENS DE MÚLTIPLA ESCOLHA DAS PROVAS DO ENADE 2016

 KAIZÔ IWAKAMI BELTRÃO^I

 MONICA CERBELLA FREIRE MANDARINO^{II}

^I Fundação Getulio Vargas (FGV), Rio de Janeiro-RJ, Brasil; kaizo.beltrao@fgv.br

^{II} Fundação Cesgranrio, Rio de Janeiro-RJ, Brasil; monica.cerbella@gmail.com

RESUMO

Investiga-se se o formato de um item, o nível cognitivo exigido e a posição do item na prova do Enade 2016 afetam os resultados e a probabilidade de anulação. Os itens foram classificados quanto à forma e ao nível cognitivo, com base na literatura e na taxionomia de Bloom. Utilizaram-se medidas da pós-aplicação das provas: índices de facilidade e de discriminação. Para o formato, verificou-se que os itens de asserção-razão foram mais difíceis, os de resposta única tiveram uma capacidade de discriminação em média um pouco maior e um terço dos itens de asserção-razão foi anulado. Quanto ao nível cognitivo, os dois índices têm médias mais baixas nos itens de análise e a proporção de itens anulados desse nível é maior. A hipótese de a posição do item na prova afetar o resultado foi rejeitada.

PALAVRAS-CHAVE ENSINO SUPERIOR • ENADE • TÉCNICAS DE MENSURAÇÃO • ANÁLISE DE ITEM.

COMO CITAR:

Beltrão, K. I., & Mandarino, M. C. F. (2023). Análise dos itens de múltipla escolha das provas do Enade 2016. *Estudos em Avaliação Educacional*, 34, Artigo e07951. <https://doi.org/10.18222/ea.v34.7951>

ANÁLISIS DE LOS ÍTEMS DE MÚLTIPLE SELECCIÓN DE LAS PRUEBAS ENADE 2016

RESUMEN

Se investiga si la forma de un ítem, el nivel cognitivo requerido y la posición del ítem en la prueba Enade 2016 afectan los resultados, incluso la probabilidad de cancelación. Los ítems se clasificaron según la forma y el nivel cognitivo, en base a la literatura y la Taxonomía de Bloom. Se utilizaron medidas de post-aplicación de las pruebas: índices de facilidad y de discriminación. En lo que se refiere al formato, se verificó que los ítems de aserción-razón fueron más difíciles, los de única respuesta tuvieron un promedio de capacidad de discriminación un poco mayor y se anuló una tercera parte de los ítems de aserción-razón. En cuanto al nivel cognitivo, se encontró que los dos índices tienen menores promedios en los ítems de análisis y que la proporción de cancelación de ítems de este nivel es mayor. Se rechazó la hipótesis de que la posición del ítem en la prueba influye en el resultado.

PALABRAS CLAVE EDUCACIÓN SUPERIOR • ENADE • TÉCNICAS DE MEDICIÓN • ANÁLISIS DEL ÍTEM.

ANALYSIS OF THE MULTIPLE CHOICE ITEMS OF THE ENADE 2016 TESTS

ABSTRACT

The objective was to investigate whether the format of an item, the required cognitive level and the item's position in the Enade 2016 test affect the results, including the probability of invalidation. The items were classified according to format and cognitive level, based on the literature and Bloom's Taxonomy. Post-test administration measures were used: difficulty and discrimination indices. Regarding format, it was found that assertion-reason items were more difficult, single answer items had a slightly greater average for discrimination capacity, and one third of the assertion-reason items were invalidated. As for cognitive level, the two indices have lower averages in items of analysis, and the proportion of invalidated items of this level is higher. The hypothesis that the position of an item in the test affects the result was rejected.

KEYWORDS HIGHER EDUCATION • ENADE • MEASUREMENT TECHNIQUES • ITEM ANALYSIS.

Recebido em: 11 NOVEMBRO 2020

Aprovado para publicação em: 30 NOVEMBRO 2022



Este é um artigo de acesso aberto distribuído nos termos da licença Creative Commons do tipo BY-NC.

INTRODUÇÃO

Neste artigo,¹ é apresentada uma meta-avaliação dos itens da parte objetiva das provas do Exame Nacional de Desempenho dos Estudantes (Enade) aplicadas aos concluintes de 2016.

As áreas de bacharelado que participaram daquela edição do Enade foram: agronomia; biomedicina; educação física; enfermagem; farmácia; fisioterapia; fonoaudiologia; medicina; medicina veterinária; nutrição; odontologia; serviço social; e zootecnia. As áreas participantes que formam tecnólogos foram: tecnologia em agronegócio; tecnologia em estética e cosmética; tecnologia em gestão ambiental; tecnologia em gestão hospitalar; e tecnologia em radiologia.

Ao longo da experiência referente à aplicação de provas objetivas, algumas questões sempre parecem permear a discussão dos resultados. Uma delas, por exemplo, é se o formato do item influencia na dificuldade para a sua resolução. Também parece quase natural que um item que exija um nível cognitivo mais elaborado seja mais difícil do que outro com nível cognitivo menos exigente. Outra dúvida é se, ao longo da prova, talvez por cansaço, os erros aumentam. O estudo toma essas hipóteses como questões de pesquisa e recorre a índices obtidos a partir dos resultados da aplicação das provas objetivas do Enade 2016.

Assim, o artigo busca analisar se as características intrínsecas ao item (tipo de item e nível cognitivo avaliado pelo item) ou a sua localização na prova influenciam os resultados aferidos empiricamente, pós-aplicação da prova: índices de facilidade² e de discriminação. Por meio de uma regressão linear, também procura-se verificar se a localização de um item na prova tem influência no desempenho daquele item. Por fim, apresentam-se conclusões e recomendações.

METODOLOGIA

As tabelas e as descrições dos índices utilizados são oriundas dos relatórios de área do Enade (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [Inep], 2017a). O *índice de facilidade* é o percentual de acerto de cada item objetivo, o que leva a uma classificação dos itens apresentada na Tabela 1. Itens acertados por 86% dos estudantes, ou mais, são consideradas *muito fáceis*. No extremo oposto, aqueles com percentual de acerto igual ou inferior a 15% são considerados *muito difíceis*.

- 1 Os autores agradecem imensamente aos pareceristas pelos comentários construtivos que fizeram, eliminando erros e deixando o texto mais claro.
- 2 Na Teoria Clássica de Testes, o índice de acertos é usualmente denominado índice de dificuldade. No entanto, neste artigo, bem como em todos os relatórios do Enade (por exemplo, Inep, 2017b), usa-se o termo índice de facilidade, para uma interpretação mais fácil e intuitiva.

TABELA 1
Classificação de questões, segundo o índice de facilidade

ÍNDICE DE FACILIDADE	CLASSIFICAÇÃO
$\geq 0,86$	Muito fácil
0,61 a 0,85	Fácil
0,41 a 0,60	Médio
0,16 a 0,40	Difícil
$\leq 0,15$	Muito difícil

Fonte: Ministério da Educação (MEC)/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)/Diretoria de Avaliação da Educação Superior (Daes) – Enade 2016.

O *índice de discriminação* permite avaliar se um item objetivo tem um nível mínimo de poder para distinguir estudantes com diferentes patamares de desempenhos (no caso do Enade, avaliados pela mesma prova). Para ser considerada apta a avaliar os alunos dos cursos, uma questão deve ser mais acertada por alunos que tiveram bom desempenho do que pelos que tiveram desempenho ruim. Um índice que mede essa capacidade das questões e que foi escolhido para ser utilizado no Enade é denominado índice de discriminação, que é calculado pela correlação ponto-biserial, usualmente representada por r_{pb} .

Para cada combinação de área participante e componente da prova (componente de formação geral e componente de conhecimentos específicos), a correlação ponto-biserial de um item objetivo é calculada pela fórmula a seguir:

$$r_{pb} = \frac{\bar{C}_A - \bar{C}_T}{S_T} \sqrt{\frac{p}{q}} \quad (1)$$

onde:

\bar{C}_A é a nota média da parte objetiva de um componente da prova (formação geral ou conhecimentos específicos) pelos alunos que acertaram o item;

\bar{C}_T é a nota média da prova por todos os alunos da área;

S_T é o desvio padrão das notas para todos os alunos da área;

p é a proporção de estudantes que acertaram o item (número de alunos que acertaram a questão, dividido pelo número total de alunos que compareceram à prova);

$q = 1 - p$ é a proporção de estudantes que erraram o item.

A Tabela 2 traz o índice de discriminação (para itens de múltipla escolha com cinco opções) classificado por faixas de valores. Um item com discriminação fraca deve ser eliminado do cômputo das notas. Elimina-se o item que apresentou menor índice de discriminação e recalculam-se as notas da prova e os índices de discriminação dos itens restantes. Mais de um item pode ser suprimido da prova,

repetindo-se essa estratégia até que permaneçam somente itens com índice de discriminação relevantes.

Na Tabela 2, apresenta-se a classificação, segundo o índice de discriminação utilizado (ponto-bisserial).

TABELA 2
Classificação de questões, segundo o índice de discriminação

ÍNDICE DE DISCRIMINAÇÃO	CLASSIFICAÇÃO
$\geq 0,40$	Muito bom
0,30 a 0,39	Bom
0,20 a 0,29	Médio
$\leq 0,19$	Fraco

Fonte: MEC/Inep/Daes - Enade 2016.

CLASSIFICAÇÃO QUANTO À ESTRUTURA DOS ITENS

A tipologia utilizada

Considerando os itens da parte objetiva das provas dos cursos avaliados no Enade 2016, foi realizada uma classificação segundo os três grandes tipos de estrutura de itens previstos no documento *Guia de elaboração e revisão de itens* do Inep (2017a).

Os três grandes tipos de itens objetivos previstos no guia que orienta o trabalho dos elaboradores do Banco Nacional de Itens (BNI) são:

- tipo 1 – itens de interpretação com resposta única;
- tipo 2 – itens de resposta múltipla;
- tipo 3 – itens de asserção-razão.

O tipo 1 é definido no documento com base na estratégia de formulação:

O item de múltipla escolha do tipo interpretação com resposta única é formulado a partir de uma situação-estímulo que compõe o texto-base. A partir da situação-estímulo, que traz um problema a ser resolvido, o estudante organiza as ideias, os dados ou as informações necessárias para resolvê-lo. É nesse momento que ele mobiliza diversos recursos, articulando habilidades, saberes, conhecimentos, entre outros, que vão oportunizar a visibilidade de determinada competência ou característica do perfil profissional esperado. O enunciado desse tipo de item pode ser apresentado tanto como frase incompleta, a ser finalizada pelas alternativas de resposta, quanto em forma de pergunta. (Inep, 2017a, p. 21).

A definição do tipo 2 também é dada no documento pela estrutura de formulação:

Os itens de resposta múltipla apresentam uma situação-estímulo e de três a cinco afirmações relacionadas à situação e ao enunciado. As afirmações são julgadas, independentemente umas das outras, como corretas ou incorretas. O estudante deve analisar as afirmações, identificar a(s) afirmação(ões) correta(s) para, em seguida, escolher uma alternativa em uma 'chave de respostas'. (Inep, 2017a, p. 22).

Para o tipo 3 – asserção-razão –, a apresentação também se apoia na estrutura de elaboração e há preocupação com a elaboração das alternativas, pela dificuldade própria desse tipo de item.

O item de múltipla escolha do tipo asserção-razão constitui-se de duas proposições que podem ou não ser corretas e, além disso, podem ou não estabelecer uma relação de causa e consequência entre si. Desse modo, são apresentadas duas proposições, interligadas pela palavra PORQUE, sendo que o estudante deve avaliar se a segunda proposição constitui razão ou justificativa da primeira.

Compõe o item, ainda, uma chave de respostas padrão, em que são descritas as alternativas de resposta propriamente ditas. Cada uma das alternativas contém uma afirmação sobre a veracidade ou a falsidade de cada uma das proposições e, também, sobre a existência ou não de relação de causalidade entre elas. Esse tipo de questão é indicado para avaliação de habilidades complexas, pois além de avaliar a veracidade de cada proposição, o estudante necessita ser capaz de estabelecer relação de causa-consequência entre elas. A seguir, apresentam-se exemplos de itens do tipo asserção-razão. (Inep, 2017a, p. 24).

Entre as recomendações gerais (item 5.1 do *Guia*) para o estudo aqui apresentado, merecem destaque:

iii. Os itens deverão ser embasados em situações-estímulo, que podem ser constituídas de situações-problema, estudos de caso, simulacros ou mesmo textos de contextualização do conteúdo a ser avaliado. Quanto mais próximas das situações trabalhadas na graduação e mesmo na área de formação, melhor será a qualidade do item e a possibilidade de avaliar o desempenho do aluno;

v. Deve-se dar preferência a textos de fonte primária, de autoria explícita, que sejam curtos e de fácil compreensão, com linguagem apropriada ao nível de dificuldade esperado para os estudantes, que sejam adequados aos objetivos da questão e possibilitem a criação das alternativas de resposta: uma correta e as demais incorretas, mas plausíveis. (Inep, 2017a, p. 32).

Classificação dos itens, segundo a tipologia da estrutura

No Enade 2016, foram avaliadas 13 áreas que formam bacharéis e cinco que formam tecnólogos, totalizando 18 provas. Foram classificadas, segundo sua estrutura, todas as 27 questões de cada uma das 18 das provas de conhecimentos específicos. Para isso, usou-se a tipologia do *Guia* do Inep, descrita na seção anterior.

Para evitar viés, as classificações dos 486 itens foram realizadas por três técnicos experientes no campo da elaboração e revisão de itens objetivos. Nos casos em que surgiram discrepâncias, houve uma reunião com a participação de um quarto profissional, quando então se decidiu a tipologia definitiva.

Os resultados estão apresentados na Tabela 3. Considerando-se todos os itens objetivos da parte de conhecimentos específicos das provas (um total 486 itens), constatou-se que 55,56% eram do tipo 2 – item de resposta múltipla –, 32,72% do tipo 1 – resposta única – e 11,7% do tipo 3 – asserção-razão.

TABELA 3

Itens utilizados por tipologia do item, segundo áreas participantes do Enade 2016

ÁREAS	TIPO 1	TIPO 2	TIPO 3
Agronomia	8	19	-
Biomedicina	7	18	2
Educação Física	7	13	7
Enfermagem	13	10	4
Farmácia	10	13	4
Fisioterapia	6	20	1
Fonoaudiologia	7	16	4
Medicina	21	5	1
Medicina Veterinária	7	20	-
Nutrição	9	16	2
Odontologia	8	14	5
Serviço Social	12	11	4
Zootecnia	10	15	2
Tecnologia em Agronegócio	6	16	5
Tecnologia em Estética e Cosmética	7	12	8
Tecnologia em Gestão Ambiental	10	13	4
Tecnologia em Gestão Hospitalar	2	23	2
Tecnologia em Radiologia	9	16	2
TOTAL	159 (32,72%)	270 (55,56%)	57 (11,73%)

Fonte: Elaboração dos autores, 2020.

As oito questões objetivas de formação geral foram também classificadas e eram dos seguintes tipos:

- três do tipo 1 (questões 2, 4 e 7);
- quatro do tipo 2 (questões 1, 3, 6 e 8);
- um do tipo 3 (questão 5).

Para responder à primeira questão de pesquisa – se a estrutura de elaboração de um item influencia seu nível de dificuldade e seu poder de discriminação –, três tipos de análise são apresentados a seguir.

Primeiramente, buscou-se verificar a situação que seria mais grave, ou seja, se há um tipo de item que leva a um maior número de anulações, tanto anulações realizadas pela própria Comissão Assessora de Área, quanto aquelas que se originam do cálculo do índice de discriminação (critério ponto bisserial).

Posteriormente, apresentam-se as estatísticas básicas da distribuição dos índices de facilidade e de discriminação, por tipo de item (média, desvio padrão, moda e mediana), considerando-se, em conjunto, os itens de todas as provas objetivas de conhecimentos específicos de 2016. Também são apresentados os gráficos da distribuição percentual dos índices, por tipo de item.

Finalmente, como a quantidade de itens de cada tipo é bastante diferente, recorreu-se ao gráfico de distribuição acumulada do índice de facilidade, por tipo de item.

Tipologia das questões anuladas

A Tabela 4 traz a quantidade de itens anulados por área, tipologia e critério de eliminação (CAA: anulado pelas Comissões Assessoras das Áreas ou IDF: índice de discriminação fraco).

TABELA 4

Itens anulados por tipologia do item e motivo de anulação, segundo áreas participantes do Enade 2016

ÁREAS	TIPO 1		TIPO 2		TIPO 3	
	CAA	IDF	CAA	IDF	CAA	IDF
Agronomia	1	0	1	1	0	0
Biomedicina	1	1	0	4	0	1
Educação Física	0	0	2	0	0	1
Enfermagem	0	2	0	3	0	0
Farmácia	0	3	0	3	0	2
Fisioterapia	0	1	2	3	0	0
Fonoaudiologia	0	1	2	3	0	2
Medicina	0	4	0	0	0	1
Medicina Veterinária	0	1	0	3	0	0
Nutrição	0	2	0	2	0	0
Odontologia	0	0	0	3	0	1

(continua)

(continuação)

ÁREAS	TIPO 1		TIPO 2		TIPO 3	
	CAA	IDF	CAA	IDF	CAA	IDF
Serviço Social	2	2	2	0	0	2
Zootecnia	0	2	1	3	0	0
Tecnologia em Agronegócio	0	2	0	5	0	3
Tecnologia em Estética e Cosmética	0	2	1	0	0	3
Tecnologia em Gestão Ambiental	0	1	0	2	0	1
Tecnologia em Gestão Hospitalar	0	0	0	6	0	1
Tecnologia em Radiologia	0	3	0	5	0	0
TOTAL	4	27	11	46	0	18
% relativo ao total de itens (486)	0,82	5,56	2,26	9,47	0,0	3,70
% relativo ao total de anuladas (106)	3,77	25,47	10,38	43,40	0,0	16,98
% relativo ao total de itens de cada tipo	2,52	16,98	4,07	17,04	0,0	31,58

Fonte: Elaboração dos autores, 2020.

Dos 486 itens de conhecimentos específicos, 15 (3,09% do total) foram anulados pelas Comissões Assessoras de Área e 91 (18,72% do total) por terem índice de discriminação fraco.

Assim, entre as questões objetivas de conhecimentos específicos, 106 (21,81% do total) não foram computadas nas notas finais dos estudantes, resultando nos seguintes percentuais de anulações por tipo de item:

- 29,25% do tipo 1 (resposta única);
- 53,77% do tipo 2 (respostas múltiplas);
- 16,98% do tipo 3 (asserção-razão).

Já os percentuais de itens eliminados relativamente à quantidade de itens de cada tipo foram:

- 19,50 % no tipo 1 (31 em 159);
- 21,11 % no tipo 2 (57 em 270);
- 31,58% no tipo 3 (18 em 57).

Estatísticas da distribuição dos índices de facilidade e de discriminação (bisserial), por tipo de item

As tabelas 5 e 6 apresentam algumas estatísticas dos índices de facilidade e de discriminação, por tipo do item (somente os 471 itens não anulados pelas CAA). Como descrito anteriormente, o índice de facilidade varia de 0 a 1, já que é o percentual de acertos do item, enquanto o índice de discriminação varia de -1 a 1, uma vez que se trata de um índice de correlação.

Na Figura 1, encontram-se as médias, com os respectivos intervalos de confiança de 95%, do índice de facilidade, por tipo de item. A sobreposição desses

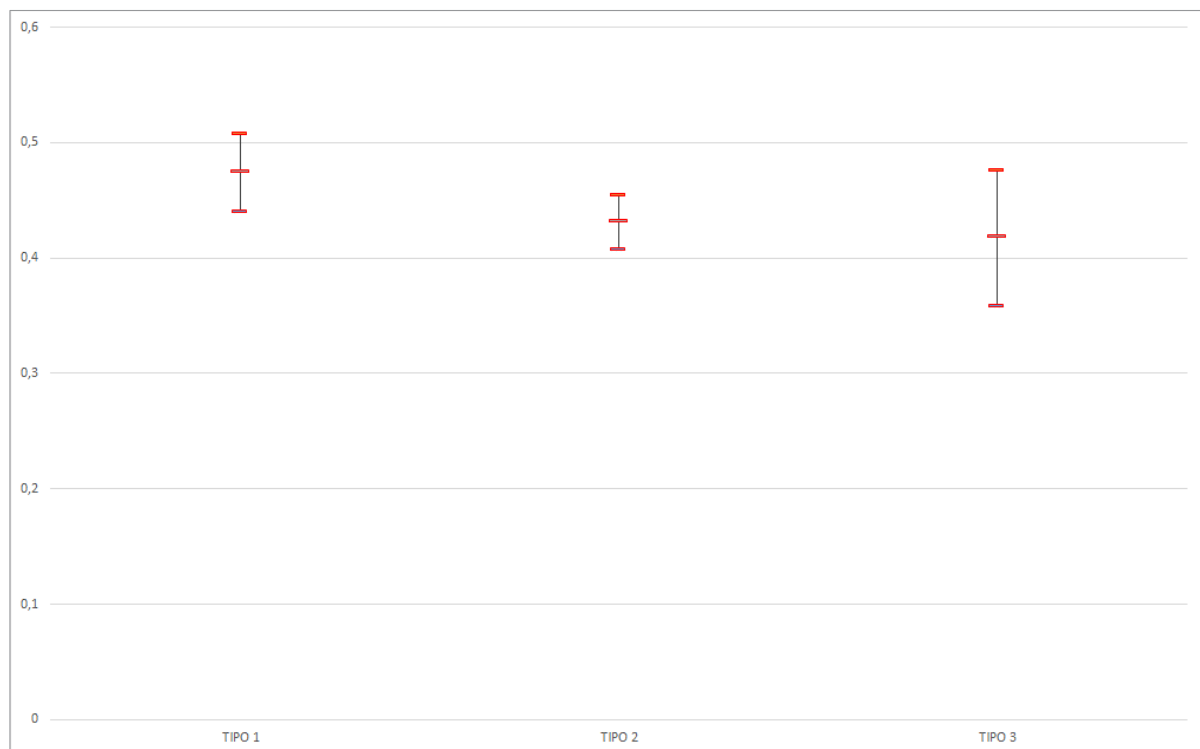
intervalos indica que não há diferença estatisticamente significativa³ entre as médias dos diferentes tipos, ainda que os valores médios do indicador decaiam com a complexidade do tipo de item.

TABELA 5
Índice de facilidade, segundo tipo de item

TIPO	N	MEDIANA	MÉDIA	IC 95%	
				LI	LS
1	155	0,44	0,47	0,44	0,51
2	259	0,41	0,43	0,41	0,46
3	57	0,39	0,42	0,36	0,48

Fonte: Elaboração dos autores, 2020.

FIGURA 1
Média do índice de facilidade e IC de 95%, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

A Figura 2 traz as médias, com os respectivos intervalos de confiança de 95%, do índice de discriminação, por tipo de item. Pode-se observar que, semelhantemente

3 Os intervalos de confiança individuais das médias sugerem sobreposições que foram testadas usando o intervalo de confiança da diferença para validar essa indicação.

ao apurado quanto ao índice de facilidade (Figura 1), os valores médios do indicador decaem com a complexidade do tipo de item, observando-se, porém, uma diferença estatisticamente significativa entre as médias dos tipos 1 e 3.

TABELA 6
Índice de discriminação, segundo tipo de item

TIPO	N	MEDIANA	MÉDIA	IC 95%	
				LI	LS
1	155	0,29	0,28	0,27	0,30
2	259	0,28	0,27	0,26	0,28
3	57	0,26	0,23	0,20	0,26

Fonte: Elaboração dos autores, 2020.

FIGURA 2
Média do índice de discriminação e IC de 95%, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

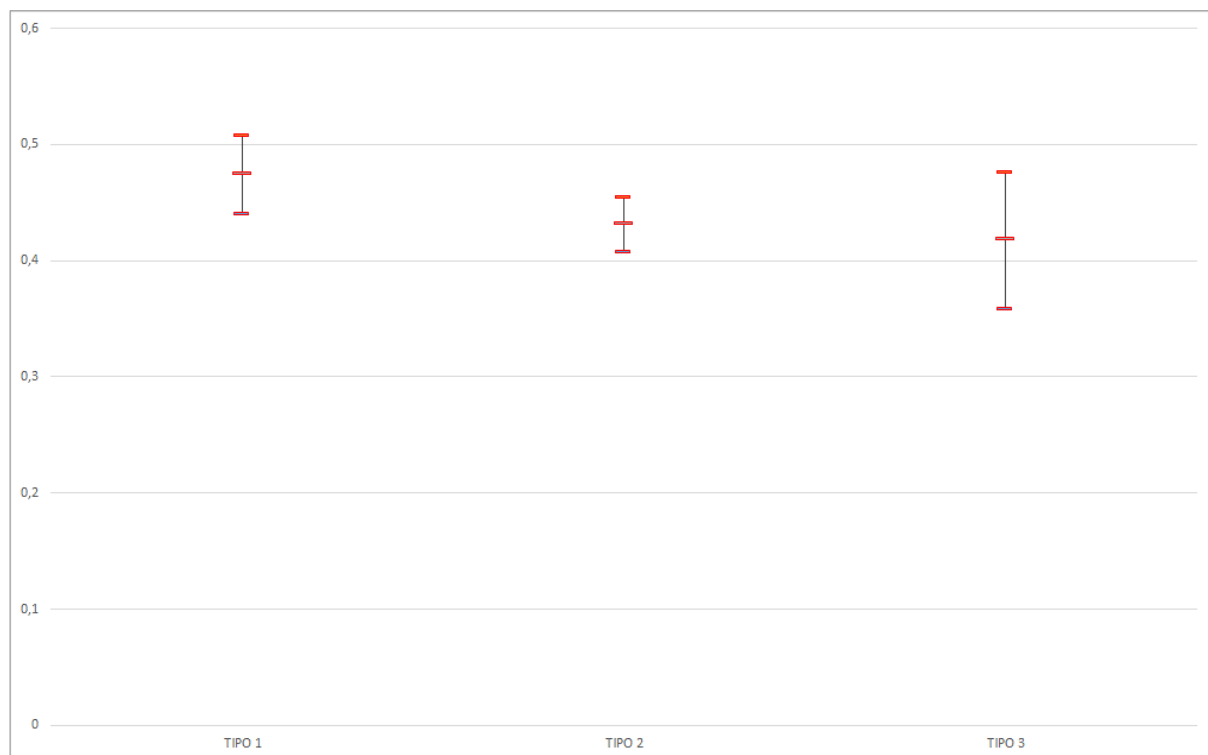
Nas figuras 3 e 4, são apresentadas, respectivamente, as distribuições de frequência relativa dos índices de facilidade e de discriminação, em intervalos de comprimento 0,1. O intervalo modal do índice de facilidade (Figura 3) é (0,3; 0,4] para os três tipos de itens. Em relação ao índice de discriminação (Figura 4), o intervalo

modal é (0,2; 0,3] para os tipos 2 e 3, e a distribuição é bimodal para o tipo 1, com modas nos intervalos (0,2; 0,3] e (0,3; 0,4].

Empregando as classificações apresentadas nas tabelas 1 e 2, *grosso modo*, pode-se concluir que, quanto ao índice de facilidade, os itens dos três tipos foram, em sua maioria, *difíceis* para os estudantes, já que a moda ocorre no intervalo (0,3; 0,4], apesar de as médias estarem um pouco acima de 0,4 (nível de facilidade médio). No que se refere ao índice de discriminação, os três tipos de item têm modas no intervalo (0,2; 0,3], ou seja, concentração em itens de média capacidade de discriminação. Mas, no caso do tipo 1, por ser bimodal, a concentração se aplica também ao intervalo (0,3; 0,4], com 37,4% de itens bons, percentual relativo maior se comparado aos tipos 2 e 3 (31,7% e 24,6%, respectivamente), o que indica que, proporcionalmente, itens do tipo 1 se destacam em relação à capacidade de discriminação, nessa edição do Enade.

FIGURA 3

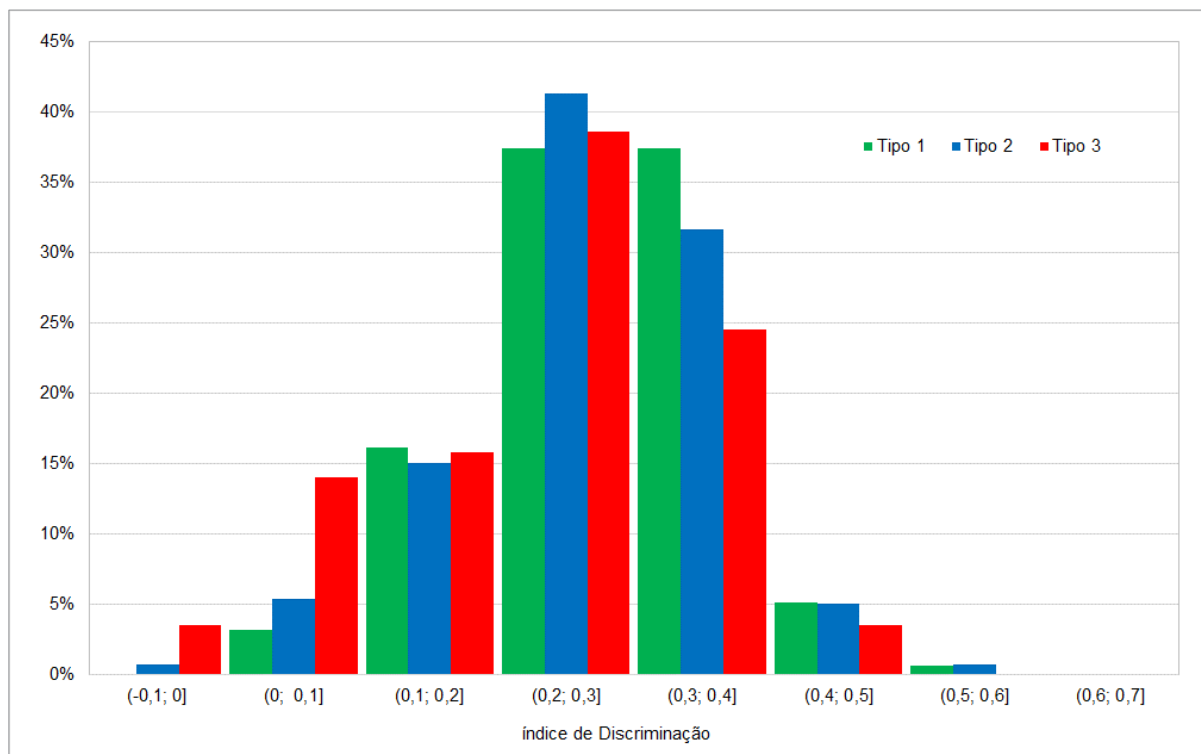
Distribuição do índice de facilidade, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

FIGURA 4

Distribuição do índice de discriminação, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

Distribuição acumulada dos índices de facilidade e de discriminação, por tipo de item

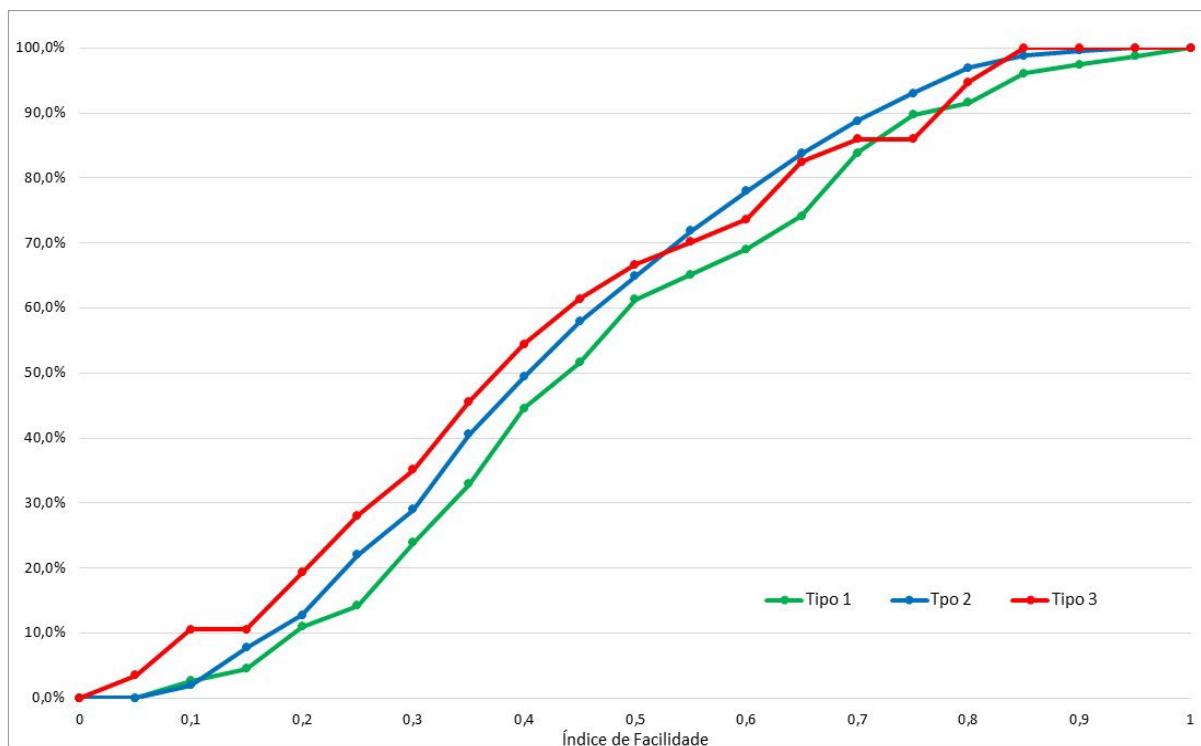
Na Figura 5, é apresentada a distribuição cumulativa do índice de facilidade, segundo o tipo de estrutura do item. Em um gráfico de distribuição cumulativa, em cada valor de abscissa (eixo x), o valor apresentado no eixo das ordenadas (eixo y) é o valor acumulado da variável. Nesse caso, a variável independente x é o índice de facilidade, e a variável dependente y é a proporção acumulada de itens, segundo o tipo, para cada valor de x. Por exemplo, com índice de facilidade menor ou igual a 0,4, foram computados em torno de 55% dos itens do tipo 3, 50% dos itens do tipo 2 e 45% dos itens do tipo 1.

Em linhas gerais, em um gráfico de distribuição cumulativa, poligonais mais à esquerda representam itens de tipo mais difícil (índices de facilidade menores) e poligonais mais à direita indicam itens de tipo mais fáceis (índices de facilidade maiores). Logo, pode-se concluir que, em relação aos 471 itens válidos das provas do Enade 2016, os do tipo 3 foram mais difíceis do que os do tipo 2, que, por sua vez, foram mais difíceis do que os do tipo 1. Essa ordenação se altera um pouco ao se acumularem os itens com índice de facilidade acima de 0,5 (mais fáceis), ponto a partir do qual os do tipo 2 parecem ser mais difíceis do que os do tipo 3.

Cabe destacar que a facilidade ou dificuldade de um item objetivo sofre influência de outras características, tais como o nível cognitivo exigido (que será discutido a seguir) e, especialmente, o conteúdo abordado, que pode envolver tanto conhecimentos básicos na área quanto conhecimentos mais aprofundados ou, até mesmo, periféricos.

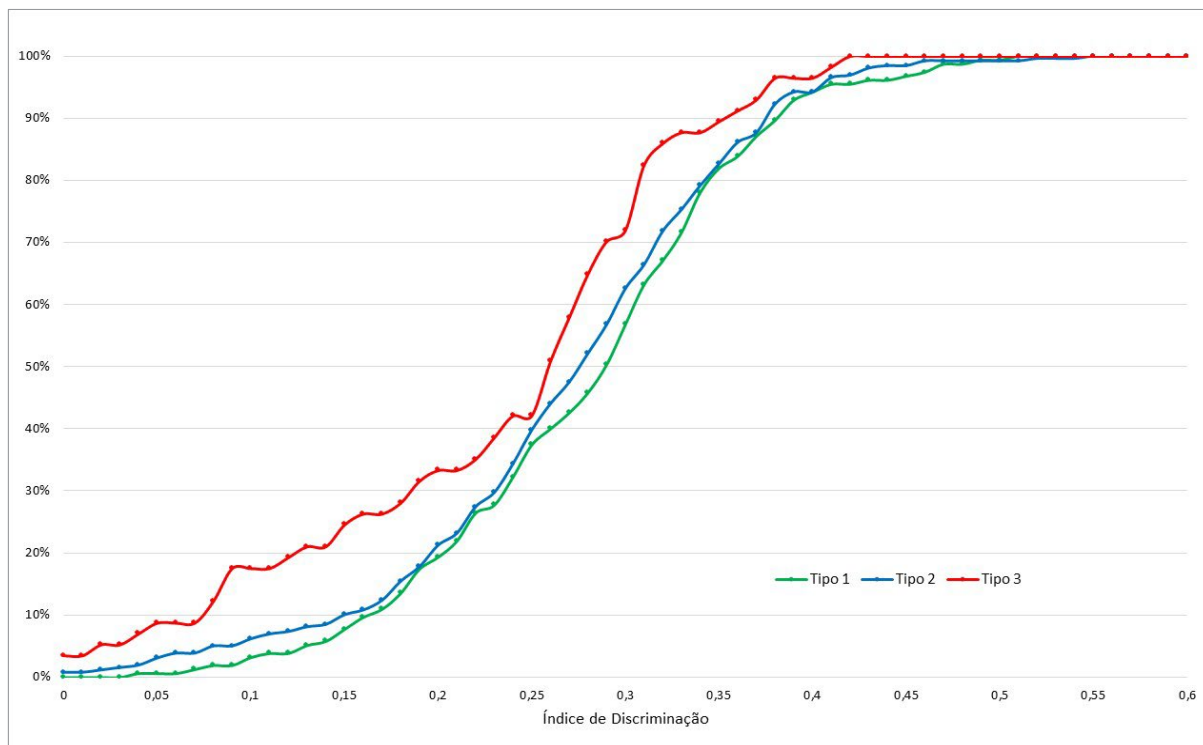
FIGURA 5

Distribuição acumulada do índice de facilidade, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

Na Figura 6, apresentam-se as distribuições cumulativas do índice de discriminação dos itens, segundo o tipo de item. A curva correspondente aos itens do tipo 3 (asserção-razão) está sempre mais à esquerda, indicando que esses foram os com menores índices de discriminação. Destaca-se que o maior valor deste índice para o tipo 3 foi 0,41 e, por isso, a curva atinge 100% neste valor. As poligonais para os outros dois tipos são bem parecidas e próximas em todo o intervalo de dados, com itens do tipo 2 (resposta múltipla) um pouco mais à esquerda do que a curva verde dos itens do tipo 1 (resposta única).

FIGURA 6**Distribuição acumulada do índice de discriminação, por tipo de item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016**

Fonte: Elaboração dos autores, 2020.

CLASSIFICAÇÃO QUANTO AO NÍVEL COGNITIVO DOS ITENS**As categorias utilizadas**

Para estabelecer uma classificação, alguns modelos foram utilizados como fonte. As principais fontes foram a tradicional taxionomia de Bloom (Bloom, 1949) e uma nova versão, conhecida como taxionomia de Bloom revisada, publicada em 2001 (Anderson & Krathwohl, 2001).

Os estudos de Bloom, que tiveram origem no final da década de 1940, nos Estados Unidos, chegaram ao Brasil nos anos 1960 e foram uma das bases do tecnicismo educacional (anos 1960-1970). Embora sua base teórica tenha ficado no esquecimento durante as duas décadas seguintes, sua abordagem voltou a ser objeto de novos estudos na atualidade (Oliveira & Costa, 2001; Rodrigues, 2009; Castro & Moreira, 2010) em função dos exames em larga escala, especialmente os relativos à avaliação e ao acesso ao ensino superior – o antigo Provão, o Enade e o Enem.

Na definição de categorias para a classificação dos itens do ponto de vista do nível cognitivo, a taxionomia de Bloom revisada foi de grande importância. Além de essa nova proposta trazer diversas reflexões sobre a aplicação da taxionomia original, a adequação é apresentada de forma detalhada.

Também foram consultados alguns manuais de elaboração de itens disponíveis na internet, de diversas instituições do campo da avaliação (López Padilla, et al., 2013; QEdu, 2019; Rodrigues, 2002). Apesar de tais documentos não serem teóricos, vale destacar que apresentam categorias baseadas na taxionomia de Bloom, com algumas adaptações e escolhas que se justificam pela característica de serem documentos de trabalho para a elaboração de itens de múltipla escolha.

A partir do estudo dessas referências, foram estabelecidas quatro categorias, definidas a seguir, que possibilitaram a análise pelos mesmos técnicos que realizaram a classificação por tipo.

Seguem as categorias definidas e utilizadas para a classificação dos itens das provas das áreas do Enade 2016.

Memorização

Um item de múltipla escolha é identificado na categoria memorização se a evocação for o processo fundamental para a sua resolução, ou seja, aplicando-se a taxionomia de Bloom – subcategorias da categoria conhecimento –, se o item exige meramente o reconhecimento de informações limitadas a aspectos pontuais ou isolados, como o reconhecimento de terminologia, ou se alcança o nível de informações inter-relacionadas e organizadas, como o reconhecimento de teorias e estruturas.

Compreensão

Um item de múltipla escolha é identificado na categoria compreensão se a sua resolução exige entendimento e apreensão de informações, sem necessidade de relacionar com outras informações / materiais, com implicações mais complexas ou com o todo em que se inserem.

Um item requer compreensão se solicitar do respondente, entre outras habilidades, a reorganização ou o reordenamento de ideias, informações, materiais, além de exigir:

- codificação ou paráfrase que preserve a ideia original, com precisão, ou seja, realizar tradução de uma linguagem para outra;
- explicação ou resumo, que implicam fazer inferência ou extrair conclusões, com base em elementos apresentados;
- significados para além das informações fornecidas, o que envolve determinar implicações, consequências, corolários e efeitos consistentes com as informações originais.

Aplicação/ resolução de problema

Aplicação implica transferência de conhecimentos em situações novas, envolve capacidade de generalizar. Exige que o respondente, utilizando abstrações

pertinentes, recorra ao conhecimento e à compreensão de terminologia / fatos específicos / convenções / tendências e sequências / classificações e categorias / critérios / metodologias / princípios e generalizações / teorias e estruturas.

Para ser classificado na categoria aplicação, o item deve apresentar situações fictícias, porque é necessário que as situações propostas sejam desconhecidas, além de conter elementos novos ou problemas em perspectiva diferente.

A resolução de problema requer raciocínio dedutivo e indutivo, envolvendo capacidade de tomar decisões.

Na taxionomia de Bloom, resolver problema é uma habilidade incluída em aplicação.

Análise

Análise implica fracionar o todo em suas partes constitutivas, perceber as relações entre as partes e evidenciar a compreensão sobre a forma em que se organizam.

O respondente deve demonstrar capacidade de distinguir fatos de hipóteses, identificar conclusões de modo fundamentado, discriminar materiais relevantes e estabelecer relações entre ideias.

É importante verificar se a resolução do problema proposto consiste em identificar:

- elementos que o constituem, considerando a pertinência, a importância e a significação de características, funções, fatos, intenções, valores, entre outros, envolvidos nas formulações;
- conexões e interações entre elementos do problema, material ou situação, bem como relações entre as suas partes e a relevância entre elementos ou partes com a ideia central;
- estrutura ou organização implícita que se configura em situações, problemas, textos, conjunto de dados, relatos de experiência, entre outros, possível, em geral, por meio de inferência.

Classificação dos itens, segundo o nível cognitivo

Todas as 27 questões de cada uma das áreas participantes do Enade 2016 foram classificadas segundo as categorias brevemente apresentadas na seção anterior. A metodologia para o estabelecimento das classificações seguiu a mesma estratégia já descrita no caso da tipologia por estrutura do item. Ou seja, inicialmente três técnicos em avaliação fizeram as classificações individualmente, e os casos de discrepância foram resolvidos em reunião, com a participação de um quarto profissional.

ME – memorização

CO – compreensão

AP – aplicação

AN – análise

Os resultados estão apresentados na Tabela 7. A classificação realizada mostrou que as provas privilegiaram itens de memorização, ou seja, aqueles que exigem apenas o resgate de informações relativas a aspectos pontuais ou isolados, como uma terminologia, ou informações inter-relacionadas e organizadas, como o reconhecimento de teorias ou estruturas, que podem ser memorizadas. Considerando-se o conjunto das provas, constatou-se que, dos 486 itens, 276 (56,79%), mais da metade, portanto, eram de memorização. Em segundo lugar, estiveram presentes os itens de compreensão (19,34%), seguidos pelos de aplicação (16,67%) e, em quantidade bem inferior, pelos de análise (7,20%). Níveis mais complexos de processos cognitivos tiveram, gradativamente, menor participação nas provas. A situação foi um pouco diferente em relação ao componente de formação geral.

TABELA 7

Itens utilizados por nível cognitivo, segundo áreas participantes do Enade 2016

ÁREAS	ME	CO	AP	AN
Agronomia	23	1	3	-
Biomedicina	19	4	4	-
Educação Física	18	7	-	2
Enfermagem	17	3	6	1
Farmácia	11	2	8	6
Fisioterapia	8	4	13	2
Fonoaudiologia	11	9	7	-
Medicina	9	6	12	-
Medicina Veterinária	22	1	4	-
Nutrição	23	4	-	-
Odontologia	11	7	8	1
Serviço Social	4	18	1	4
Zootecnia	19	2	4	2
Tecnologia em Agronegócio	14	3	4	6
Tecnologia em Estética e Cosmética	18	7	2	-
Tecnologia em Gestão Ambiental	14	3	2	8
Tecnologia em Gestão Hospitalar	18	7	1	1
Tecnologia em Radiologia	17	6	2	2
TOTAL	276 (56,79%)	94 (19,34%)	81 (16,67%)	35 (7,20%)

Fonte: Elaboração dos autores, 2020.

As oito questões objetivas de formação geral eram dos seguintes tipos:

- três de memorização (questões 2, 4 e 8);
- quatro de compreensão (questões 1, 3, 6 e 7);
- nenhum item de aplicação;
- uma de análise (questão 5).

Classificação das questões anuladas

Primeiramente, buscou-se caracterizar o nível cognitivo dos itens que foram anulados pelas Comissões Assessoras de Área (CAA) e por índice de discriminação fraco (IDF), o que pode ser observado na Tabela 8. Tendo em vista que as quantidades de itens dos quatro níveis cognitivos utilizados neste estudo eram muito diferentes, vale observar o percentual de anulados de cada tipo em relação ao total de itens daquele mesmo tipo (ver última linha da Tabela 8). Como resultado, foram anulados 34,29% dos itens de análise, 23,40% dos itens de compreensão, 20,99% dos itens de aplicação e 19,93% dos itens de memorização.

TABELA 8

Itens anulados por nível cognitivo e motivo de anulação, segundo áreas participantes do Enade 2016

ÁREAS	ME		CO		AP		AN	
	CAA	IDF	CAA	IDF	CAA	IDF	CAA	IDF
Agronomia	1	1	0	0	1	0	0	0
Biomedicina	0	4	1	0	0	2	0	0
Educação Física	1	0	0	1	0	0	1	0
Enfermagem	0	3	0	2	0	0	0	0
Farmácia	0	2	0	0	0	3	0	3
Fisioterapia	0	2	1	1	1	1	0	0
Fonoaudiologia	1	3	1	1	0	2	0	0
Medicina	0	3	0	0	0	2	0	0
Medicina Veterinária	0	4	0	0	0	0	0	0
Nutrição	0	4	0	0	0	0	0	0
Odontologia	0	2	0	2	0	0	0	0
Serviço Social	1	0	2	3	0	0	1	1
Zootecnia	1	5	0	0	0	0	0	0
Tecnologia em Agronegócio	0	3	0	1	0	3	0	3
Tecnologia em Estética e Cosmética	1	4	0	0	0	1	0	0

(continua)

(continuação)

ÁREAS	ME		CO		AP		AN	
	CAA	IDF	CAA	IDF	CAA	IDF	CAA	IDF
Tecnologia em Gestão Ambiental	0	1	0	1	0	0	0	2
Tecnologia em Gestão Hospitalar	0	4	0	2	0	0	0	1
Tecnologia em Radiologia	0	4	0	3	0	1	0	0
TOTAL	6	49	5	17	2	15	2	10
% relativo ao total de itens (486)	1,23	10,8	1,03	3,50	0,41	3,09	0,41	2,06
% relativo ao total de anuladas (106)	5,66	46,23	4,74	16,04	1,89	14,15	1,89	9,43
% relativo aos itens da mesma categoria	2,17	17,75	5,32	18,09	2,47	18,52	5,71	28,57

Fonte: Elaboração dos autores, 2020.

Como registrado na análise dos tipos de estrutura dos itens, dos 486 itens de conhecimentos específicos, 15 (3,09% do total) foram anulados pelas CAA e 91 (18,52% do total) pelo índice de discriminação. Assim, como já comentado na análise por tipo de item, 106 itens (21,81% do total) não foram computados nas notas dos estudantes.

Na classificação por nível cognitivo dos itens anulados, observa-se que a maior quantidade é do tipo memorização, o que está associado à maior presença dessa categoria de item nas provas objetivas do Enade 2016. Considerando o total de 106 itens anulados, os percentuais de anulações por tipo de item foram:

- 51,89% de ME;
- 20,75% de CO;
- 16,04% de AP;
- 11,29% de AN.

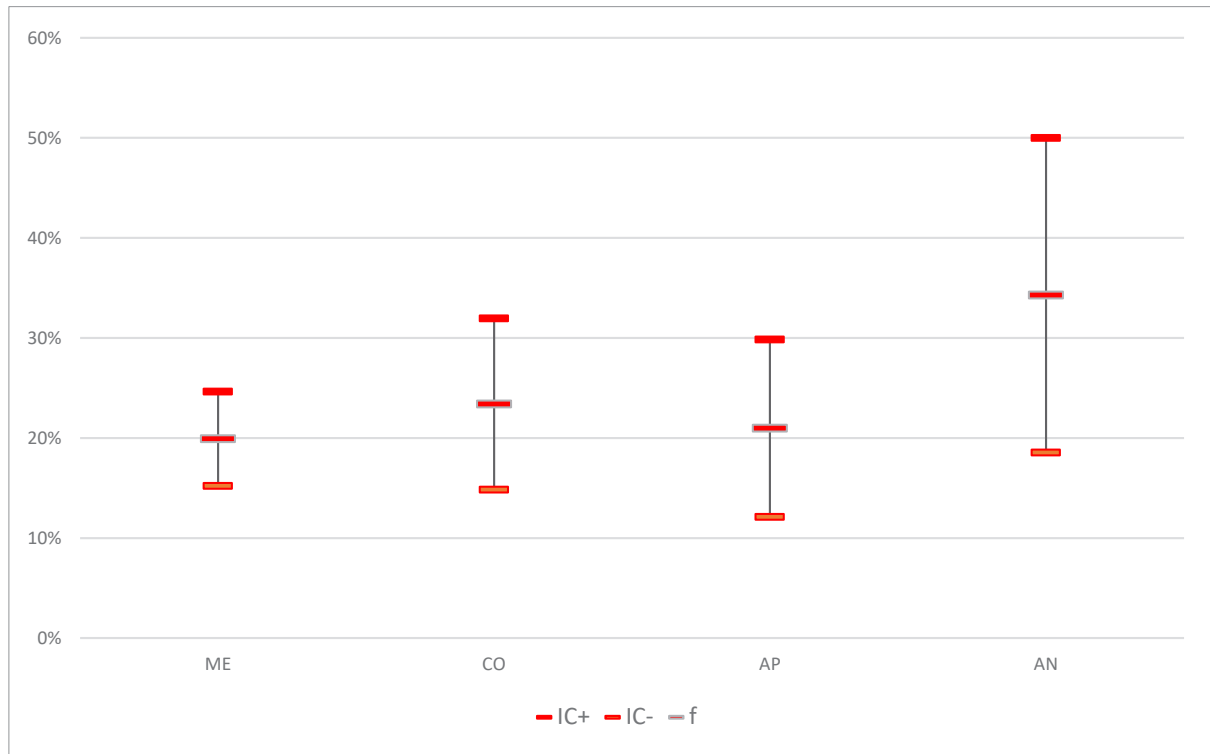
Os percentuais de itens eliminados relativamente à quantidade por nível cognitivo foram:

- 19,93% de nível ME (55 em 276);
- 23,40% de nível CO (22 em 94);
- 20,99% de nível AP (17 em 81);
- 34,29% de nível AN (12 em 35).

Na Figura 7, apresentam-se as proporções de questões anuladas com os respectivos intervalos de confiança de 95%, por nível do processo cognitivo. A superposição dos intervalos indica que não há diferenças estatisticamente significativas entre essas proporções (confirmado por testes paramétricos). Assim, não se pode dizer que questões de algum nível cognitivo específico teriam maior ou menor probabilidade de serem anuladas, nessa edição do Enade.

FIGURA 7

Proporção de questões anuladas, segundo o nível cognitivo dos itens das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

Estatísticas da distribuição dos índices de facilidade e de discriminação (bisserial), por nível cognitivo do item

Nas tabelas 9 e 10, podem ser observadas, respectivamente, as estatísticas dos índices de facilidade e de discriminação do total de 471 itens não anulados pelas CAA, segundo o nível cognitivo. As menores médias, para ambos os índices, foram as dos itens com nível de análise.

TABELA 9

Índice de facilidade, segundo nível cognitivo

NÍVEL COGNITIVO	N	MEDIANA	MÉDIA	IC 95%	
				LI	LS
ME	270	0,42	0,45	0,43	0,48
CO	89	0,39	0,42	0,38	0,46
AP	79	0,47	0,48	0,43	0,53
AN	33	0,31	0,34	0,28	0,41

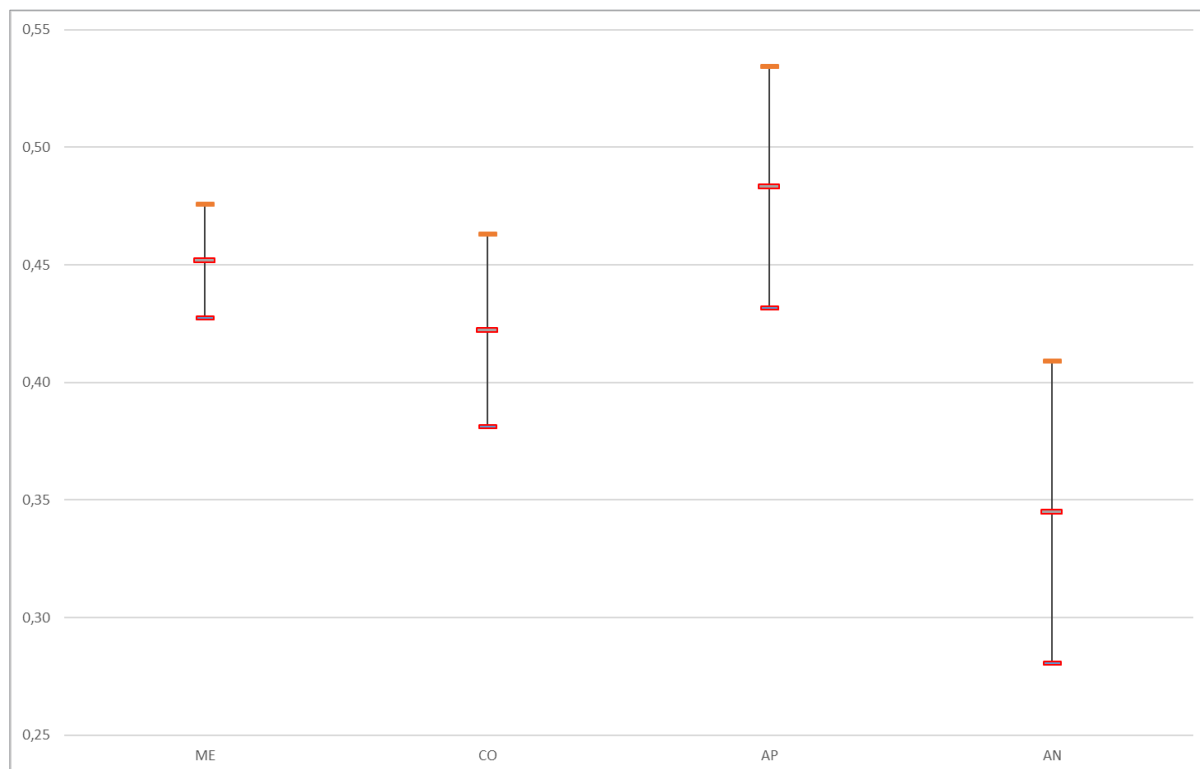
Fonte: Elaboração dos autores, 2020.

Na Figura 8, são apresentadas as médias dos índices de facilidade, com os respectivos intervalos de confiança de 95%, por nível cognitivo do item. Como ocorreu

para o índice de facilidade, não há diferença estatisticamente significativa entre as médias dos itens dos diferentes níveis cognitivos.

FIGURA 8

Média do índice de facilidade e IC de 95%, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

TABELA 10

Índice de discriminação, segundo nível cognitivo

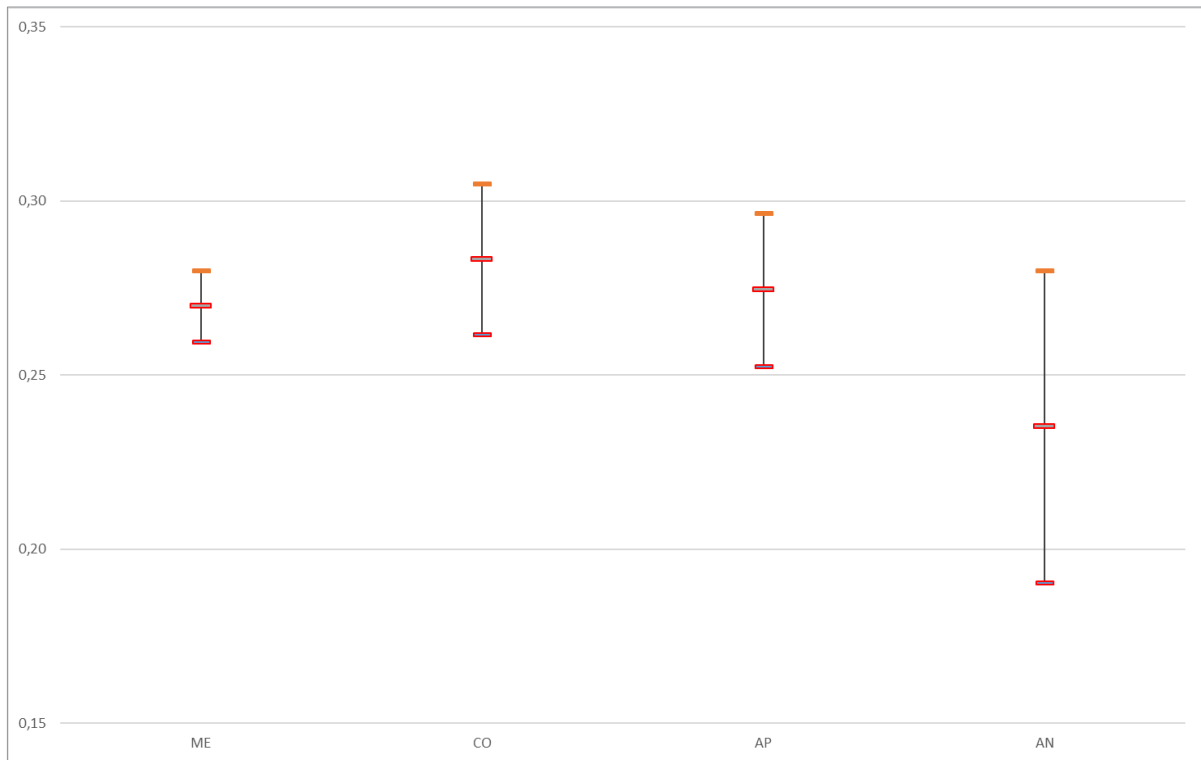
NÍVEL COGNITIVO	N	MEDIANA	MÉDIA	IC 95%	
				LI	LS
ME	221	0,28	0,27	0,26	0,28
CO	72	0,29	0,28	0,26	0,30
AP	64	0,28	0,27	0,25	0,30
AN	23	0,26	0,24	0,19	0,28

Fonte: Elaboração dos autores, 2020.

Na Figura 9, são apresentadas as médias dos índices de discriminação com os respectivos intervalos de confiança de 95%, por nível cognitivo do item. Não se pode afirmar que as diferenças das médias sejam estatisticamente significativas, na comparação entre quaisquer níveis cognitivos.

FIGURA 9

Média do índice de discriminação e IC de 95%, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

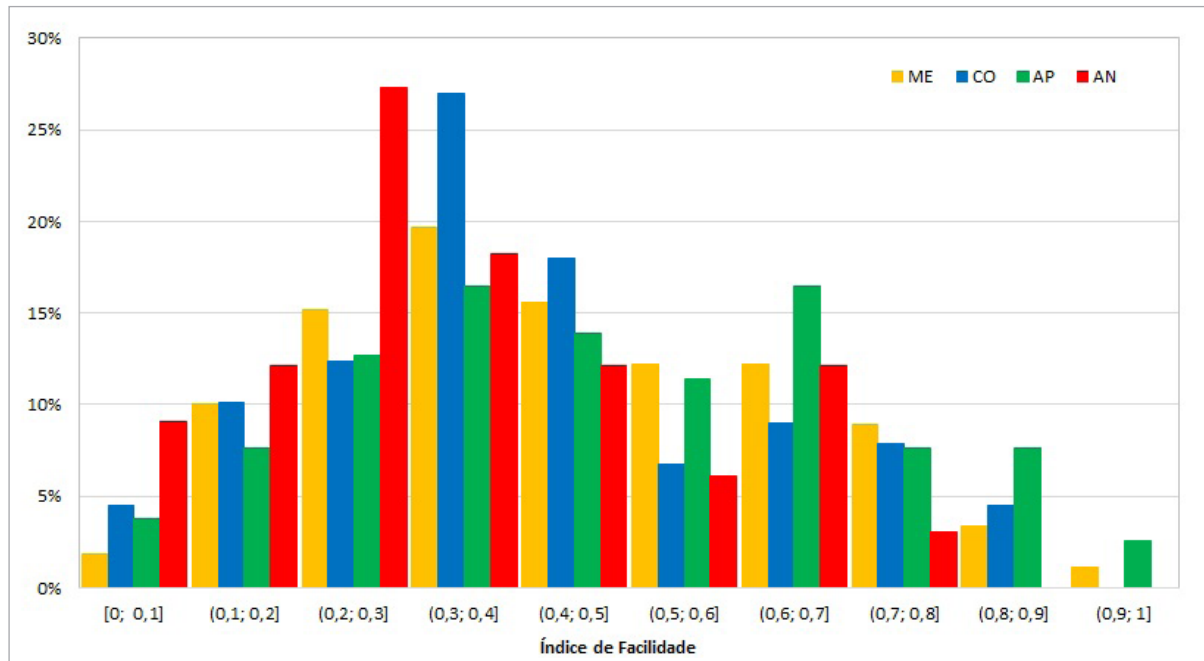
Nas figuras 10 e 11, apresentam-se, respectivamente, as distribuições de frequência relativa dos índices de facilidade e de discriminação, em intervalos, segundo o nível cognitivo dos itens.

Considerados os valores agrupados em intervalos, a moda do índice de facilidade para itens de análise ficou no intervalo (0,2; 0,3]. Para os itens de memorização e compreensão, a moda ficou no intervalo imediatamente superior, (0,3; 0,4], e para os itens de aplicação, o intervalo modal foi bem superior, (0,6; 0,7], seguido de perto pelo intervalo (0,3; 0,4].

Merece destaque, também, que não houve item de compreensão e de análise com índice de facilidade no intervalo máximo. As distribuições dos itens desses níveis também são mais concentradas em torno da moda, enquanto as distribuições dos itens dos níveis de memorização e de aplicação são mais espalhadas pelos intervalos.

FIGURA 10

Distribuição do índice de facilidade, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016

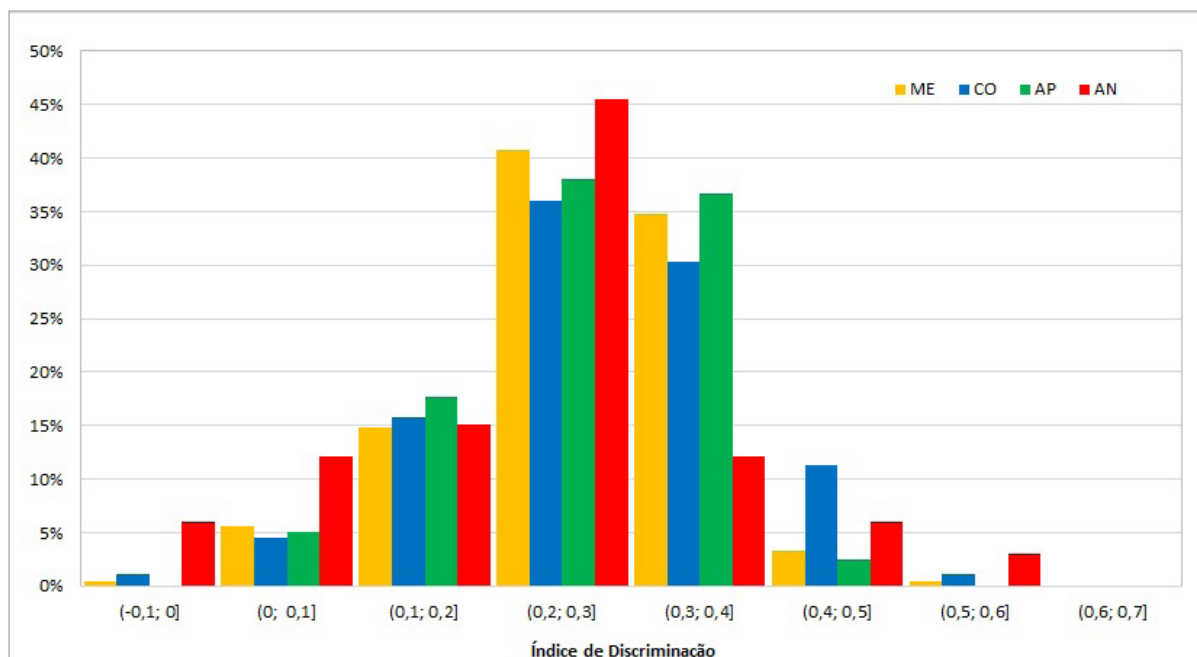


Fonte: Elaboração dos autores, 2020.

Quanto ao índice de discriminação, observa-se que o intervalo modal para os quatro tipos é (0,2; 0,3], sendo que para os itens de aplicação, o intervalo (0,3; 0,4] tem frequência relativa próxima da do intervalo modal. Apesar de a distribuição do índice de discriminação dos itens de análise estar mais à esquerda (índices mais baixos), no intervalo máximo, o dos itens com maior poder de discriminação, percentualmente, há mais itens de análise, seguidos pelos de compreensão e de memorização. Não houve registro de nenhum item de aplicação.

FIGURA 11

Distribuição do índice de discriminação, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

Distribuição acumulada dos índices de facilidade e de discriminação, por nível cognitivo do item

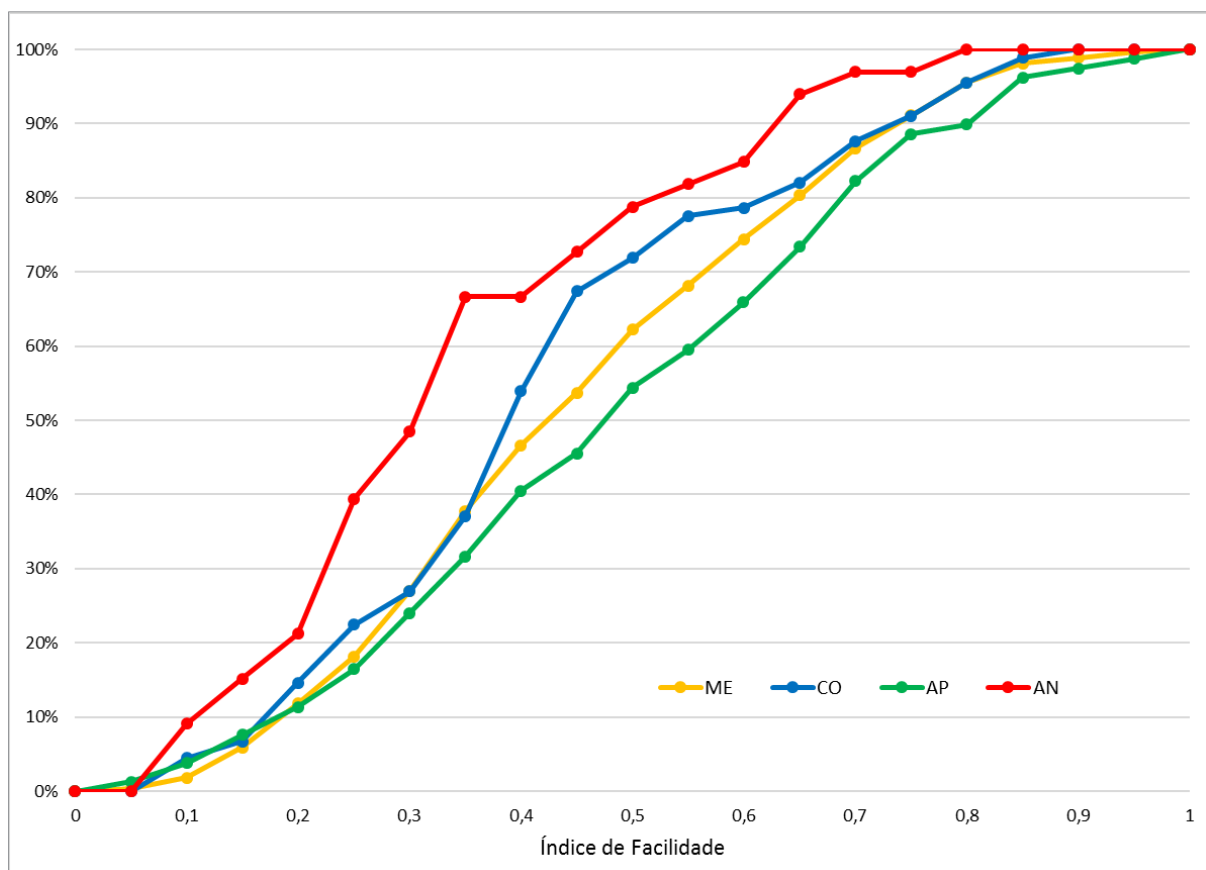
Na Figura 12, assim como na Figura 5, apresenta-se a distribuição cumulativa do índice de facilidade, agora segundo a classificação de nível cognitivo do item. Em cada valor de abscissa (eixo x), o valor apresentado no eixo das ordenadas (eixo y) é o valor acumulado de y. Por exemplo, em torno de 68% dos itens com nível cognitivo de análise (AN) tiveram índice de facilidade menor ou igual a 0,4. Também com índice de facilidade menor ou igual a 0,4 ficaram 55% dos itens de compreensão (CO), 48% dos itens de memorização (ME) e 40% dos itens de aplicação (AP).

Em linhas gerais, em um gráfico de distribuição cumulativa, poligonais mais à esquerda representam itens mais difíceis (índices de facilidade menores), enquanto poligonais mais à direita indicam itens mais fáceis (índices de facilidade maiores). Logo, pode-se concluir que, em relação aos 471 itens das provas de Enade 2016, os de nível cognitivo de análise foram mais difíceis do que os de compreensão, memorização e aplicação, nessa ordem. Pode-se observar na Figura 12 que as distribuições cumulativas do índice de facilidade dos itens de memorização e de compreensão se superpõem em alguns intervalos. Já a frequência acumulada por índice de facilidade de itens de análise está sempre mais à esquerda, enquanto a de itens de aplicação está sempre mais à direita, indicando, respectivamente, os mais difíceis e os mais fáceis.

Cabe lembrar, mais uma vez, que a facilidade ou a dificuldade de um item objetivo é afetada por outras características, especialmente pelo conteúdo abordado. Um item de memorização, por exemplo, pode cobrar um conhecimento básico, ao passo que outro, também de memorização, pode exigir conhecimentos mais aprofundados ou, até mesmo, periféricos.

FIGURA 12

Distribuição acumulada do índice de facilidade, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016

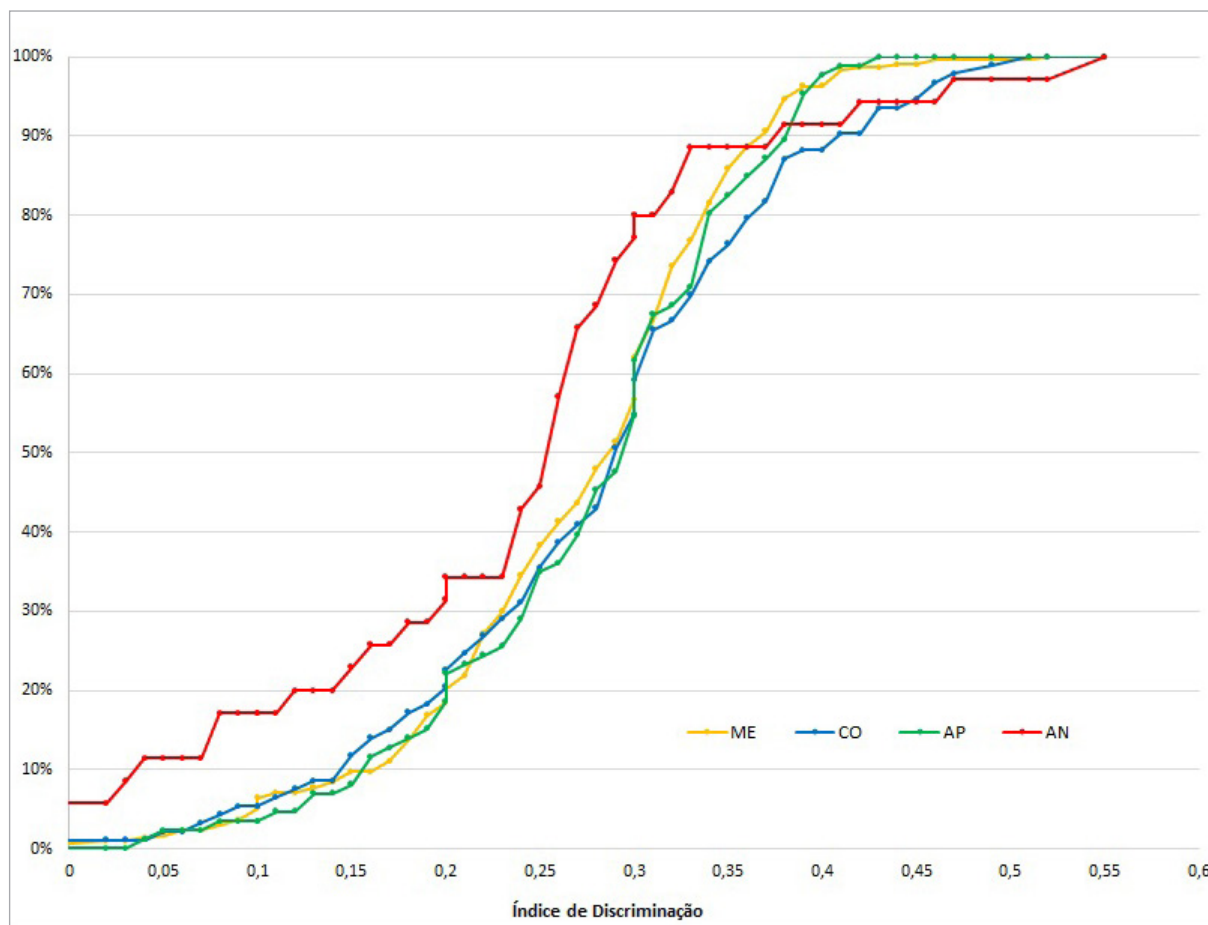


Fonte: Elaboração dos autores, 2020.

Na Figura 13, apresentam-se as distribuições cumulativas do índice de discriminação dos itens, segundo o nível cognitivo. A curva correspondente aos itens de análise está sempre mais à esquerda, indicando que esses foram os com menores índices. As poligonais para os outros três níveis cognitivos – memorização, compreensão e aplicação – se entrecruzam na maior parte do intervalo de dados. A exceção acontece nos valores mais altos do índice, praticamente só com casos de análise e de compreensão.

FIGURA 13

Distribuição acumulada do índice de discriminação, por nível cognitivo do item das provas objetivas de conhecimentos específicos das áreas participantes do Enade 2016



Fonte: Elaboração dos autores, 2020.

ANÁLISE DOS ITENS PELA ORDEM DE APRESENTAÇÃO NAS PROVAS

Uma hipótese frequentemente apontada em provas, em particular no Enade, é a de que a ordem de apresentação dos itens na prova pode influenciar o resultado. Supõe-se que o cansaço possa levar a uma espécie de correlação entre a ordem do item e o aumento dos erros.

Para verificar a hipótese de que a facilidade possui relação linear com a ordem dos itens, recorreu-se ao modelo de regressão linear simples, ajustado para cada área, dado por

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (2)$$

onde:

Y_i é a variável que se pretende explicar (dependente) – índice de facilidade do i -ésimo item de uma determinada área;

X_i é a variável explicativa (independente) – ordem do item na prova;

α é uma constante, valor em que a reta de regressão cruza o eixo vertical (quando $x = 0$);

β é a inclinação (coeficiente angular) da reta de regressão;

ε_i representa todos os fatores residuais e erros de medição.

Se não houver relação entre as variáveis, então $\beta = 0$ e, se houver, $\beta \neq 0$. Os resultados estão apresentados na Tabela 11.

TABELA 11

Efeito estimado (β) da ordem dos itens sobre o índice de facilidade, segundo áreas¹ participantes do Enade 2016

ÁREAS	β	v (β)
Agronomia	0,00548	0,00385
Biomedicina	-0,00942	0,06110
Educação Física	0,00312	0,00274
Enfermagem	0,00382*	0,00171
Farmácia	0,00145	0,00334
Fisioterapia	-0,00228	0,01662
Fonoaudiologia	-0,00755	0,07674
Medicina	-0,00083	0,00841
Medicina Veterinária	-0,00044	0,00466
Nutrição	-0,00996	0,00843
Odontologia	0,00234	0,00451
Serviço Social	-0,00061	0,00896
Zootecnia	0,00284	0,00734
Tecnologia em Agronegócio	0,00371*	0,00155
Tecnologia em Estética e Cosmética	-0,00477	0,08134
Tecnologia em Gestão Ambiental	0,00029	0,00272
Tecnologia em Gestão Hospitalar	-0,01372	0,00753
Tecnologia em Radiologia	-0,00363	0,00412

Fonte: Elaboração dos autores, 2020.

¹ Estimativa via regressão linear simples.

* Estatisticamente significativo a 5%.

A ordem de apresentação dos itens na prova do Enade 2016, no atual modelo, parece não ter influenciado o índice de facilidade dos itens, já que a estimativa do coeficiente β no ajuste linear proposto não foi estatisticamente relevante em 16 das 18 áreas. Subjacente ao teste, estava a hipótese de que a taxa de acertos decairia com a ordem dos itens (inclinação negativa). Somente nas áreas de enfermagem e tecnologia em agronegócios, o coeficiente angular foi significativo, com valores um pouco abaixo de 0,4%, mas positivo, indicando a situação inversa, ou seja, mais acertos nas questões de maior ordem.

CONCLUSÕES E RECOMENDAÇÕES

Antes de se apresentarem conclusões sobre os resultados do estudo, para contextualizá-los de forma adequada, três aspectos são destacados e comentados nos parágrafos a seguir.

O primeiro destaque é para o fato de este ser um estudo tipicamente empírico, ou seja, realizado a partir da realização das provas pelos formandos de cada área do Enade 2016. Assim, foi possível recorrer a duas medidas associadas às respostas dadas pelos participantes aos itens: o índice de facilidade e o índice de discriminação.

Em segundo lugar, na mesma linha, as classificações por tipo de estrutura dos itens e por nível cognitivo associado aos itens foram realizadas por nossas equipes de técnicos, após a prova ter sido aplicada e divulgada. Até onde temos informação, tais classificações não são feitas previamente nem são usadas para a composição das provas, nem pelos técnicos do Inep nem pelos docentes das Comissões Assessoras das diferentes áreas, que se responsabilizam, junto com o Inep, pela construção do instrumento. Assim, como foi possível observar, não há uma distribuição semelhante de itens por tipo ou por nível nas provas das diferentes áreas. Ao contrário, pode-se observar certa concentração, nas duas classificações, em um tipo (item de resposta múltipla, com 55,56% de presença nas provas) e em um nível (item de memorização, com 56,79% dos itens das provas). Nossa hipótese é a de que esse aspecto parece mais associado à facilidade de elaboração de itens com tais características pelos docentes que participam das equipes de elaboradores do Banco Nacional dos Itens (BNI) do que a uma escolha justificada por alguma base teórica.

Por fim, o terceiro aspecto que precisa ser destacado, antes de nossas considerações sobre os resultados, é a necessidade de ampliação deste estudo para se analisarem os conteúdos avaliados, no sentido de verificar se houve uma preocupação em contemplar uma cobertura ampla dos currículos, prescritos pelas Diretrizes Curriculares Nacionais (DCN) dos cursos, ou se houve algum tipo de concentração ou ênfase que possa ter influenciado os resultados. Cabe salientar que o Inep prescreve que a seleção dos itens seja pautada por uma matriz curricular que inter Cruze prescrições das DCN, considere o perfil do formando, mobilize recursos associados às competências e habilidades e selecione os conteúdos de ensino a serem avaliados. Sem dúvida, como já comentado, os índices de facilidade e de discriminação têm correlação com a dificuldade do conteúdo avaliado, a centralidade dos conceitos na formação, a sua atualidade e a ênfase de abordagem que é dada, ou não, nos diferentes cursos, por exemplo.

Levando-se tais aspectos em conta, a partir das análises apresentadas até aqui, é importante voltar às questões iniciais que motivaram este estudo. Primeiramente, chamou a atenção a primazia de itens de respostas múltiplas (tipo 2) e de memorização nas provas. Como mais da metade das provas são baseadas em itens como esses,

considera-se que a avaliação é, de certa forma, tendenciosa. É sinalizado aos cursos que conteúdos que exigem a simples memorização são os mais importantes e que se deve valorizar a habilidade de ler e interpretar afirmações para decidir quais são falsas ou verdadeiras (tipo 2 de item).

A partir da associação do índice de facilidade ao tipo de item, apesar da ênfase destacada no parágrafo anterior, *grosso modo*, foi possível identificar que os três tipos de item foram, em sua maioria, *difíceis* para os estudantes, já que a moda ocorre no intervalo (0,3; 0,4] e as médias foram bem próximas de 0,4 (ver Tabela 1). No entanto, observou-se que os itens do tipo 3 foram mais difíceis do que os do tipo 2, que, por sua vez, foram mais difíceis do que os do tipo 1 (ver Figura 5). Ao se analisarem as estatísticas básicas do índice de facilidade, por tipo de item, verificou-se que os valores médios do indicador decaem com a complexidade do tipo de item, apesar de não haver diferença estatisticamente significativa, ao nível de 95%, entre tais médias (ver Figura 1).

Quanto ao índice de discriminação (ver Tabela 2), a maioria dos itens teve capacidade média de diferenciar os estudantes, sendo que os itens do tipo 1, os de resposta única, tiveram uma capacidade de discriminação um pouco maior do que a dos itens dos outros dois. Merece destaque que, também para o índice de discriminação, os valores médios decaem com a complexidade do tipo de item (ver Figura 1). E, nesse caso, observa-se uma diferença estatisticamente significativa entre as médias da capacidade de discriminar dos itens dos tipos 1 (resposta única) e tipo 3 (asserção-razão).

Cabe observar que, para itens do tipo 2, o fato de os intervalos de confiança das médias, tanto do índice de facilidade quanto do índice de discriminação, serem menores, está associado à primazia da presença de itens desse tipo nas provas.

Uma segunda questão de pesquisa proposta foi verificar a hipótese de parecer natural que um item que exija um nível cognitivo mais elaborado seja mais difícil do que outro com nível cognitivo menos exigente. Como discutido a seguir, o estudo confirma parcialmente tal hipótese. A classificação utilizada no estudo aqui apresentado se distribuiu em quatro níveis cognitivos – memorização, compreensão, aplicação e análise – que, segundo as teorias desse campo, aumentam de complexidade, nessa ordem.

Verificou-se que tanto o índice de facilidade quanto o índice de discriminação têm médias mais baixas nos itens do nível de análise. Na Figura 10, a da distribuição do índice de facilidade, foi possível observar que, nos itens de análise, o intervalo modal foi (0,2; 0,3]. Nos itens de memorização e compreensão, a moda ficou no intervalo imediatamente superior, (0,3; 0,4] e nos itens de aplicação o intervalo modal foi bem superior: (0,6; 0,7]. Tal resultado indica que os itens mais fáceis para os respondentes foram os de aplicação e os mais difíceis foram os de análise, seguidos

pelos de compreensão (não houve item de compreensão e de análise com índice de facilidade no intervalo máximo).

Esse resultado é corroborado pela Figura 11, de distribuição cumulativa do índice de facilidade por nível cognitivo, no qual se constata que da esquerda para a direita (dos itens com menor facilidade para a maior) a ordem é análise, compreensão, memorização e aplicação. De certa forma, o fato de os itens de aplicação serem os mais acertados pode estar associado às atividades mais valorizadas durante a formação. Especialmente no final dos cursos, é natural que a aplicabilidade dos conhecimentos receba mais atenção.

Quanto aos índices de discriminação dos itens por nível cognitivo, os itens de aplicação, os mais fáceis, têm menor capacidade de discriminar os estudantes por grupos de desempenho. Já os de nível cognitivo de análise, seguidos pelos de compreensão, têm maior poder de discriminação.

Grosso modo, considerando-se que as médias dos índices de discriminação por nível cognitivo dos itens variaram de 0,24 a 0,27, consultando-se a Tabela 2, é possível afirmar que os quatro tipos de itens parecem ter tido capacidade média de discriminação.

Também buscou-se identificar se algum tipo de item teria maior probabilidade de ser anulado pelo bisserial e se itens de algum nível cognitivo poderiam ter essa chance aumentada. Quanto ao tipo de estrutura do item, no conjunto dos itens de tipo 3 (asserção-razão), 31,58%, o equivalente a um terço, foram anulados, enquanto a proporção de anulação dos outros dois tipos ficou em torno de um quinto. Apesar disso, as probabilidades de um item ser anulado em função de seu tipo não possuem diferenças estatisticamente significativas ao nível de 95%, o que, em grande medida, é devido ao pequeno quantitativo de itens do tipo 3 presentes na prova.

Considerando-se o nível cognitivo, a proporção de anulação dentro dos conjuntos de itens de um mesmo nível parece indicar que os de análise sofreram mais anulações (34,29%), em relação aos demais grupos: memorização (19,93%), compreensão (23,40%) e aplicação (20,99%). Apesar disso, as probabilidades de um item ser anulado não apresentam diferenças estatisticamente significativas ao nível de 95%, segundo o nível cognitivo envolvido no processo de solução da questão.

Também se investigou a hipótese de que os acertos tenderiam a diminuir ao longo da prova, ou seja, de que poderia haver um efeito de cansaço ou de tempo insuficiente e, assim, itens colocados mais para o final seriam menos acertados do que itens do início da prova. Essa hipótese foi rejeitada. Um ajuste linear teve coeficientes estatisticamente não significativos (diferentes de zero) em 16 das 18 áreas analisadas. Em enfermagem e tecnologia em agronegócios, o coeficiente angular foi significativo, com valores um pouco abaixo de 4%, porém positivo, indicando a situação inversa, ou seja, mais acertos nas questões de maior ordem.

O estudo apresentado neste trabalho leva a apontar algumas recomendações, umas visando ao aprimoramento das análises aqui apresentadas e outras na perspectiva da elaboração de itens e da seleção dos itens que serão utilizados na construção das provas.

Na primeira perspectiva, recomenda-se que novas análises possam contar com quantitativos semelhantes de tipos de item e de nível cognitivo dos mesmos. Observou-se, na análise dos mais de quase 500 itens das provas do Enade 2016, uma concentração de tipo de item (tipo 2 – respostas múltiplas) e de nível cognitivo (memorização). Sem dúvida, esse tipo de concentração pode causar viés nos resultados. Idealmente, seria bom que para todos os tipos houvesse itens de todos os níveis cognitivos, de tal modo que fosse possível verificar se há correlações que melhorem a capacidade de discriminar, por exemplo.

Na segunda perspectiva, seria importante que os elaboradores fossem capacitados para apresentar a classificação de nível cognitivo que o item pretende avaliar, além de indicar o grau de dificuldade que se supõe que o item possua, e também que fossem estimulados a elaborar uma quantidade equivalente de itens de todos os tipos recomendados. Nossa experiência com equipes de elaboradores tem mostrado que há tipos que podem ser mais facilmente construídos a partir de um texto básico (resposta múltipla), enquanto outros (resposta única) parecem desafiar mais os elaboradores pela necessidade de se listarem quatro distratores verossímeis. Destaca-se, ainda, a necessidade de delimitação clara e bastante específica do conteúdo que se pretende avaliar. A ideia de um banco de itens com produção permanente, contendo boa quantidade e diversidade de itens, é o que pode contribuir para a construção de boas provas.

REFERÊNCIAS

- Anderson, L. W., & Krathwohl, D. R. (2001). *Uma taxonomia para aprender, ensinar e avaliar: Uma revisão da taxonomia de objetivos educacionais da Bloom*. Longman.
- Bloom, B. S. (1972). *Taxionomia de objetivos educacionais 1: Domínio cognitivo*. Globo.
- Castro, R. da S., & Moreira, E. V. (2010). Exame Nacional de Desempenho de Estudantes: Avaliação ou exame? *EDUCAmazônia*, 4(1), 1-10.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2017a). *Guia de elaboração e revisão de itens*. Inep.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2017b). *Enade 2017 – Relatório síntese de área: Arquitetura e Urbanismo*. Inep.
- López Padilla, A., Sánchez Restrepo, H. S., Espinosa Rodríguez, J. D., & Carmona Soto, M. B. (2013). *Elaboración de ítems de opción múltiple*. Instituto Nacional de Evaluación Educativa.
- Oliveira, E. da S. G. de. & Costa, M. de A. (2001). Por um zero consciente? Discussão proativa do Exame Nacional de Cursos. *Revista Brasileira de Política e Administração da Educação*, 17(2), 221-232.

QEdU Academia. (2019). *Guia prático de elaboração de itens: Breves recomendações sobre como criar itens melhores para as suas provas*. QEdU.

Rodrigues, J. F. (2009). *Avaliação do estudante universitário*. Senac.

Rodrigues, S. G. (2002). Texto adaptado de documento do MEC/Inep para elaboração de itens. In L. Prazeres (Org.), *Normas técnicas para a elaboração de itens*. Universidade Federal de Minas Gerais.