



INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

NOTA TÉCNICA Nº 33/2017/CGCQES/DAES

PROCESSO Nº 23036.004395/2017-64

INTERESSADO: DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO SUPERIOR

1. OBJETIVO

O objetivo desta Nota Técnica é apresentar a metodologia de cálculo do Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD), referente ao ano de 2016.

2. INTRODUÇÃO

O IDD é um indicador de qualidade que mede o valor agregado pelo curso ao desenvolvimento dos estudantes concluintes, considerando seus desempenhos no Enade e suas características de desenvolvimento ao ingressar no curso de graduação avaliado.

O IDD é calculado e divulgado para cada curso de graduação avaliado, identificado pelo código de curso constante no sistema e-MEC utilizado pelas IES para inscrição de estudantes e enquadramento em uma das áreas de avaliação do Enade elencadas no artigo 1º da Portaria Normativa do MEC nº 5, de 9 de março de 2016 (disponíveis no Anexo I), de acordo com a metodologia explicitada nesta Nota Técnica. Até 2015, o IDD não era divulgado individualmente, mas apenas como um dos componentes do Conceito Preliminar de Curso (CPC).

3. INFORMAÇÕES UTILIZADAS PARA O CÁLCULO DO IDD

O cálculo do IDD, realizado por código de curso, leva em consideração as seguintes informações:

- a) número de estudantes concluintes participantes no Enade;
- b) desempenho geral dos estudantes concluintes no Enade;
- c) desempenho dos estudantes no Enem nas áreas de Ciências da Natureza, Ciências Humanas, Linguagens e Códigos e Matemática e suas Tecnologias;
- d) número de concluintes participantes no Enade com nota do Enem recuperada para o cálculo do IDD.

4. CONDIÇÕES PARA QUE UM CURSO TENHA O IDD CALCULADO

Para que um curso tenha o IDD calculado, é preciso que ele atenda às seguintes condições:

- a) Possuir no mínimo 2 (dois) estudantes concluintes participantes do Enade com dados recuperados da base de dados do Enem no período entre o ano de ingresso no curso avaliado e os 3 (três) anos anteriores;
- b) Atingir 20% (vinte por cento) do total de estudantes concluintes participantes do Enade com dados recuperados da base de dados do Enem.

5. NOTA DO INDICADOR DA DIFERENÇA ENTRE OS DESEMPENHOS OBSERVADO E ESPERADO ($NIDD_c$)

Um dos aspectos importantes na avaliação da qualidade de um curso de graduação está na mensuração de sua efetiva contribuição para o desenvolvimento de competências, habilidades e conhecimento dos estudantes, o que tem sido chamado de *valor agregado* pelo processo formativo oferecido pelo curso de graduação.

Tendo-se em vista que o desempenho dos estudantes concluintes no Enade não pode ser explicado exclusivamente pela qualidade das condições de oferta dos processos formativos, torna-se importante destacar outro fator interveniente: o perfil dos estudantes concluintes ao ingressarem na graduação (no curso referente ao Enade realizado).

Nesse sentido, o Indicador da Diferença entre os Desempenhos Observado e Esperado busca aferir aquilo que diz respeito especificamente ao valor agregado pelo curso ao desenvolvimento dos estudantes concluintes, considerando seus desempenhos no Enade e suas características de desenvolvimento ao ingressar no curso de graduação avaliado.

Conceitualmente, os fatores que determinam o desempenho dos concluintes de cursos de graduação podem estar relacionados a:

- a) características de desenvolvimento do estudante concluinte ao ingressar na Educação Superior;
- b) qualidade das condições do processo formativo oferecido pelos cursos; e
- c) outros elementos que afetam o desempenho do estudante concluinte, captados por um termo de erro.

Assim sendo, o desempenho de cada estudante concluinte no Enade poderia ser decomposto em função dos referidos três aspectos, como mostra a equação 1.

$$C = I + Q + \varepsilon \quad (1)$$

Onde:

C é o desempenho observado do estudante concluinte;

I é a parte do desempenho do estudante concluinte decorrente de suas características quando ingressante no curso;

Q é a parte do desempenho do estudante concluinte decorrente da qualidade das condições de oferta do processo formativo do curso; e

ε é o termo de erro, com a hipótese usual de que $E[\varepsilon|I, Q] = 0$.

Partindo-se dos fatores que determinam o desempenho dos concluintes, o IDD poderia ser expresso pela equação 2:

$$IDD = C - \hat{I} \quad (2)$$

Onde:

IDD é a estimativa da parte do desempenho do estudante concluinte decorrente da qualidade das condições de oferta do processo formativo do curso;

C é o desempenho observado do estudante concluinte; e

\hat{I} é a estimativa da parte do desempenho do estudante concluinte decorrente de suas características quando ingressante no curso.

Para a estimativa do termo $\hat{\lambda}$ é necessária uma medida relativa às características do estudante concluinte quando de seu ingresso no curso de graduação. Os resultados do ENEM, utilizados para a obtenção dessa medida, são recuperados a partir do número de CPF do estudante, informação obrigatória para a inscrição no exame desde 2009.

A medida de desempenho obtida a partir do Enem se configura como uma *proxy* das condições de desenvolvimento do estudante quando de seu ingresso na graduação, como uma medida de boa qualidade para a estimação do IDD, tendo em vista que ela é referente ao próprio estudante concluinte. Aqui são consideradas as notas obtidas em escala TRI (Teoria de Resposta ao Item) referentes às quatro áreas avaliadas pelo Enem: ciências da natureza, ciências humanas, linguagens e códigos e matemática e suas tecnologias.

Nesse processo foi utilizado o modelo de regressão linear multinível ou hierárquico, do tipo paramétrico para dados agregados, com o uso de parâmetros de efeitos fixos e efeitos aleatórios. Esse modelo descreve a relação entre uma variável dependente contínua (Desempenho dos estudantes concluintes no Enade) e variáveis independentes ou explicativas (Desempenho dos estudantes nas quatro áreas no Enem), fazendo uso de parâmetros de *efeitos fixos* associados a uma ou mais covariáveis e de *efeitos aleatórios* associados a um ou mais fatores aleatórios. As regressões foram estimadas com base no modelo da máxima verossimilhança restrita.

Os testes realizados para a definição da metodologia de cálculo do IDD apontam para a utilização das 4 (quatro) notas do estudante no Enem como variáveis explicativas do modelo, com efeito fixo em seus parâmetros; e com efeito aleatório no intercepto, variando para cada curso de graduação.

Diante disso, buscou-se localizar na base de dados do Enem as notas dos estudantes concluintes com valor atribuído às 4 (quatro) provas do exame, limitando-se a busca aos 3 (três) anos anteriores ao ingresso no curso avaliado ou no referido ano.

Tendo sido localizadas mais de uma participação no Enem para parte dos estudantes concluintes, foram definidos os seguintes critérios de seleção dos dados a serem utilizados no cálculo do IDD, sequenciados da mesma forma como são apresentados:

- I - Dentre os dados do estudante referentes aos 3 (três) anos anteriores ao seu ingresso no curso de graduação avaliado no Enade, foram selecionados aqueles referentes ao ano mais próximo ao seu ingresso no curso;
- II - Não sendo localizados dados de participação do estudante no Enem nos três anos anteriores ao seu ingresso no curso de graduação em questão, foram selecionados os dados referentes ao ano do ingresso, desde que esse não seja o mesmo ano de conclusão do curso (2016).

Selecionados os dados referentes aos estudantes elegíveis para o cálculo do IDD, buscou-se identificar os cursos de graduação que atendessem às seguintes condições para ter esse indicador calculado:

- a) Ter o mínimo de 2 (dois) estudantes concluintes participantes do Enade com dados recuperados da base de dados do Enem.
- b) Ter atingido 20% (vinte por cento) do total de estudantes concluintes participantes do Enade com dados recuperados da base de dados do Enem.

No cálculo do IDD, o modelo de regressão multinível foi especificado em dois níveis: (1) o estudante, identificado pelo subscrito i ; e (2) o curso, identificado pelo subscrito c . Para todas as etapas do processo de cálculo do IDD, as regressões e as estimativas de seus parâmetros ocorreram por área de avaliação do Enade.

O primeiro passo no processo de cálculo do IDD foi estimar os parâmetros do modelo de regressão utilizado para cada área de avaliação do Enade, conforme especificado nas equações 3 e 4, que caracterizam o modelo de regressão multinível para a variável de medida de desempenho obtida a partir

do Enade. Para estimar esses parâmetros, são utilizados apenas os dados dos estudantes concluintes dos cursos que atenderam às condições de cálculo do IDD.

A equação 7 expressa o primeiro nível da regressão, relativa ao estudante, onde foram considerados o intercepto e as medidas de desempenho do estudante no Enem.

$$C_{ic} = \beta_{0c} + \beta_1 \cdot CN_{ic} + \beta_2 \cdot CH_{ic} + \beta_3 \cdot LC_{ic} + \beta_4 \cdot MT_{ic} + \lambda_{ic} \quad (3)$$

Onde:

C_{ic} é medida de desempenho do estudante concluinte i no Enade, ponderada pelas notas no componente específico (75%) e na formação geral (25%), do curso c ;

CN_{ic} é a medida de desempenho do estudante concluinte i do curso c na prova do Enem de ciências naturais e suas tecnologias;

CH_{ic} é a medida de desempenho do estudante concluinte i do curso c na prova do Enem de ciências humanas e suas tecnologias;

LC_{ic} é a medida de desempenho do estudante concluinte i do curso c na prova do Enem de linguagens e códigos e suas tecnologias;

MT_{ic} é a medida de desempenho do estudante concluinte i do curso c na prova do Enem de matemática e suas tecnologias; e

λ_{ic} é o efeito aleatório associado ao estudante concluinte i do curso c .

Já a equação 4 expressa o segundo nível da regressão, relativo ao curso *de graduação*.

$$\beta_{0c} = \beta_{00} + u_{0c} \quad (4)$$

Onde:

β_{00} representa a média ou o valor do intercepto geral, que é constante entre os cursos; e

u_{0c} é o efeito aleatório associado ao curso de graduação c .

O modelo de regressão apresentado nas equações 3 e 4 é estimado duas vezes. Na primeira, estima-se a regressão, computa-se o resíduo e, então, calcula-se o resíduo padronizado. A partir do resíduo padronizado são identificados os estudantes com resíduos considerados discrepantes (*outliers*), ou seja, aqueles que apresentaram o resíduo padronizado com valor absoluto maior que 3. Na segunda estimativa, desconsiderando-se estudantes *outliers*, são obtidos os parâmetros definitivos que serão utilizados no cálculo de \hat{I}_{ic} , necessário ao cálculo do IDD.

Obtidos os parâmetros da regressão, calcula-se \hat{I}_{ic} para cada estudante concluinte, como mostra a equação 5.

$$\hat{I}_{ic} = \beta_{0c} + \beta_1 \cdot CN_{ic} + \beta_2 \cdot CH_{ic} + \beta_3 \cdot LC_{ic} + \beta_4 \cdot MT_{ic} \quad (5)$$

Onde:

\hat{I}_{ic} é estimativa da parte do desempenho do estudante concluinte i do curso de graduação c no Enade, decorrente de suas características quando ingressante no curso;

CN_{ic} é a medida de desempenho do estudante concluinte i do curso de graduação c na prova do Enem de ciências naturais e suas tecnologias;

CH_{ic} é a medida de desempenho do estudante concluinte i do curso de graduação c na prova do Enem de ciências humanas e suas tecnologias;

LC_{ic} é a medida de desempenho do estudante concluinte i do curso de graduação c na prova do Enem de linguagens e códigos e suas tecnologias; e

MT_{ic} é a medida de desempenho do estudante concluinte i do curso de graduação c na prova do Enem de matemática e suas tecnologias.

O IDD bruto foi calculado para cada estudante i do curso de graduação c como exposto na equação 6.

$$IDD_{ic} = C_{ic} - \hat{I}_{ic} \quad (6)$$

Onde:

IDD_{ic} é o IDD do estudante i do curso de graduação c ;

C_{ic} é medida de desempenho do estudante concluinte i no Enade, ponderada pelas notas no componente específico (75%) e na formação geral (25%), do curso de graduação c ; e

\hat{I}_{ic} é estimativa da parte do desempenho do estudante concluinte i do curso de graduação c no Enade decorrente de suas características quando ingressante no curso.

Calculado o IDD_{ic} de cada estudante concluinte, é necessário fazer o cálculo do IDD_c dos cursos de graduação que atenderam às condições indicadas anteriormente. Para tanto, calculou-se a média dos IDD_{ic} conforme indica a equação seguinte.

$$IDD_c = \frac{\sum_{c=1}^n IDD_{ic}}{n} \quad (7)$$

Onde:

IDD_c é a média dos dos estudantes concluintes do curso de graduação c ;

IDD_{ic} é o IDD do estudante i do curso de graduação c ; e

n é a quantidade de estudantes concluintes com IDD_{ic} calculado para o curso de graduação c .

6. PADRONIZAÇÃO E REESCALONAMENTO

Uma vez calculado o IDD_c bruto, procedeu-se à padronização e ao reescalamento descritos a seguir para se obter a Nota Padronizada do IDD ($NIDD_c$), que assume valores de 0 (zero) a 5 (cinco), na forma de variável contínua.

O processo de padronização e reescalamento passa por duas etapas:

- a) cálculo do afastamento padronizado de cada curso de graduação, fazendo-se uso das médias e dos desvios-padrão calculados por área de avaliação, como mostram as equações 8, 9 e 10;
- b) transformação dos afastamentos padronizados em notas padronizadas que também podem variar de 0 (zero) a 5 (cinco), como especifica a equação 11.

Para o cálculo das médias, utilizou-se a seguinte equação:

$$\overline{IDD}_{ck} = \frac{\sum_{c=1}^T IDD_{ck}}{T} \quad (8)$$

Onde:

\overline{IDD}_{ck} é a média do IDD_c da área de avaliação k ;

IDD_{ck} é a nota bruta do IDD_c do curso de graduação c da área de avaliação k ; e

T é o número de cursos da área k .

Para o cálculo dos desvios-padrão, adotou-se:

$$S_{IDD_k} = \sqrt{\frac{\sum_{c=1}^T (IDD_{ck} - \overline{IDD}_{ck})^2}{T-1}} \quad (9)$$

Onde:

S_{IDD_k} é o desvio-padrão do IDD_c da área k ;

IDD_{ck} é a nota bruta do IDD_c do curso de graduação c da área de avaliação k ;

\overline{IDD}_{ck} é a média do IDD_c da área de avaliação k ; e

T é o número de cursos da área de avaliação k .

Para o cálculo dos afastamentos padronizados, fez-se uso da seguinte equação:

$$Z_{IDD_c} = \frac{IDD_{ck} - \overline{IDD}_k}{S_{IDD_k}} \quad (10)$$

Onde:

Z_{IDD_c} é o afastamento padronizado do IDD_c do curso de graduação c ;

IDD_{ck} é a nota bruta do IDD_c do curso de graduação c da área de avaliação k ;

\overline{IDD}_{ck} é a média do IDD_c da área de avaliação k ; e

S_{IDD_k} é o desvio-padrão do IDD_c da área de avaliação k .

Para que todos os cursos de graduação tenham suas notas de IDD numa escala de 0 (zero) a 5 (cinco), efetua-se a interpolação linear, expressa na equação 11, obtendo-se, assim, respectivamente, as Notas Padronizadas de IDD de cada curso de graduação c . Os cursos de graduação com afastamento padronizado menor que -3 e maior que +3 recebem nota padronizada igual a 0 (zero) e 5 (cinco), respectivamente, que não são utilizadas como mínimo ou máximo no cálculo do IDD, por serem consideradas discrepantes (*outliers*) em relação aos demais.

$$NIDD_c = 5 \cdot \left(\frac{Z_{IDD_c} - Z_{IDD_k \min}}{Z_{IDD_k \max} - Z_{IDD_k \min}} \right) \quad (11)$$

Onde:

$NIDD_c$ é a nota padronizada do IDD_c do curso de graduação c ;

Z_{IDD_c} é o afastamento padronizado do IDD_c do curso de graduação c ;

$Z_{IDD_k \min}$ é o afastamento padronizado mínimo do IDD_c da área de avaliação k ; e

$Z_{IDD_k \max}$ é o afastamento padronizado máximo do IDD_c da área de avaliação k .

7. DISTRIBUIÇÃO POR FAIXAS

O IDD é uma variável discreta que assume valores de 1 a 5, resultante da conversão da *Nata Padronizada do IDD* do curso de graduação c ($NIDD_c$), realizada conforme definido na Tabela 1.

Tabela 1 – Parâmetros de conversão do $NIDD_c$ em IDD por faixas

IDD (Faixa)	$NIDD_c$ (Valor Contínuo)
1	$0 \leq NIDD_c < 0,945$
2	$0,945 \leq NIDD_c < 1,945$
3	$1,945 \leq NIDD_c < 2,945$
4	$2,945 \leq NIDD_c < 3,945$
5	$3,945 \leq NIDD_c \leq 5$

Fonte: Inep/Daes

Equipe Técnica

Mariangela Abrão Robson Quintilio

Ulysses Tavares Teixeira

De acordo,

Suzi Mesquita Vargas

Coordenadora de Estatísticas e Indicadores da Educação Superior

Renato Augusto dos Santos

Diretor de Avaliação da Educação Superior - substituto

8.

ANEXO I - ÁREAS DE AVALIAÇÃO DO ENADE EM 2016

CÓDIGO	ÁREA DE ENQUADRAMENTO
17	AGRONOMIA
55	BIOMEDICINA
35	EDUCAÇÃO FÍSICA
23	ENFERMAGEM
19	FARMÁCIA
36	FISIOTERAPIA
27	FONOAUDIOLOGIA
12	MEDICINA
5	MEDICINA VETERINÁRIA
28	NUTRIÇÃO
6	ODONTOLOGIA
38	SERVIÇO SOCIAL
51	ZOOTECNIA
90	TECNOLOGIA EM AGRONEGÓCIO
95	TECNOLOGIA EM ESTÉTICA E COSMÉTICA
92	TECNOLOGIA EM GESTÃO AMBIENTAL
91	TECNOLOGIA EM GESTÃO HOSPITALAR
69	TECNOLOGIA EM RADIOLOGIA



Documento assinado eletronicamente por **Ulysses Tavares Teixeira, Servidor Público Federal**, em 09/06/2017, às 10:32, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Robson Quintílio, Servidor Público Federal**, em 09/06/2017, às 10:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mariângela Abrão, Servidor Público Federal**, em 09/06/2017, às 17:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Suzi Mesquita Vargas, Coordenador(a)**, em 09/06/2017, às 17:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Renato Augusto dos Santos, Diretor(a), Substituto(a)**, em 09/06/2017, às 18:08, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.inep.gov.br/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0055924** e o código CRC **D8A05FA9**.

