

VERIFICAÇÃO DA ADEQUABILIDADE DO ÍNDICE DE DIFERENÇA DE DESEMPENHO (IDD) DO INEP**Musbah Koleilat Câmara**CEFET/RJ Campus Nova Iguaçu
Estr. de Adrianópolis, 1.317 - Nova Iguaçu – RJ
koleilat@cefet-rj.br**Ana Luiza Lima de Souza**CEFET/RJ Campus Nova Iguaçu
Estr. de Adrianópolis, 1.317 - Nova Iguaçu – RJ
ana.souza@cefet-rj.br**Marcos Pereira Estellita Lins**UNIRIO - Centro de Ciências Exatas e Tecnológicas
Avenida Pasteur, 458 - Urca - Rio de Janeiro – RJ
marcos.lins@uniriotec.br**RESUMO**

As medições da aprendizagem dos estudantes desempenham um papel essencial na determinação da qualidade do ensino superior no Brasil e no mundo, o que evidencia a relevância de se avaliar a validade do Indicador de Diferença de Desempenho (IDD) do INEP, ou seja, verificar se esse indicador mede o valor educacional agregado pelos cursos aos seus alunos. Experimentos foram realizados no modelo de regressão linear multinível do IDD, com o intuito de verificar sua adequabilidade e consistência em avaliar o ganho de aprendizagem. Foi detectada inconsistência, nas estimativas para ranqueamento de instituições.

Palavra-chave: Modelo Linear Multinível; IDD do Inep; Ganho De Aprendizagem.

ABSTRACT

Measurements of student learning play a essential role in determining the quality of higher education in Brazil and around the world, which evidences the relevance of assess the validity of the Performance Difference Indicator (IDD), ie, checking whether this indicator measures the educational value added by the courses to their students. Experiments were performed on the IDD's multilevel linear regression model, in order to verify their appropriateness and consistency to assess learning gain. Inconsistency was detected, in estimates to rank institutions.

KEYWORDS: Multilevel Linear Models; INEP's IDD; Learning Gain.

Como Citar:

KOLEILAT, Musbah; SOUZA, Ana Luiza L.; LINS, Marcos P. E. Verificação da Adequabilidade do Índice de Diferença de Desempenho (IDD) do INEP. *In: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA*, 19., 2019, Rio de Janeiro, RJ. **Anais** [...]. Rio de Janeiro: Centro de Análises de Sistemas Navais, 2019.

1 INTRODUÇÃO

As medições da aprendizagem dos estudantes estão desempenhando um papel cada vez mais significativo na determinação da qualidade do ensino superior[CITATION COA09 \l 1046]. Yorke e Knight [CITATION Yor06 \n \t \l 1046] analisam o papel das instituições de ensino superior em termos de contribuição para os ganhos de aprendizado dos alunos e a importância desse aspecto para a empregabilidade e o sucesso econômico de diferentes países. As funções econômicas e sociais do ensino superior expandiram-se e diversificaram-se rapidamente nas últimas décadas e os estudantes devem atuar de forma produtiva na economia global do conhecimento. Isso enfatiza a necessidade crescente do governo, gestores universitários, empresas, estudantes e o público em geral de acesso a informações mais focadas no valor que o ensino superior agrega ao aprendizado[CITATION COA09 \l 1046].

O ganho de aprendizagem busca mensurar o desempenho do aluno, ou seja, o conhecimento adquirido entre duas etapas de sua vida acadêmica. Para os propósitos deste estudo, o conceito de ganho de aprendizagem é análogo ao conceito de valor agregado ou simplesmente aprendizado, como observado por[CITATION MCG15 \l 1046]. Douglass, Thomson e Zhao [CITATION Dou12 \n \t \l 1046] relatam que, em geral, a mensuração dos ganhos de aprendizagem é vista por muitos como um método para avaliar o valor agregado aos estudantes por suas respectivas universidades. Portanto, para fornecer uma comparação justa do desempenho de diferentes instituições educacionais, é necessário medir o valor agregado ao aluno, o que justificaria o investimento do governo na melhoria da qualidade dos serviços e na ampliação da capacidade das instituições de ensino superior [CITATION Rod05 \l 1046].

Portanto, se torna necessário, o desenvolvimento de métodos robustos para medir o ganho de aprendizagem, pois é importante medir o desempenho das instituições de ensino [CITATION Bog13 \l 1046]. A eficácia desses indicadores pode ser aprimorada pelo uso de métodos adequados e criteriosos de julgamento para mensurar os ganhos e comparar as instituições [CITATION Boy12 \l 1046]. Cunha e Miller [CITATION Cun14 \n \t \l 1046] destacam a importância de desenvolver medidas de valor agregado aplicadas ao ensino superior, como apoio à tomada de decisões sobre políticas educacionais, para fins de financiamento e incentivos. À luz do crescente interesse em usar medidas de valor agregado no ensino superior, esses indicadores oferecem lições impares para os formuladores de políticas.

Estudos que analisam o nível de conhecimento ou perfil dos alunos podem utilizar testes padronizados aplicados em diferentes momentos da trajetória acadêmica e, a partir dos resultados obtidos, buscam inferir sobre algumas características como, por exemplo, a qualidade do ensino das instituições, desempenho dos professores, sucesso do aluno com base em seu perfil social, etc. Modelos matemáticos como Análise Envoltória de Dados (DEA) e regressão linear multivariada e/ou multinível são usados para medir o valor agregado de estudantes de nível superior usando dados obtidos através de testes padronizados.

2 CONTEXTUALIZAÇÃO

Mesmo que se compare com os padrões latino-americanos, constata-se que o Brasil entrou tarde na esfera da educação. As primeiras universidades datam da década de 1930, e até a década de 1950 nenhuma tentativa séria foi feita para criar um sistema nacional de educação pública universal [CITATION Sch13 \l 1046]. No início dos anos 1950, foram criados o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Essas agências governamentais tornaram-se vitais, para o desenvolvimento de estudos nos níveis de pós-graduação, pesquisa científica e apoio financeiro à ciência brasileira, e ainda contribuíram com o intercâmbio de pesquisadores com instituições do exterior [CITATION Cou12 \m Pic16 \l 1046]. O ensino superior no Brasil, até a década de 1960, foi marcado por um número limitado de vagas em universidades e grande número de candidatos, o que se tornou um gargalo muito estreito para a maioria dos estudantes. Nas décadas de 1970 e 1980, ocorreu a primeira expansão do ensino superior, liderada pelo setor privado, para aumentar o número de instituições de ensino superior e a consequente oferta de cursos e vagas [CITATION Cam18 \l 1046].

Em meados da década de 1990, o governo brasileiro transferiu funções e serviços públicos para a iniciativa privada, um movimento conhecido como privatização, e entre os setores afetados estava o ensino superior. Essa estratégia tinha o princípio de mudar o papel do Estado, de controlador para regulador, usando sistemas regulatórios [CITATION Pol09 \l 1046]. Esses sistemas regulatórios exigem instrumentos técnicos e dados comparáveis para avaliar o desempenho das instituições, assim os exames gerais ganharam importância. De acordo com Rosa Becker e Costa [CITATION Ros13 \n \t \l 1046], o atual sistema de avaliação inclui uma avaliação do desempenho dos alunos, nos moldes do extinto Provão, e foi batizado como Exame Nacional de Desempenho dos Estudantes (ENADE). O ENADE é um exame obrigatório para o ensino superior, sendo aplicado em alunos que estão próximos da conclusão do curso, e foi projetado para medir o conhecimento geral e específico da área de estudo desses alunos para os diferentes cursos, como por exemplo, engenharia, medicina, direito, etc [CITATION Mel16 \l 1046]. Ainda faz parte desse sistema de avaliação, uma pesquisa de opinião de estudantes, aplicada em conjunto com o ENADE; e informação estatística sobre a proporção de professores com pós-graduação, regime de trabalho, entre outros [CITATION Mag15 \l 1046].

Uma comparação feita pelo Indicador de Diferença de Desempenho (IDD), mais um componente desse sistema, que leva em consideração o desempenho do aluno ao ingressar (medido pelo ENEM), e completar o curso (dados do ENADE), é usada para estimar o valor educacional adicionado pelos cursos aos seus alunos [CITATION Mag15 \l 1046]. Atualmente, o teste de admissão utilizado pela maioria das universidades públicas e privadas no Brasil é o Exame Nacional do Ensino Médio (ENEM), criado em 1998 pelo governo federal, inicialmente com o objetivo de avaliar a qualidade do ensino médio no país, e que tem participação voluntária de estudantes [CITATION Her15 \l 1046]. Todas essas informações são quantificadas e usadas para produzir uma classificação combinada chamada CPC (Conceito Preliminar de Curso), um indicador com escala de cinco pontos que agrega, utilizando pesos, todas essas métricas para avaliar cursos e, em seguida, essas pontuações são combinadas novamente com informações sobre o avaliação da pós-graduação para produzir uma classificação geral para a instituição de ensino superior como um todo [CITATION Mag15 \l 1046]. Esses rankings são publicados e utilizados pelo Ministério da Educação (MEC) para fins regulatórios e apoio às políticas educacionais, além de serem consultados por alunos, pais, acadêmicos e público em geral para orientar suas decisões sobre cursos e instituições [CITATION Ros13 \l 1046 \m Sat17].

O Indicador de Diferença de Desempenho (IDD) é o instrumento utilizado como índice de valor agregado na avaliação do ensino superior brasileiro. Surgiu em 2005 como resposta a críticas que sugeriam que o sistema de avaliação, tal como foi construído, favorecia as instituições públicas, devido à maior exigência na admissão das mesmas [CITATION Bit09 \l 1046]. O IDD, se entendido como medida de valor agregado, é considerada como o principal avanço, inovação e diferencial, quando comparado ao sistema de avaliação anterior baseado no ENC [CITATION Bit08 \l 1046 \m Gar14]{Amadeu

Garcia, 2014 #17;Bittencourt, 2008 #18}. Se considerado isoladamente, o IDD parece ser mais justo em teoria do que o ENADE, a fim de avaliar as instituições de ensino superior (IES), uma vez que leva em conta o contexto anterior a realização do curso. Na prática, o IDD, que é um dos elementos que compõem o CPC, tornou-se uma espécie de penalidade para as universidades públicas, uma vez nivelada a condição de entrada, levando a um maior equilíbrio em termos de competição entre instituições públicas e privadas [CITATION Bit08 \l 1046].

Dada a importância deste indicador de valor agregado em termos de prestígio institucional tanto para o governo como para a sociedade em geral, o objetivo deste estudo é avaliar a validade do IDD, ou seja, se esse indicador mede adequadamente o que está sendo proposto. Se a métrica for apropriada, deve permitir que qualquer instituição possa chegar ao topo do ranking de avaliação. Além disso, a pesquisa pretende avaliar se a metodologia é confiável e se realmente revela aspectos importantes para melhorias nas IES.

Este artigo está estruturado em seis tópicos, a saber: a introdução que apresenta o objetivo e a ideia central do trabalho, a contextualização que embasa o leitor na realidade do problema e a revisão sobre ganhos de aprendizado e avaliação de valor agregado no ensino superior, são os três primeiros. A metodologia adotada, as técnicas de coleta e análise dos dados utilizados na pesquisa empírica são descritas na quarta seção. Na quinta seção, os resultados da pesquisa são descritos. E, por fim, associadas às conclusões e discussões, registram-se as limitações do estudo e as recomendações para estudos futuros.

3 REVISÃO DO GANHO DE APRENDIZAGEM E VALOR AGREGADO

O uso e os efeitos de uma variedade de métodos de avaliação no ensino superior foram investigados por Pereira, Flores e Niklasson [CITATION Per16 \n \t \l 1046], particularmente aqueles que apontam para os chamados métodos alternativos de avaliação. De acordo com os requisitos do Processo de Bolonha (um processo de reforma intergovernamental europeia com uma série de acordos relacionados com o ensino superior), são necessários novos métodos e estratégias educacionais para melhorar o ensino centrado no aluno. Fernandes, Flores e Lima [CITATION Fer12 \n \t \l 1046] exploram as percepções dos estudantes sobre os procedimentos e processos de avaliação e apresentam algumas implicações para melhorar as práticas de avaliação.

O ganho de aprendizado procura medir o desempenho do aluno, ou aprendizado adquirido, entre dois estágios de sua vida acadêmica. Este conceito pode estar relacionado ao desenvolvimento pessoal, ao conteúdo do conhecimento e aos avanços nas ferramentas e habilidades adquiridas. Entre seus principais benefícios, avaliar o ganho de aprendizagem pode melhorar a qualidade do ensino e facilitar o aprendizado; também promovendo valores como responsabilidade, transparência e cidadania [CITATION MCG15 \l 1046]; e permitindo a comparabilidade do ensino superior. Banta [CITATION Ban12 \n \t \l 1046] em seu estudo, por exemplo, enfoca o desenvolvimento de métodos de comparação interestaduais de aprendizagem dos alunos nos Estados Unidos que permitem a avaliação da qualidade institucional.

Os testes padronizados são usados para obter informações sobre os cursos de graduação e podem ser aplicados aos alunos no início ou no final de seus cursos, com o objetivo de comparar e avaliar alunos e instituições. Testes padronizados podem ser usados para medir o conhecimento dos alunos em termos genéricos ou especializados [CITATION MCG15 \l 1046]. Com relação aos testes de disciplina específicos, podemos mencionar os Testes de Progresso administrados em algumas escolas médicas na Inglaterra, bem como o projeto AHELO da OCDE, que também analisa a viabilidade de desenvolver uma medida comparável de resultados de aprendizagem [CITATION COA09 \m Tre12 \l 1046]. Testes

de habilidades genéricas tendem a medir habilidades gerais, como pensamento crítico, que não são específicos da disciplina. Alguns modelos são o Collegiate Learning Assessment (CLA), que se baseia em questões abertas, e o Measure for Academic Proficiency and Progress (MAPP), que é o teste mais tradicional baseado em questões de múltipla escolha. Ambos os exemplos foram originados e são aplicados nos Estados Unidos [CITATION Cun14 \l 1046].

Vários estudos, como o desenvolvido por Roohr, Liu, Liu [CITATION Roo17 \n \t \l 1046] analisam os ganhos de aprendizagem de universitários através da estimação do cálculo das diferenças de pontuação em aplicações de entrada e saída em testes padronizados. Nesse sentido, o estudo de Stowell [CITATION Sto04 \n \t \l 1046] examina a confiabilidade de um instrumento para medir a capacidade de aprendizagem de estudantes na universidade por meio de testes educacionais nos Estados Unidos. Uma pesquisa realizada na Universidade da Califórnia discutiu os resultados da aprendizagem, a validade dos testes padronizados e como tirar proveito das pesquisas com os alunos. Também abordou a questão do conflito entre os desejos dos governos e as necessidades das universidades. Entre as desvantagens dos testes padronizados estão a gama de comparabilidade entre disciplinas e o nível de precisão para a medição do ganho de aprendizagem. Além disso, os testes padronizados precisam ser claramente especificados para fornecer uma medida significativa do ganho de aprendizagem, de modo a não criar uma percepção negativa do setor educacional, pois são usados para classificar alunos e instituições e, por outro lado, precisam fornecer informações úteis sobre o contexto educacional, recursos, programas e processos pedagógicos. Também deve haver cautela no significado dos testes usando amostras representativas de estudantes [CITATION Tre12 \l 1046].

A avaliação do valor agregado está se tornando cada vez mais popular na avaliação do impacto da universidade em estudantes do ensino superior. Argumenta-se que métodos apropriados devem ser empregados para comparações de instituições e podem ser mais ou menos apropriados dependendo das várias condições sistêmicas, estabelecendo padrões para julgar ganhos observados e fazer inferências causais [CITATION BOR08 \l 1046 \m Boy12]. O conceito de valor agregado mede a diferença entre a pontuação real de um aluno em um segundo momento e sua pontuação prevista em um primeiro momento (que pode ser baseada no ponto de saída médio ou mediana alcançado por outros alunos com o mesmo ponto de partida ou similar). O valor adicionado tem sido utilizado mais recentemente em estudos de educação e sua mensuração é realizada por meio de modelos estatísticos de diferentes complexidades. Existem diferentes entendimentos sobre o ganho de aprendizado. Para os propósitos deste estudo, como já mencionado, o conceito de ganho de aprendizagem é análogo ao conceito de valor agregado de MacGrath, Guerin, et al. [CITATION MCG15 \n \t \l 1046].

Goldenstein e Thomas [CITATION Gil96 \n \t \l 1046] definiram o valor agregado como o aumento quantitativo em cada instituição de ensino, através de suas práticas e políticas, do desempenho acadêmico dos estudantes. O potencial para medir o valor agregado é uma ferramenta de diagnóstico para melhoria institucional, fornecendo à sociedade uma ideia da qualidade da instituição, bem como oferecendo informações úteis e melhorias no sistema de avaliação nacional das IESs [CITATION Soa01 \l 1046]. Grayson [CITATION Gra99 \n \t \l 1046] argumenta que o melhor método para medir o valor agregado na educação é através de análises longitudinais usando medidas subjetivas e objetivas de habilidades com grupos de controle apropriados que podem comparar as habilidades de entrada e saída dos alunos, o estudo centra-se na Universidade de York (Ontário). Nesse sentido, Pampaka, Williams e Hutcheson [CITATION Pam12 \n \t \l 1046] realizaram uma pesquisa longitudinal com alunos que ingressaram em diferentes programas em cinco universidades e mediram a relação da pedagogia com os resultados de

aprendizagem de estudantes de matemática através das percepções dos alunos. Pike [CITATION Pik16 \n \t \l 1046] aborda o uso de modelos de valor agregado na avaliação de ensino superior nos EUA e levanta questões sobre a aprendizagem dos alunos, a necessidade de evidência rigorosa de valor agregado para a educação e o monitoramento regular da qualidade técnica dos modelos de valor agregado.

Com base nesse cenário, é importante desenvolver um método robusto para medir o valor agregado que deve apresentar quatro fatores essenciais: levantamentos longitudinais ou transversais; validade estatística; disponibilidade de dados; e comparabilidade, uma vez que a comparação entre o desempenho de estudantes, disciplinas, instituições e países é potencialmente válida neste tipo de estudo [CITATION MCG15 \l 1046]. Vale ressaltar ainda que, segundo Ferrante [CITATION Fer17 \n \t \l 1046], por meio de um estudo utilizando testes padronizados de 24 faculdades de engenharia italianas, confirmou-se que não levar em conta a qualidade recebida dos alunos ingressantes poderia gerar distorções significativas na avaliação da produtividade acadêmica das universidades, que impactam na alocação de recursos públicos, especialmente na forma de bolsas de estudo, e agravam a polarização dos sistemas educacionais. O mesmo resultado pode ser gerado pelo uso indevido de classificações de universidades com base em medidas de desempenho não ajustadas. Além disso, Zhang [CITATION Zha09 \n \t \l 1046], que estima a qualidade média das faculdades americanas com base no valor agregado, expõe que vários aspectos da política estadual de ensino superior podem afetar o desempenho das universidades americanas. Em uma análise crítica, Kelly-Woessner e Woessner [CITATION Kel11 \n \t \l 1046] levantam a questão de que as universidades falham em produzir um aumento significativo no valor agregado do pensamento crítico e do raciocínio complexo entre os estudantes. Isso repercute no questionamento da contribuição social dessas instituições.

Finalmente, Johnes [CITATION Joh18 \n \t \l 1046] afirma que os rankings universitários buscam medir o desempenho das instituições de ensino superior e são usados por diferentes interessados, por exemplo: estudantes que procuram universidades de renome; acadêmicos que procuram empregos universitários; gestores universitários que precisam manter sua posição para o recrutamento de estudantes; e governos que querem saber se os investimentos nas universidades estão fornecendo um sistema de ensino superior de qualidade.

4 MODELOS MULTINÍVEL, IDD E METODOLOGIA DE VALIDAÇÃO

Apresentamos nesta seção uma breve explanação sobre a utilização do modelo de regressão linear multinível, descrevemos o modelo matemático utilizado para a obtenção do Índice de Diferença de Desempenho (IDD), que é baseado no modelo multinível. E, por último, especificamos o método de validação utilizado para verificar a adequabilidade e consistência da aplicação de tal modelo para avaliar o ganho de aprendizagem.

4.1 MODELOS LINEARES MULTINÍVEL

Modelos lineares multinível são comumente usados para obter medidas de valor agregado na educação. Essas medidas foram inicialmente usadas para avaliar escolas secundárias, mas podem ser facilmente adaptadas para avaliar instituições de ensino superior. Tipicamente, essas aplicações são baseadas na pontuação de um teste de conhecimento individual em uma primeira etapa, e após um período de tempo em que se considera que o indivíduo foi submetido a um processo de aprendizagem, outra avaliação (segunda etapa) é feita, novamente usando um teste escrito e pontuações, em que se estima o quanto, em termos de conhecimento, foi agregado ao aluno. Ao final, a partir desta informação, comparações entre instituições, que podem diferir substancialmente em sua efetividade, são feitas [CITATION Rab08 \l 1046].

Um aspecto importante desse processo é a tentativa, a partir da aplicação de um primeiro teste, de estimar o que realmente se agrega em termos de conhecimento, já que os aspectos formativos, sociais e culturais que permitiram ao indivíduo atingir certo nível de desempenho no estágio anterior são sendo capturado no modelo. Esse recurso às vezes é ignorado em exames que examinam apenas o desempenho final do aluno. Como justificativa para o uso deste modelo, argumenta-se que o desempenho dos alunos que estão prestes a concluir o curso não pode ser explicado exclusivamente pela qualidade da instituição de ensino superior à qual está vinculado, sendo importante reconhecer e também considerar o nível de conhecimento que o aluno tinha antes de ingressar no ensino superior. Esta análise comparativa recebe o nome de valor agregado. O modelo utilizado pelo INEP é equivalente a um modelo geral de regressão multinível com interceptos aleatórios, tal como

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \xi_{ij}$$

onde y_{ij} é a variável dependente, x_{1ij} até x_{pij} são variáveis explicativas e ξ_{ij} é o resíduo.

Pode ser pouco realista assumir que os resultados obtidos por alunos de uma mesma instituição não estejam correlacionados, ou que os resíduos ξ_{ij} e $\xi_{i'j}$ não possuam correlação. Nesse caso, podemos superar essa limitação ao dividir o resíduo ξ_{ij} em dois componentes não correlacionados ϵ_{ij} e u_j , respectivamente resíduos de nível 1 e 2:

$$\xi_{ij} = u_j + \epsilon_{ij}$$

O resíduo de nível 1, ϵ_{ij} , é específico para cada aluno i . O resíduo de nível 2, u_j , também chamado de intercepto aleatório, é específico para cada uma das j instituições, e é constante para cada um dos seus i alunos. Substituindo o termo ξ_{ij} no modelo de regressão multinível, nós obtemos o modelo linear de interceptos aleatórios

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + (u_j + \epsilon_{ij}) = (\beta_0 + u_j) + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \epsilon_{ij}$$

Este modelo pode ser visto como um modelo de regressão com resíduos de segundo nível u_j , ou com interceptos específicos para instituições $\beta_0 + u_j$. Uma formulação também utilizada e útil na especificação e interpretação de modelos de dois ou mais níveis, é apresentada

$$\text{Nível 1: } y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \epsilon_{ij} \quad \text{Nível 2: } \beta_{0j} = \beta_0 + u_j$$

4.2 ÍNDICE DE DIFERENÇA DE DESEMPENHO (IDD)

O INEP/MEC define como modelo de valor agregado para o seu sistema de avaliação, o modelo linear multinível de interceptos aleatórios especificado na Nota Técnica DAES/INEP nº 58/2015, que utiliza as pontuações do ENEM para os alunos entrantes e a pontuação do ENADE para os alunos concluintes. A métrica designada como indicador de diferença de desempenho (IDD) faz parte de um índice mais amplo chamado de Conceito Preliminar de Curso (CPC) que, por sua vez, é a medida do INEP/MEC que pretende avaliar a qualidade dos cursos superiores. Esse ranking é de suma importância para as instituições de ensino superior no Brasil, pois regula os processos de reconhecimento e perpetuidade dos cursos [CITATION BRA06 \m BRA08 \l 1046]. Em vista dos impactos causados pelo CPC, temos que garantir que os diversos indicadores que o compõem sejam apropriados e consistentes, a fim de garantir uma avaliação justa.

Nesse sentido, descreveremos com mais detalhes a metodologia do IDD, que possui

o maior peso (35%) entre os indicadores que compõem o CPC, e que teve sua consistência verificada. De acordo com a Nota Técnica DAES/INEP nº 33/2017, que trata da metodologia de cálculo do IDD, o indicador de diferença entre o desempenho observado e o esperado mede o valor agregado pelo curso (IDD) no desenvolvimento dos alunos, considerando seu desempenho como aluno recém-chegado, com base no ENEM, e no final do curso, utilizando o escore do ENADE. O modelo matemático utilizado pelo INEP (2016) para traduzir o conceito de valor agregado bruto, IDD bruto, é representado pela seguinte equação:

$$IDD_{ij} = C_{ij} - \hat{I}_{ij}$$

onde IDD_{ij} é o valor do IDD para o indivíduo, neste caso o estudante i , matriculado na instituição j ; C_{ij} é a pontuação do ENADE, do estudante i na instituição j ; e \hat{I}_{ij} é a regressão linear multinível (tipo paramétrico, com máxima verossimilhança restrita) do desempenho estimado para ao entrar no curso superior, com base nos resultados dos testes ENEM, e que é representado nas seguintes equações:

$$\text{Nível 1: } C_{ij} = \beta_{0j} + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij} + \lambda_{ij} \quad \text{Nível 2: } \beta_{0j} = \beta_{00} + u_{0j}$$

onde C_{ij} é a pontuação do ENADE do aluno i ; CN_{ij} é a pontuação do estudante i no teste de ciências naturais do ENEM, CH_{ij} é a pontuação do estudante i no teste de ciências humanas do ENEM; LC_{ij} é a pontuação do estudante i no teste de linguagens do ENEM; MT_{ij} é a pontuação do estudante i no teste de matemática do ENEM. Além disso, λ_{ij} é o efeito aleatório associado ao estudante i do curso j . β_{00} é o intercepto geral; e u_{0j} é o intercepto aleatório associado com o curso de graduação j .

Esta regressão multinível foi realizada considerando apenas os dados dos alunos na área de conhecimento especificado (por exemplo, engenheiro mecânico, direito, medicina, arquitetura, psicologia, etc.) de seu respectivo curso, e o nível 1 foi relativo ao aluno e nível 2 a o curso em instituição. Considerando de podemos expandir a regressão de nível 1 para:

$$C_{ij} = (\beta_{00} + u_{0j}) + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij} + \lambda_{ij}$$

então, ao estimar os coeficientes da regressão, calculamos o ENADE estimado \hat{I}_{ij} como

$$\hat{I}_{ij} = \beta_{00} + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij}$$

Cabe observar que no ENADE estimado \hat{I}_{ij} , não está sendo considerado justamente o efeito aleatório associado a instituição u_{0j} . Dessa forma, ao calcular IDD_{ij} temos:

$$IDD_{ij} = C_{ij} - \hat{I}_{ij} \\ = \left[(\beta_{00} + u_{0j}) + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij} + \lambda_{ij} \right] - \left[\beta_{00} + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij} \right] \\ = u_{0j} + \lambda_{ij}$$

Ou seja, o IDD bruto IDD_{ij} é igual ao valor do intercepto aleatório u_{0j} associado a instituição de ensino superior j para cada um dos estudantes i .

Esse modelo é estimado duas vezes, pois na primeira estimativa calcula-se o resíduo padronizado visando eliminar os outliers ao eliminar registros que possuam resíduo maior em valor absoluto do que 3. Na segunda estimativa, sem os outliers, são obtidos os

parâmetros que serão utilizados para calcular o \hat{I}_{ij} .

$$\hat{I}_{ij} = \beta_0 + \beta_1 \cdot CN_{ij} + \beta_2 \cdot CH_{ij} + \beta_3 \cdot LC_{ij} + \beta_4 \cdot MT_{ij} \quad (1)$$

O \hat{I}_{ij} é a característica do desempenho do estudante i no ENADE decorrente de suas características quando ingressou no curso j . Calculamos o IDD do curso a partir da média dos IDD individuais dos alunos da respectiva instituição, conforme a equação (2).

$$IDD_j = \frac{\sum_{i=1}^n IDD_{ij}}{n} \quad (2)$$

onde IDD_j é a média do IDD dos estudantes em seus respectivos cursos.

Portanto, após os procedimentos apresentados obtemos um valor numérico positivo ou negativo, que será padronizado e colocado em uma escala contínua de valores entre 0 e 5.

4.3 MÉTODO DE VALIDAÇÃO

O experimento de validação visa verificar eventual inconsistência quanto ao método empregado no cálculo do IDD. Analisando a priori os resultados oficiais do IDD, questionou-se o modelo¹ quanto a possibilidade de instituições com processos de admissão extremamente concorridos (instituições com nota de corte no ENEM muito elevadas) serem prejudicadas por terem pouca margem de melhoria no indicador ENADE, e se tal característica permitiria as mesmas de competir em pé de igualdade com as outras instituições, com menores notas de corte. Em outras palavras, se deseja verificar a possibilidade ou impossibilidade matemática da obtenção, por instituições muito concorridas, de um valor de IDD contínuo, que as permita competir em pé de igualdade com as outras instituições. Caso seja possível obtenção de tais valores de IDD que possibilite que a instituição assuma eventualmente a primeira posição, caso seja melhor que as demais no indicador em estudo, então o modelo é válido. Caso seja comprovada tal impossibilidade matemática, então o modelo apresenta inconsistência, e deve ser reformulado.

Foi feito inicialmente procedimento de reprodução dos resultados de IDD divulgados pelo INEP, para os cursos de engenharia de produção. Para levar a diante esta tarefa, utilizamos o software estatístico StataSE 13 e o software de manipulação de planilhas Microsoft Excel®, além da base de dados do IDD de 2014. As etapas realizadas para a obtenção de resultados equivalentes aos divulgados pelo INEP, são de forma resumida apresentados a seguir:

Etapla 1: Obter e ajustar arquivo de dados de entrada.

Etapla 2: Realizar primeira estimativa dos valores do ENADE com base no ENEM.

Etapla 3: Eliminar Outliers utilizando os resíduos padronizados.

Etapla 4: Realizar a segunda estimativa, de obtendo os valores finais dos parâmetros.

Etapla 5: Obtemos \hat{I}_{ij} de cada aluno.

Etapla 6: IDD bruto é calculado como exposto na equação, $IDD_{ij} = C_{ij} - \hat{I}_{ij}$.

Etapla 7: O IDD_j das IES são calculados pela média dos IDD_{ij} .

Ao comparar os resultados entre os dados gerados pelo estudo e os dados oficiais divulgados pelo INEP, temos que os dados obtidos foram exatamente iguais, divergindo

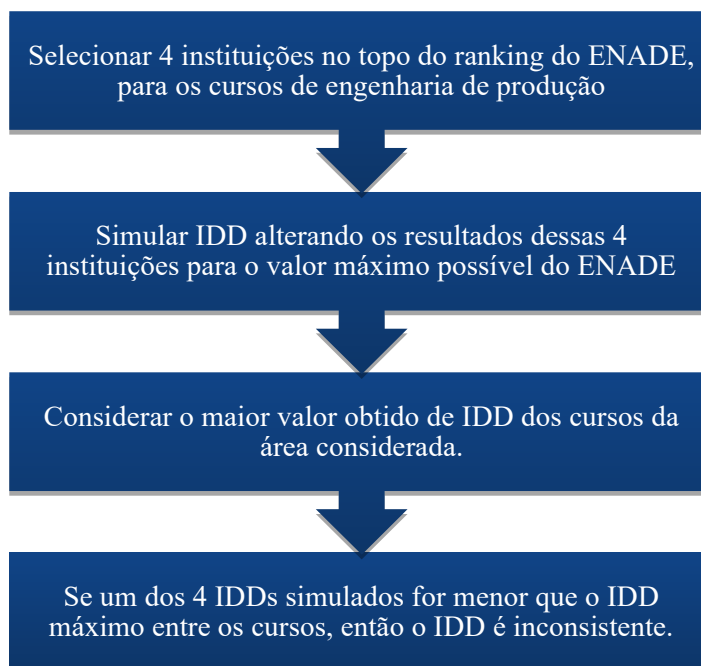
¹ Questionamento levantado por grupo de pesquisa PSIGMA do Programa de Engenharia de Produção da COPPE/UFRJ, ao analisar os indicadores de qualidade na educação superior quanto aos resultados dos indicadores de avaliação do MEC publicados em 2014.

apenas após a 5ª casa decimal. Essa pequena diferença era esperada devido ao eventual uso de diferentes pacotes de software, arredondamentos intermediários, ou mesmo atualizações ou reconsiderações feitas pelo INEP para algumas instituições antes da publicação do IDD, que pode ter provocado diferença quase insignificante entre os parâmetros da regressão. Não obstante, a obtenção de resultados com este nível de proximidade se comparados aos resultados fornecidos pelo INEP, nos deixou confortáveis para avançar com as próximas fases do experimento.

Considerando agora as questões conceituais por trás do experimento de validação propriamente dito, partimos de algumas considerações. A primeira delas é que todas as unidades de observação estão competindo entre si para obter o maior valor de IDD, caso uma determinada instituição tenha uma nota de corte do ENEM muito elevada, ainda assim, deve existir a possibilidade matemática de que a instituição possa competir com as demais. Pois caso uma instituição saiba de antemão que não conseguirá atingir o topo do ranking por melhor que ela seja nos itens de avaliação, então, nesse caso, não se justifica ou pode não ser interessante, que ela continue participando de tal competição. Afinal de contas, a instituição está sendo avaliada por esse indicador, e este indicador possui o maior peso no sistema de avaliação.

Depois, é de conhecimento público que a instituição de ensino não tem ingerência sobre a nota de corte do ENEM. A nota de corte é definida pela quantidade de vagas ofertadas (normalmente constante), e sobretudo, pela procura dos alunos em ingressar na referida instituição. Não raro em instituições públicas, para os cursos mais concorridos, as notas de corte ficam muito próximas das notas máximas. Desta forma, remetendo a PIDD (1998), temos que o desempenho no ENADE está sob o controle da instituição, sendo assim uma variável controlável. Em contrapartida, a nota de corte do ENEM foge a alçada institucional sendo, portanto, variável que não permite seu controle. Ou seja, os gestores de uma instituição não têm ingerência sob tal indicador.

Temos também que o IDD, como o próprio nome sugere, busca mensurar a diferença de desempenho entre o ENEM e o ENADE. Considerando que a nota de corte do ENEM de uma instituição seja próxima da nota máxima, então mesmo que a sua nota no ENADE (variável controlável) seja máxima, muito provavelmente (sairemos da natureza especulativa adiante) a diferença obtida será insuficiente para que a mesma possa competir com as demais instituições. Tendo em vista as questões colocadas, nos resta agora analisar, conforme a situação apresentada, se uma instituição com pontuação na nota de corte do ENEM próxima da nota máxima pode possuir uma nota de IDD inferior as outras instituições, mesmo que ela consiga atingir a nota máxima no ENADE. O experimento de validação consiste na sequência de procedimentos listados a seguir, sendo esses procedimentos aplicados em um grupo focal que consiste em algumas instituições que tiveram excelente desempenho no ENADE em 2014. Apresentamos abaixo diagrama que contém a lógica do experimento.



5 RESULTADO

Já tendo validado o procedimento de cálculo do IDD e definido o procedimento de validação, executamos o experimento e obtemos os resultados.

Selecionamos as 4 instituições no topo do ranking do ENADE, para os cursos de engenharia de produção no ENADE 2014, conforme a Tabela 2.

Tabela 2 - Seleção dos Melhores do ENADE 2014 na Engenharia de Produção

Ranking ENADE	Instituição de Ensino	Código da Instituição	Código do Município	Nº. de concluintes do ENADE	Nº. De concluintes com ENEM	IDD Bruto*
1º	UFRJ	586	3304557	114	47	0,6335
2º	UNB	2	5300108	27	7	3,6373
3º	UFMTS	694	5002704	16	15	3,9821
4º	UNIRIO	693	3304557	3	2	4,3679

Fonte: INEP [CITATION INE15 \n \t \l 1046]

Nota: * - Não foi padronizado ou reescalado

Simulamos o IDD alterando os resultados dessas 4 instituições para o valor máximo do ENADE, para que fosse feita a segunda estimativa, de onde se obtêm os valores finais dos parâmetros da regressão. Em seguida, calculamos o IDD bruto como exposto na equação,

$ID D_{ij} = C_{ij} - \hat{I}_{ij}$. Com isso, foi possível o cálculo do IDD da instituição, $ID D_j$, que são calculados pela média dos $ID D_{ij}$. Esse é, portanto, o IDD bruto (sem padronização e reescalonamento) que estávamos buscando.

O procedimento exposto acima foi repetido outras três vezes, considerando além da UFRJ, as instituições UNB, UFMTS e UNIRIO, respectivamente 2º, 3º e 4º colocados no ENADE. Contudo, antes de apresentar o resultado do experimento, é importante consultar a lista das instituições com maiores valores de IDD, para que se possa identificar a pontuação necessária para atingir a primeira posição na métrica do IDD. Essa informação é apresentada na Tabela 3.

Tabela 3 - Maiores IDDs na Engenharia de Produção em 2014

Instituição de Ensino Superior	Sigla	Concluinte ENADE	Concluinte com Nota ENEM	Ranking ENADE	IDD Bruto	IDD Padronizado
Universidade Federal de Rondônia	UNIR	3	3	36º	1,79	5,00
Faculdade Arnaldo Horácio Ferreira	FAAHF	11	3	32º	0,41	5,00
Universidade Paulista	UNIP	6	3	23º	7,79	5,00
Centro Univers. Vale do Itajaí	UNIDAVI	13	6	50º	7,78	5,00
Centro Universitário Moura Lacerda	CUML	8	2	80º	6,97	5,00
Universidade Santa Cecília	UNISANTA	13	7	74º	6,83	4,94

Fonte: INEP [CITATION INE15 \n \t \l 1046]

É importante observar que, embora a primeira colocada tenha obtido uma pontuação de IDD bruto igual 11,79, ao ter a nota padronizada, esse valor foi limitado ao teto estipulado pelo INEP cujo valor é igual a 5. Isso ocorre devido a regra introduzida na nota técnica DAES/INEP nº 58/2015 para o CPC 2014 [CITATION INE15 \l 1046], onde é informado que para escores z do IDD bruto menores que -3 ou maiores que 3, estes são considerados outliers e recebem respectivamente, as notas padronizadas 0 e 5. Portanto, a partir do escore z do IDD maior do que 3, ou a partir do IDD bruto maior do que 7,10 (valor calculado utilizando interpolação linear para esse caso específico), temos que uma determinada instituição assumira a primeira posição mesmo que esteja empatada com outras instituições, também na primeira colocação. Cabe observar que a instituição CUML, indicada na Tabela , possui nota igual a 5,00 devido sua nota ter sido arredondada para cima.

Estando ciente de que para uma instituição que almeja atingir a primeira colocação precisa obter um IDD bruto maior ou igual a 7,10. Apresentamos os resultados considerando as 4 instituições com maiores notas de ENADE, que conforme observado na Tabela 2, não estão na primeira colocação no ranking do IDD. Descobrimos pelo experimento em que posição estariam essas instituições, no caso de, por hipótese, conseguirem obter valor máximo no exame ENADE para todos os seus estudantes. Os resultados podem ser verificados na Tabela 3 a seguir.

Tabela 3 - Resultado do Experimento Considerando Pontuação Máxima

Instituição de Ensino Superior	Ranking ENADE	IDD Bruto Original	IDD Bruto Máximo (Dados Simulados)	Ranking IDD Padronizado Original	Ranking IDD Padronizado (Dados Simulados)
UFRJ	1º	0,6335	3,73	114º	22º
UNB	2º	3,6373	21,88	22º	1º
UFMTS	3º	3,9821	12,17	16º	1º
UNIRIO	4º	4,3679	30,52	9º	1º

Fonte: Elaborado pelo autor.

Os resultados da Tabela mostram que das 4 instituições analisadas 3 delas iriam para a primeira colocação da métrica do IDD, caso seus alunos tirassem nota máxima no ENADE, o que seria esperado em todos os casos. Entretanto, com uma delas isso não ocorreu, e é esta unidade que revela e comprova aquilo que queríamos demonstrar. A unidade em questão é a UFRJ. Esta instituição obteve em 2014, a maior pontuação no exame ENADE, para a área da engenharia de produção, mas também, a maior nota de corte do SISU (ENEM) para ingresso em seu quadro discente [CITATION VCS14 \l 1046]. Ao observar o resultado dessa instituição, verificamos que mesmo que todos os seus alunos atingissem nota máxima, ainda assim a instituição não conseguiria alcançar a primeira posição no ranking do IDD. Na verdade, se todos os alunos dessa instituição tirassem nota máxima, a posição dessa instituição no ranking seria a 22ª colocação, ou seja, muito abaixo das primeiras posições, o que contrasta com o que seria esperado.

Isso ocorre, principalmente, porque para que seja possível a obtenção de um valor de IDD alto suficiente para competir nas primeiras colocações, é necessário que a nota de corte no SISU não seja extremamente elevada, o que no caso da UFRJ não acontece. A nota do ENADE é limitada, possui um teto, e a nota máxima possível é 100 pontos. Então, é impossível para uma instituição atingir nota do ENEM acima de 100 pontos, mesmo que fosse necessário, para obter um IDD da mesma magnitude dos outros competidores. Essa é uma importante inconsistência do modelo desenvolvido para o IDD, como se queria demonstrar, e supõem-se que a mesma passou despercebida nas fases de conceitualização e validação do mesmo. Tendo sido descoberta essa deficiência, se faz necessária sua reformulação visando adequá-lo.

6 CONCLUSÃO

Estudos que buscam mensurar o ganho de aprendizagem são sem dúvida instrumentos importantes para traçar caminhos que busquem a melhoria do ensino. Entretanto o modelo utilizado para o cálculo do IDD pelo INEP se mostrou inadequado com relação a sua utilização para a determinação de rankings devido a deficiência em sua construção conceitual, visto que o modelo não permite que instituições com desempenhos excepcionais no ENEM, consigam por melhor que sejam, o resultado necessário no ENADE para que possam competir com outras instituições, simplesmente devido ao fato das notas do ENADE terem uma determinada pontuação máxima, ou seja, um limitante superior. Além disso, a forma como vem sendo calculado acarreta em baixo nível de confiança na determinação das suas estimativas, muito em função do pequeno tamanho das amostras obtidas para cada uma das instituições avaliadas.

Os modelos de regressão linear continuam sendo valiosos no intuito de medir esse ganho de aprendizado, mas devemos reforçar que apresentou inconsistência quanto ao uso dado pelo INEP, na tentativa de gerar estimativas que busquem classificar ou fazer ranking

de instituições. Devemos ter em mente, que esse instrumento pode causar impacto negativo no prestígio de instituições de ensino tradicionais e conceituadas e, além disso, é levado em consideração para determinar se as instituições avaliadas podem ter suas atividades mantidas.

Sugerimos como resultado desse estudo, caso haja interesse na continuidade da métrica do IDD pelo MEC, que a verificação apresentada no estudo seja realizada entre o grupo de instituições com maior nota de corte no ENEM, de forma a garantir que tal instituição possa ao menos ser capaz de superar suas competidoras matematicamente. Resgatando questão apontada por Trembley, Lalancette e Roseveare [CITATION Tre12 \n \t \l 1046], que já tinha sido mencionada na revisão de literatura, ao analisar outros estudos, é importante também no modelo do IDD, que seja ampliado significativamente o tamanho mínimo amostral de alunos (atualmente 3 alunos), de forma a obtermos uma amostra mais representativa de alunos de cada instituição, aumentando assim a confiança nos resultados.

Outro aspecto que deveria ser levado em consideração é a equivalência em termos de grau de esforço requerido por uma instituição, para agregar conhecimento em alunos de diferentes perfis. Pode-se alegar, por exemplo, que se agregar 1 ponto de IDD a um aluno muito bom seja muito mais custoso do que agregar esse mesmo valor a um aluno mediano. Fazendo um paralelo com o atletismo, é muito mais fácil conseguir uma melhora de 3 segundos no tempo de um atleta que corre 100 metros rasos em 30 segundos, do que melhorar em 3 segundos o tempo de um atleta que corre os 100 metros rasos em 12 segundos, sabedores que o recorde mundial dos 100 metros é de 9,58 segundos. Ou seja, está problemática está muito provavelmente atrelada a diferentes retornos de escala e diferentes fronteiras de produtividade. Esses conceitos são adequadamente tratados pela metodologia de análise envoltória de dados.

Por fim, acreditamos ter atingido o objetivo proposto neste trabalho de se verificar a validade do IDD, e acredito que trouxemos contribuição não apenas quanto à necessidade de aprimoramento quanto ao aspecto avaliado pelo experimento, mas também nos demais aspectos citados nessa seção, além de aumentar a compreensão quanto à existência e importância dos indicadores de ganho de aprendizagem de uma forma mais ampla.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- BANTA, T. W. Addressing a question of credibility. **Assessment Update**, v. 24, n. 6, p. 3-4, 2012.
- BITTENCOURT, H. R. et al. Uma análise da relação entre os conceitos Enade e IDD. **Estudos em avaliação educacional**, v. 19, n. 40, p. 247-262, 2008.
- BITTENCOURT, H. R.; OLIVEIRA CASARTELLI, A.; MORAIS RODRIGUES, A. C. Sobre o índice geral de cursos (IGC). **Avaliação: Revista da Avaliação da Educação Superior**, v. 14, n. 3, 2009.
- BOGOYA, J. D.; BOGOYA, J. M. An academic value-added mathematical model for higher education in Colombia: Caso de la educación superior en Colombia. **Ingeniería e investigación**, v. 33, n. 2, p. 76-81, 2013.
- BORDEN, V. M. H.; YOUNG, J. W. Measurement validity and accountability for student learning. **New Directions for Institutional Research**, p. 19-37, 2008.
- BOYAS, E.; BRYAN, L. D.; LEE, T. Conditions affecting the usefulness of pre-and post-tests for assessment purposes. **Assessment & Evaluation in Higher Education**, v. 37, n. 4, p. 427-437, 2012.
- BRASIL, L. Decreto Nº 5.773. de 9 de maio de 2006, que dispõe sobre o exercício das funções de regulação, supervisão e avaliação de instituições de educação superior e cursos superiores de graduação e sequenciais no sistema federal de ensino. **Presidência da República**, Brasília, 2006.

- CAMPOS, S. R. M.; HENRIQUES, R.; YANAZE, M. H. Higher education in Brazil: an exploratory study based on supply and demand conditions. **Universal Access in the Information Society**, v. 17, n. 4, p. 711-733, 2018.
- COATES, H. What's the difference? A model for measuring the value added by higher education in Australia. **Higher Education Management and Policy**, v. 21, n. 1, p. 1-20, 2009.
- COUTINHO, R. X. et al. Brazilian scientific production in science education. **Scientometrics**, v. 92, n. 3, p. 697-710, 2012.
- CUNHA, J. M.; MILLER, T. Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. **Economics of Education Review**, v. 42, p. 64-77, 2014.
- CUNHA, J. M.; MILLER, T. Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. **Economics of Education Review**, v. 42, p. 64-77, 2014.
- DOUGLASS, J. A.; THOMSON, G.; ZHAO, C.-M. The learning outcomes race: The value of self-reported gains in large research universities. **Higher education**, v. 64, n. 3, p. 317-335, 2012.
- FERNANDES, S.; FLORES, M. A.; LIMA, R. M. Students' views of assessment in project-led engineering education: findings from a case study in Portugal. **Assessment & Evaluation in Higher Education**, v. 37, n. 2, p. 163-178, 2012.
- FERRANTE, F. Assessing quality in higher education: Some caveats. **Social Indicators Research**, v. 131, n. 2, p. 727-743, 2017.
- GARCIA, A. A.; NICOLINI, A. M.; ANDRADE, R. O. B. Valor agregado aos estudantes nos cursos de administração: uma comparação do desempenho entre instituições de ensino superior públicas e privadas. **Revista Gestão Universitária na América Latina-GUAL**, v. 7, n. 2, p. 24-46, 2014.
- GOLDSTEIN, H.; THOMAS, S. Using examination results as indicators of school and college performance. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, v. 159, n. 1, p. 149-163, 1996.
- GRAYSON, J. P. Using surveys to measure 'value added' in skills in four faculties. **Canadian Journal of Higher Education**, v. 29, n. 1, 1999.
- GREGORY, S. et al. Gregory, Simon, et al. "Exploring the potential uses of value-added metrics in the context of postgraduate medical education. **Medical teacher**, v. 38, n. 10, p. 1003-1010, 2016.
- HERINGER, R. Affirmative action and the expansion of higher education in Brazil. In: _____ **Race, Politics, and Education in Brazil**. [S.l.]: Springer, 2015. p. 111-131.
- INEP. Conceito Preliminar de Curso (CPC). INEP, 2015. Disponível em: <<http://inep.gov.br/conceito-preliminar-de-curso-cpc>>. Acesso em: 5 Setembro 2017.
- INEP. Documentos e Legislação. INEP, 2016. Disponível em: <http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2016/nota_tecnica_n32_2017_cgqes_daes_calculo_conceito_enade.pdf>. Acesso em: 05 Agosto 2017.
- JOHNES, J. Measuring efficiency: A comparison of multilevel modelling and data envelopment analysis in the context of higher education. **Bulletin of Economic Research**, v. 58, n. 2, p. 75-104, 2006.
- JOHNES, J. University rankings: What do they really show? **Scientometrics**, v. 115, n. 1, p. 585-606, 2018.
- KELLY-WOESSNER, A.; WOESSNER, M. Promoting a "House of study". **Society**, v. 48, n. 3, p. 227-231, 2011.
- KONG, W.-H.; FU, T.-T. Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added performance indicators. **Omega**, v. 40, n. 5, p. 541--549, 2012.

LETA, J.; GLÄNZEL, W.; THIJS, B. Science in Brazil. Part 2: Sectoral and institutional research profiles. **Scientometrics**, v. 67, n. 1, p. 87-105, 2006.

LIU, O. L. Measuring value-added in higher education: conditions and caveats--results from using the Measure of Academic Proficiency and Progress (MAPPTM). **Assessment & Evaluation in Higher Education**, v. 36, n. 1, p. 81-94, 2011.

MAGALHÃES CASTRO, M. H. Higher education policies in Brazil: A case of failure in market regulation. In: _____ **Higher education in the BRICS countries**. [S.l.]: Springer, 2015. p. 271-289.

MCGRATH, C. H. et al. Learning gain in higher education. **Santa Monica, CA: RAND Corporation**, 2015.

MELGUIZO, T.; WAINER, J. Toward a set of measures of student learning outcomes in higher education: evidence from Brazil. **Higher Education**, v. 72, n. 3, p. 381-401, 2016.

MILLA, J.; MARTÍN, E. S.; VAN BELLEGEM, S. Higher education value added using multiple outcomes. **Journal of Educational Measurement**, v. 53, n. 3, p. 368-400, 2016.

MINISTÉRIO DA EDUCAÇÃO. Portaria Normativa nº 4, de 5 de agosto de 2008. Regulamenta a aplicação do conceito preliminar de cursos superiores-CPC, para fins dos processos de renovação de reconhecimento respectivos, no âmbito do ciclo avaliativo do Sinaes instaurado pela Portaria. **Diário Oficial da República Federativa do Brasil**, 2008.

PAMPAKA, M.; WILLIAMS, J.; HUTCHESON, G. Measuring students' transition into university and its association with learning outcomes. **British Educational Research Journal**, v. 38, n. 6, p. 1041-1071, 2012.

PEREIRA, D.; FLORES, M. A.; NIKLASSON, L. Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. **Assessment & Evaluation in Higher Education**, v. 41, n. 7, p. 1008-1032, 2016.

PICININ, C. T. et al. Comparison of performance of researchers recipients of CNPq productivity grants in the field of Brazilian production engineering. **Scientometrics**, v. 109, n. 2, p. 855-870, 2016.

PIDD, M. **Modelagem Empresarial: Ferramentas para tomada de decisão**. Porto Alegre: Bookman, 1998.

PIKE, G. R. Considerations when Using Value-Added Models in Higher Education Assessment. **Assessment Update**, v. 28, n. 5, p. 8-10, 2016.

POLIDORI, M. M. Políticas de avaliação da educação superior brasileira: Provão, SINAES, IDD, CPC, IGC e outros índices. **Avaliação: Revista da Avaliação da Educação Superior**, v. 14, n. 2, 2009.

PORTELA, M. C.; CAMANHO, A. S.; KESHVARI, A. Assessing the evolution of school performance and value-added: trends over four years. **Journal of Productivity Analysis**, v. 39, n. 1, p. 1-14, 2013.

RABE-HESKETH, S.; SKRONDAL, A. **Multilevel and longitudinal modeling using Stata**. [S.l.]: STATA press, 2008.

RODGERS, T. Measuring value added in higher education: Do any of the recent experiences in secondary education in the United Kingdom suggest a way forward? **Quality assurance in education**, v. 13, n. 2, p. 95-106, 2005.

ROOHR, K. C.; LIU, H.; LIU, O. L. Investigating student learning gains in college: a longitudinal study. **Studies in Higher Education**, v. 42, n. 12, p. 2284-2300, 2017.

ROSA BECKER, F.; COSTA, L. C. Brazil: Shift of Accountability Incentives. In: _____
Education Policy Reform Trends in G20 Members. [S.l.]: Springer, 2013. p. 173-188.

SATOLO, E. G. et al. Perception of the Evolution of Industrial Engineering Areas Based on the Brazilian ENADE-INEP Assessment System. In: _____ **Engineering Systems and Networks**. [S.l.]: [s.n.], 2017. p. 355--363.

SCHWARTZMAN, S. Uses and abuses of education assessment in Brazil. **Prospects**, v. 43, n. 3, p. 269-288, 2013.

SHAVELSON, R. J. Assessing student learning responsibly: From history to an audacious proposal. **Change: The Magazine of Higher Learning**, v. 39, n. 1, p. 26-33, 2007.

SHAVELSON, R. J. et al. On the practices and challenges of measuring higher education value added: The case of Colombia. **Assessment & Evaluation in Higher Education**, v. 41, n. 5, p. 695-720, 2016.

SOARES, J. F.; RIBEIRO, L. M.; MOURA CASTRO, C. Valor agregado de instituições de ensino superior em Minas Gerais para os cursos de Direito, Administração e Engenharia Civil, 2001.

SOBRINHO, J. D. Avaliação e transformações da educação superior brasileira (1995-2009): do provão ao SINAES. **Avaliação: Revista da Avaliação da Educação Superior**, v. 15, n. 1, 2010.

STOWELL, M. Equity, justice and standards: assessment decision making in higher education. **Assessment & Evaluation in Higher Education**, v. 29, n. 4, p. 495-510, 2004.

SULIS, I.; PORCU, M. Assessing divergences in mathematics and reading achievement in italian primary schools: a proposal of adjusted indicators of school effectiveness. **Social Indicators Research**, v. 122, n. 2, p. 607-634, 2015.

TREMBLAY, K.; LALANCETTE, D.; ROSEVEARE, D. **AHELO Feasibility Study Report. Volume 1 - Design and Implementation**. Organization for Economic Co-operation and development (OECD). [S.l.]. 2012.

VC SISTEMAS EDUCACIONAIS. Engenharia de Produção – Notas de Corte Sisu 2014 – curso de graduação. **Blog do ENEM**, 2014. Disponível em: <<https://blogdoenem.com.br/engenharia-producao-sisu-2014/>>. Acesso em: 11 Setembro 2017.

YORKE, M.; KNIGHT, P. T. Curricula for economic and social gain. **Higher Education**, v. 51, n. 4, p. 565-588, 2006.

ZHANG, L. A value-added estimate of higher education quality of US states. **Education Economics**, v. 17, n. 4, p. 469-489, 2009.