UNIVERSIDADE FEDERAL DE UBERLÂNDIA
GRADUATE PROGRAM IN ACCOUNTING
DOCTORAL STUDIES IN ACCOUNTING

VIVIAN DUARTE COUTO FERNANDES

QUALITY OF ACCOUNTING UNDERGRADUATE PROGRAMS IN BRAZIL: HOW TO ESTIMATE VALUE-ADDED?

UBERLÂNDIA – MG
2020

VIVIAN DUARTE COUTO FERNANDES

QUALITY OF ACCOUNTING UNDERGRADUATE PROGRAMS IN BRAZIL: HOW TO ESTIMATE VALUE-ADDED?

Dissertation submitted to the Graduate Program in Accounting at Universidade Federal de Uberlândia in partial fulfillment of the requirements for the degree of PhD in Accounting.

Focus area: Accounting and Controllership

Advisor: Prof. Gilberto José Miranda

Co-advisor: Prof. Nicola Alexander

UBERLÂNDIA – MG

2020

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

F363 Fernandes, Vivian Duarte Couto, 1984-
2020 Quality of Accounting Undergraduate Programs in Brazil [recurso eletrônico] : How to Estimate Value-Added / Vivian Duarte Couto Fernandes. - 2020.

Orientador: Gilberto José Miranda.
Coorientador: Nicola Alexander.
Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-Graduação em Ciências Contábeis.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.te.2020.365
Inclui bibliografia.

1. Contabilidade. I. Miranda, Gilberto José,1974-, (Orient.). II. Alexander, Nicola,1966-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-Graduação em Ciências Contábeis. IV. Título.

CDU: 657

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Ciências Contábeis
Av. João Naves de Ávila, 2121, Bloco 1F, Sala 248 - Bairro Santa Monica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3291-5904 - www.ppgcc.facic.ufu.br - ppgcc@facic.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Defesa de:	Tese de Doutorado Acadêmico Número 010- PPGCC				
Data:	10 de fevereiro de 2020	Hora de início:	13h00min	Hora de encerramento:	16h45min
Matrícula do Discente:	11613CCT010				
Nome do Discente:	Vivian Duarte Couto Fernandes				
Título do Trabalho:	Qualidade dos Cursos de Ciências Contábeis no Brasil: afinal, como calcular o valor agregado?				
Área de concentração:	Contabilidade e Controladoria				
Linha de pesquisa:	Educação Contábil				
Projeto de Pesquisa de vinculação:	FAU-UFU N° 02/2018- A eficiência dos professores sobre o desempenho dos alunos de ciências Contábeis no Brasil: uma perspectiva a partir do modelo de valor agregado				

Reuniu-se na sala 1F146 Campus Santa Mônica, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciências Contábeis, assim composta: Professores Doutores: Silvia Pereira de Castro Casa Nova (FEA/USP); Sheizi Calheira de Freitas (UFBA); Camila Lima Coimbra (UFU); Janser Moura Pereira (UFU); Nicola A. Alexander (CEHD) e Gilberto José Miranda(UFU), orientador da candidata. Os Professores Doutores Silvia Pereira de Castro Casa Nova; Sheizi Calheira de Freitas participaram da Banca examinadora por meio do sistema de web conferência. A Professora Nicola A. Alexander teve a participação epistolar.

Iniciando os trabalhos o presidente da mesa, Dr. Gilberto José Miranda, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à discente a palavra para a exposição do seu trabalho. A duração da apresentação da discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

APROVADA

Esta defesa faz parte dos requisitos necessários à obtenção do título descrito na tabela acima. O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Gilberto José Miranda**, **Presidente**, em 13/03/2020, às 16:05, 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 **Janser Moura Pereira**, **Professor(a) do Magistério Superior**, em 13/03/2020, às 16:22, 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 **Camila Lima Coimbra**, **Professor(a) do Magistério Superior**, em 13/03/2020, às 18:14, 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 **Sheizi Calheira de Freitas**, **Usuário Externo**, em 16/03/2020, às 14:17, 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 **Silvia Pereira de Castro Casa Nova**, **Usuário Externo**, em 17/03/2020, às 10:34, 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 https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1939292** e o código CRC **AEC44E56**.

Referência: Processo nº 23117.006318/2020-44 SEI nº 1939292

ACKNOWLEDGMENTS

To God in the first place, because He is and has been always good! He never forsakes His children, even when they feel lonely. He carries us, He waits, He cares, He soothes. He was the one to whom I clung when I had no further strength… He has been there to show me the best, the most accurate paths to take so I could be here.

At the same time, I thank my parents Pedro and Rosângela for teaching me the value of faith and love. For providing me and my siblings – Karin, Ingrid, and Pedro Junior – the most harmonious, loving, and wonderful home we could ever have experienced. I am grateful for such an amazing, special family. Getting here is undoubtedly the result of your commitment to our education, which has been a priority in our home since our childhood. I remember the several times we were delighted to look for solutions to difficult, controversial problems from various fields whenever any of us did not know the answer in the first place. Yeah, I am really a privileged woman!

I thank my siblings for our partnership. Words cannot express my gratitude for those times when I did not fancy calling you, and you went right after me; when I said I could not go somewhere because I was "distressed" about the dissertation, and you said, "Oh, Vivian, it never ends! You can catch up on Monday!", convincing me to leave the office. Thanks for hearing my desperate cries and calming me down by saying, "My dear, if it were easy, everyone was a PhD!" Sincerity is our forte! (*chuckles*)

My heartfelt 'thank you' to the family I have nurtured: Matheus, Rafael, and Lucas. You are the best thing in my life. You are the three men who rock my world in every single way! Matheus, thank you for being such a partner. I know it was not easy over the last four years. I showed up to virtually none of your meetings, and you were a widower of a living woman! Thank you for supporting my decision of spending one year abroad. We were apart for such a long time – I am truly sorry and thank you for overcoming this with me. Rafa and Lucas, please forgive me for being an absent mother in this period. Please remember that you are my greatest joy, motivation and reason to be a better person, always! I love you truly and deeply.

I thank Professor Gilberto, my supervisor – the best I could ever have. Thank you for your patience, reasonableness, support. For helping me become a better educator. For asking so much, for making me wonder why about so many things. I will be forever grateful for this experience under your tuition. I also thank Professor Nicola, my co-supervisor, who accepted my request in her first email and who has been kind, lovely and punctual since our first meeting. Thank you, Professor Nicola, for making me look to the unequal side of education

provision, to the impact of education policy, and to the objectivity needed in my research! Thank you for your amazing classes, from which I have learnt a lot!

I am thankful to my classmates in the graduate program. They made this walk more enjoyable, shared the joys and hardships of these years. I am especially thankful to those whom I am proud to call my friends: Neire, Thayse, Edilei, and Reiner. You are the guys! I am thankful to the professors at my undergraduate program in Universidade Federal de Uberlândia and to those at the University of Minnesota – especially Stuart Yeh and John LaVelle, who have always been available. I am also thankful to the staff at both universities, as they were responsible for every arrangement that made it happen.

Thanks to my friends from the School of Management (FAGEN) – Etienne, Luciana Cezarino, Michelle, Valeriana, Márcia, Andrea, and André – who supported me in every decision, from my leave to my one-year-long doctoral stay in the US, and in discussing some specific topics in my dissertation. You would also invite me to meet up and relax with good music and laughter. I am especially grateful to Etienne and André, both of whom listened to my regrets and soothed my heart, especially when my emotional health was shaken.

It is also with the purest feeling of love and gratitude that I thank you, my dear friends from Minnesota, US. My doctoral stay has brought me a new perspective both on life and on my studies, but it has also presented me with great and eternal friends. First of all, I thank you, Marina, for kindly greeting me at the airport and welcoming me in the early days at your place. Then, I thank you, Alessandra, Matheus, Gael, and Arthur, for giving me hope and peace in a moment of despair. You are inspiring! Marcos, Miriam, and Maria, you made my weekends livelier, and gave me precious tips. Alessandra and Miriam, you welcomed me as family – thank you so very much! Gayle and Randy, you were parents to me in Minnesota. I have lots of gratitude and love for what you did for me! Thiago, Ana, Scott, Rose, Victoria, Emily, Manuela and so many dear ones who are there: thank you! I hope to see you soon.

At last but not least, thank you, Professor Janser, for devoting some of your precious time helping me design the methodology for this dissertation. Words cannot express my gratitude for your listening to me patiently and helping me do all the statistics. And I am also thankful to all members of the qualifying committee and the examining committee. You all made valuable contributions!

Thanks to the Coordination for the Improvement of Higher Education Personnel (CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), which funded my doctoral stay in the US. I also thank to University Support Foundation (FAU - Fundação de Apoio Universitário) and to the National Council for Scientific and Technological

Development (CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the support in carrying out this research.

ABSTRACT

School quality assessment has been the object of research by economists, educators, policy makers and various stakeholders worldwide. In Brazil the National Assessment System for Higher Education (Sinaes – Sistema Nacional de Avaliação do Ensino Superior) is a sound initiative that seeks to assess the country's undergraduate programs, their faculty members, and the students' academic achievement, as well as to provide quality indicators that account for the differences between them. One of such indicators is the Indicator of Difference between Observed and Expected Achievements (IDD – Indicador de Diferença entre os Desempenhos Observado e Esperado), which measures the contribution of an undergraduate program to its students' achievements upon their undergraduate studies. The first IDD was released in 2006; since then, lawmakers have changed its estimation methodology, seeking to improve it as an accurate measure of value added. This doctoral dissertation aims to discuss such different methodologies and their impacts on the ranking of undergraduate programs in Accounting in Brazil, as provided through an indicator named Preliminary Program Quality Level (CPC – Conceito Preliminar de Curso). A quantitative design was used both to test four value-added models that included the historical records of the IDD and to identify the impact of changes in estimation methodology implemented since 2006 on the quality indicators of the undergraduate programs under scrutiny. The analysis was based on data from Brazilian Accounting students who took the 2015 National Exam of Student Achievement (Enade), completed a student questionnaire and had a valid score in the National Exam of High School Education (Enem) for admission to an undergraduate program. The sample consisted of 30,668 students from 911 undergraduate programs, which represents 46.98% of the total population. The results show that the current model is more accurate than the previous ones for estimating value added. However, both the literature and the findings indicate that the model could be improved by introducing explanatory variables for academic achievement that cannot be controlled by higher education institutions. Even though the current model is statistically more robust than the previous ones, it is still inappropriate because it misrepresents all institutions and all students as equals. A new IDD estimation methodology, Model IDD-VDCF, is proposed to estimate value added while still including significant variables of academic achievement that cannot be controlled by organization leaders. The findings point to 1) gender, marital status, and reading and study habits as control variables at the student level, and 2) type of higher education institution, learning modality, and regional location as control variables at the institution level. Such variables reduced the IDD estimate bias associated with student selection for admission to a given undergraduate program. This study suggests that the IDD is relevant to identify differences across undergraduate programs and it should bear more weight within the Sinaes. It also suggests that the CPC quality index should be abolished, as it is not able to portray the actual quality of the undergraduate programs. While the IDD points to significant differences across the undergraduate programs regardless of the methodology used, the rationale underlying the CPC estimation reduces such differences and renders distinct programs as similar in quality. Not only does this study contribute to the methodological discussion about how to estimate value added in higher education, but it also adds to the debates about how to assess the effectiveness of higher education institutions in order to support policy making. The students' learning context must be considered when comparing performance across institutions through standardized exams such as the Enade – even though the undergraduate program bears great responsibility for their academic achievement, such factors as motivation, commitment, and professional aspirations, which are not subject to control by higher education institutions, can impact the academic outcomes.

Keywords: value-added, school effectiveness, quality, higher education.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AACSB	Association to Advance Collegiate Schools of Business
ACBSP	Accreditation Council for Business Schools and Programs (ACBSP)
ACT	American College Testing
AHELO	Assessment of Higher Education Learning Outcomes
AIC	Akaike Information Criterion
AICPA	American Institute of Certified Public Accountant
ANOVA	Analysis of Variance
BIC	Bayesian Information Criterion
BRL	Brazilian Real (currency)
CAAP	Collegiate Assessment of Academic Proficiency
CAGED	Cadastro Geral de Empregados e Desempregados (Brazilian Register of Employed and Unemployed)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Coordination for the Improvement of Higher Education Personnel)
CC	Conceito de Curso (Program Quality Level)
CE	Componente Específico (Specific Training Section)
CFC	Conselho Federal de Contabilidade (Federal Board of Accountants)
CH	Ciências Humanas e Suas Tecnologias (Humanities and Their Technologies)
CI	Confidence Interval
CIA	Certified Internal Auditor
CISA	Certified Information Systems Auditor
CLA	Collegiate Learning Assessment
CMA	Certificate Management Accountant
CN	Ciências Naturais e suas Tecnologias (Natural Sciences and Their Technologies)
CONAES	National Commission for Assessment of Higher Education (Comissão Nacional de Avaliação da Educação Superior)
CPA	Comissão Própria de Avaliação (HEI-run Assessment Committee)
CPC	Conceito Preliminar de Curso (Preliminary Program Quality Level)
CTAA	Comissão Técnica de Acompanhamento da Avaliação (Technical Committee for Assessment Follow-up)
DCN	Diretrizes Curriculares Nacionais (National Curriculum Guidelines)
DF	Degrees of Freedom
Enade	Exame Nacional de Desempenho dos Estudantes (National Exam of Student Achievement)
Enem	Exame Nacional do Ensino Médio (National Exam of High School Education)
EPF	Education Production Function
FEA	Faculdade de Economia, Administração e Contabilidade
FG	Formação Geral (General Training Section)
Fies	Financiamento do Ensino Superior (Financing Higher Education)
GDP	Gross Domestic Product

HEI	Higher Education Institution
HLM	Hierarchical Linear Model
IBGE	Instituto Brasileiro de Geografia e Estatística (Brazilian Institute of Geography ans Statistics)
IACBE	International Assembly for Collegiate Business Education
ICC	Intraclass Correlation Coefficient
ICMA	Institute of Certified Management Accountant
IDD	Indicador de Diferença entre os Desempenhos Observado e Esperado (Indicator of Difference between Observed and Expected Achievements)
IFF	Infraestrutura e Instalações Físicas (Infrastructure and Physical Facilities)
IGC	Índice Geral de Cursos (General Index of Programs)
Inep	Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (National Institute of Educational Research Anísio Teixeira)
IPCA	Índice de Preços ao Consumidor Amplo (Extended Consumer Price Index)
LT	Languages, Codes, and Their Technologies
MAPP	Measure of Academic Proficiency and Progress
MBA	Master of Business Administration
MEC	Ministry of Education
MT	Matemática e suas Tecnologias (Mathematics and Its Technologies)
NASBA	National Association of State Boards of Accountancy
NCLB	No Child Left Behind Act
NT	Nota Técnica (Technical Bulletin)
OAF	Oportunidade de Ampliação da Formação (Opportunity for Further Training/Learning)
ODP	Organização Didático-Pedagógica (Pedagogical Teaching Structure)
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PDI	Plano de Desenvolvimento Institucional (Institutional Development Plan)
Pisa	Programme for International Student Assessment
Prouni	Programa Universidade para Todos (University for All)
REUNI	Restructuring and Expansion of Federal Universities (Reestruturação e Expansão das Universidades Federais)
SAM	Student Achievement Measure
SAT	Scholastic Assessment Test
SD	Standard Deviation
Sinaes	Sistema Nacional de Avaliação do Ensino Superior (National Assessment System for Higher Education)
SS	Sum of Squares
TQM	Total Quality Management
TVAAS	Tennessee Value-Added Assessment System
UFU	Universidade Federal de Uberlândia
UNESCO	United Nations Educational, Scientific and Cultural Organisation
US	United States

USAID	Agency for International Development
USP	Universidade de São Paulo
VAM	Value-Added Model
VIF	Variance Inflation Factor
VSA	Voluntary System of Accountability

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Context

Addressing quality in higher education is not a trivial task, as this concept takes on different meanings in different contexts (Lagrose et al., 2004). On the one hand, the notion of 'quality' is relative from two perspectives – from that of the speaker and from that of the context in which it is stated (Krause, 2012) – i.e., the term 'quality' can have different meanings to different people, depending on the context of its use. On the other hand, the nation of quality may be absolute or require predefined standards (Harvey & Green, 1993).

Students may find quality important, but they usually seem to take it for granted. Employers are also concerned with quality when it comes to recruiting faculty to teach and research. Policy makers, too, deem quality as relevant, particularly when circumstances change, participation levels increase, access to higher education increases, audits or assessments take place, and/or human and physical resources are too limited. In turn, private enterprises are concerned with quality when competition between higher education institutions (HEIs) intensifies, leading to disputes for market share and search for excellence-driven differentiation across organizations (Harvey & Green, 1993).

In Brazil the National Assessment System for Higher Education (Sinaes – Sistema Nacional de Avaliação do Ensino Superior) was created through Act No. 10861, as of 2004, to assess the quality of higher education institutions, their undergraduate programs, and their students' academic achievement. It targets various dimensions of education, including teaching, research, extension and outreach, social responsibility, program coordination, faculty, and facilities. A number of instruments are used to assign rates to each of these dimensions and provide a quality indicator-based overview of the undergraduate programs and higher education institutions in the country (Brasil, 2004).

One of the Sinaes quality indicators is the Indicator of Difference between Observed and Expected Achievements (IDD – Indicador de Diferença entre os Desempenhos Observado e Esperado). This is "a quality indicator that measures the value that an undergraduate program adds to the development of its seniors by probing their achievements on the Enade[1] as compared to their developmental characteristics at the beginning of their

[1] Exame Nacional de Desempenho dos Estudantes (National Exam of Student Achievement), an exam for measuring the achievement of higher education students in specific study tracks.

study track"[2] (Brasil, 2017b, p. 1). In other words, this indicator, the object of study in this doctoral dissertation, is aimed at quantifying how much each higher education institution adds to a student's achievement throughout his/her undergraduate studies.

Such notion of quality assumes that education is a process of continuous transformation of the students, one which empowers them and adds value to them (Harvey & Green, 1993). This is consistent with Austin (1980), according to whom quality in education should be considered as a continuous process that also includes assessing how much an institution contributes to its students' intellectual, cultural and social development.

The notion of value added stems from economic theory, which posits that the value of a good increase at each stage of the production process because of an activity, resource, or process that leads to the output (Koedel et al., 2015). In higher education, value added can be defined as the difference between the college seniors' achievement and that of the freshmen, which gives an estimate of how much a student has learnt in a given period (Liu, 2010). Approaches that attempt to identify value-added dimensions provide clearer insights into what has been transformed, but their downside is that they require a representative measure of the outputs (Harvey & Green, 1993).

The goal of a value-added model (VAM) is to estimate a production function that accounts for student achievement, building on the assumption that cognitive development is the output of a cumulative process influenced by a history of inputs provided by families and schools (Todd & Wolpin, 2003). Such a model, which originally emerged in basic education to explain the role played by the teacher, allows for a statistical analysis of determinants of achievement in marginal terms. To this end, such a model needs to track the students' individual progresses and relate them to the teacher (or school) responsible for their achievements.

1.2 Research Problem and Objectives

The literature is controversial regarding the value-added methodology of choice and the role that a VAM can play in assessing both teacher and educational institutions. However, a series of widely accepted research findings has shown that the VAM is the most suited model for measuring the school effects on student achievement (Koedel et al., 2015).

---

[2] The author's translation for: "*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*".

A large criticism toward the methodological basis of a VAM is the lack of consensus about the control variables to include in the model (Center for Education Policy Research, 2011). Neglecting the students' individual characteristics (e.g., engagement in extracurricular training), for one, affects the measurement of their academic achievement, and consequently the value added will not reflect the actual quality of a given school.

In both higher education and basic education, the conflicting debates over the VAM lie in its methodology. Kim and Lalancette (2013) divide the studies in the area into three groups: 1) those focused on the large-scale tests' ability to measure what is really important or valuable (e.g., Camargo et al., 2016), 2) those centered around how to measure value added, assuming that measurement errors may bias the findings (e.g., Liu, 2011; Melguizo et al., 2017), and 3) those devoted to the test result, i.e., whether it is a continuous or a categorical score, whether it defines a top achievement, whether its score is reliable and effective (e.g., Horn et al., 2017). The present study falls under group 2, as it will soon be clear.

In Brazil the National Exam of Student Achievement (Enade – Exame Nacional de Desempenho dos Estudantes) is used to assess the achievement of those graduating from undergraduate programs. Applied by the National Institute of Educational Research Anísio Teixeira (Inep – Instituto Nacional de Pesquisas Educacionais Anísio Teixeira) since 2004, it is a standardized knowledge exam that assesses 1) the students' development against the expected syllabus defined by the national curriculum guidelines for the undergraduate studies in a given domain, 2) their development of competences and skills necessary for solid professional practice, and 3) their awareness of the current state of affairs in Brazil and worldwide (Brasil, 2018a).

The present study assumes that Enade is a solid, representative exam of the academic achievement provided by undergraduate programs in Brazil. This is the reason why this study does not fit into group 1 mentioned above. Also, this dissertation does not conform to group 3, as it is not aimed at discussing implications or differences between measures related to the Enade result, their validity, or their reliability.

Inep has published Technical Bulletins (NT – Nota Técnica) on how it estimates quality indicators for higher education every year since 2004, when the Sinaes was created. As for the object of study in this dissertation – the indicator IDD, which measures value added – four different methodologies have been applied at some moment from 2004 through 2019 to estimate the contribution of undergraduate programs to student achievement. To the best of the author's knowledge, no studies have delved into their differences and the

implications thereof for understanding quality indicators for undergraduate programs in Brazil.

The literature suggests that different methods of estimating academic gain produce different findings (Kim & Lalancette, 2013; Liu, 2011; Melguizo et al., 2017; Pike, 1992; Steedle, 2012). Like any other assessment model, the VAM cannot be a standalone parameter to underlie or bear out public policies, and its modeling require caution in fitting school and family characteristics, so as not to reinforce the disadvantages of HEIs which have a relatively low percentage of students from lower socioeconomic background. Yet, it is still possible to use the VAM findings to compare the units under scrutiny by looking into the institutions' achievements against the mean, which includes all other institutions (Liu, 2011; National Research Council, 2010).

This study is relevant not only because of its theoretical contribution to using the VAM to measure the quality of undergraduate programs in Brazil, but most importantly because it scrutinizes the advantages or disadvantages that an institution may have depending on the model used for assessment. The inconsistent definition of the IDD estimation methodology since 2006 has had a significant impact on the ranking results of the institutions, with implications for their brands (meaning prestige), allocation of financial resources, and access to public funding.

Accounting has been one of the five largest fields of undergraduate studies in Brazil since 2009. It currently ranks third in number of undergraduate programs in the country and fourth in number of enrollments, only falling below Law, Pedagogy, and Administration (Inep, 2019). Yet, the numbers are disturbing in other fronts. Only 30% (11,210 out of the 37,051) applicants passed the 2019-2 proficiency exam applied by the Federal Board of Accountants (CFC – Conselho Federal de Contabilidade). Meanwhile, 1,101 undergraduate programs in Accounting had students taking the 2019 Enade and obtained the following quality levels: 50 rated 1 (worst level); 348, 2; 478, 3 (satisfactory level); 166, 4; and 42, 5 (best level) (Inep, 2019), i.e., 36% of them did not reach a satisfactory quality level according to the criteria set forth by Inep.

This dissertation aims to show the impact of changing the IDD estimation methodology on the quality levels of undergraduate programs in Accounting in Brazil. The specific objectives include:

1) compare the Sinaes value-added models that have been developed since the system's creation;

2) rank the Brazilian undergraduate programs in Accounting based on the different value-added models available;
3) identify the determinants of achievement for Accounting students in all editions of the Sinaes assessment (2006, 2009, 2012, and 2015);
4) assess the potential of a new VAM that includes the determinants of achievement in undergraduate programs in Accounting;
5) discuss what quality for undergraduate programs in Accounting is.

1.3 Relevance of the Topic and Justification

Data provided by the Higher Education Census (Inep, 2018) lay bare the complexity and importance of this level of education in Brazil, given the significant number of enrollments, undergraduate programs and institutions from North to South. The numbers for undergraduate studies have increased significantly over the last ten years in the country: overall it has increased 51% for new students, 56.4% for enrollments, and 52% for graduates (Inep, 2018).

The Ministry of Education (MEC – Ministério da Educação) estimates that the direct public investment in an undergraduate student in Brazil amounts to 109,081.66 BRL[3] (US 21,813.35) on average over the theoretical duration of his/her undergraduate studies (four years) (Inep, 2015), not to mention family expenses, prior expenditures on basic education, and the opportunity cost of staying out of the labor market. Nearly 1.2 million students (which accounts for 5% of the traditional college age population) graduated from an undergraduate program in Brazil in 2017 (Inep, 2018), from which it follows that direct public investment in higher education currently represents approximately 1% of the Brazilian GDP.[4] Indeed, understanding the effectiveness of such investment seems to be important for several stakeholders.

Undergraduate programs in the fields of Business, Law and Social Sciences are the most demanded programs in Brazil, representing 30.9% of all enrollments in higher education. Undergraduate programs in Accounting make up 13.8% of this group, with 359,840 enrollments in 1,489 higher education institutions throughout the country in 2018

[3] This number correspond to a 2015 estimate (i.e., 92,861 BRL) adjusted for inflation based on the Extended Consumer Price Index (IPCA) as of September 2019, with an update factor of 1.1747.

[4] Brazil's gross domestic product (GDP) was 6.8 trillion BRL in 2018 (IBGE, 2019). Total government spending on basic education (kindergarten, primary education, and secondary education) was 4.2% of the GDP, and total spending on higher education was 1.4% of the GDP (OECD, 2019).

(INEP, 2019a). Following the national distribution of undergraduate programs, most of them are located in the South and Southeast, i.e., in the most developed regions in the country.

Given such significant growth of higher education in Brazil, the assessment of education quality is relevant not only for national authorities, who need to follow up on the outcomes of their policies, but also for other stakeholders, including the civil society concerned with its own academic education. In this context, the Sinaes was created in 2004 to "ensure a national process for assessing higher education institutions, undergraduate programs, and student achievement"[5] (Brasil, 2004).

The Sinaes operates with three assessment dimensions: institutional assessment, which may be performed by internal or external audit depending on the specificities of the HEI; undergraduate program assessment; and student assessment. Each dimension holds its own quality indicators: the General Index of Programs (IGC – Índice Geral de Cursos) for institutional assessment, the Preliminary Program Quality Level for undergraduate program assessment, and the Enade score and the IDD for student assessment.

Such quality measures are important because they support decisions in educational funding, which plays an important role in improving the quality of educational institutions. The current federal government's model assumes that rates equal to or higher than 3, on a scale from 1 to 5, are indicative of quality while rates below 3 denote poor quality and prompt an on-site audit by a panel of experts from the National Commission for Assessment of Higher Education (CONAES) to examine the undergraduate program and set forth improvement measures before revoking its authorization if it fails to comply with the minimum requirements.

The IDD is the indicator that has undergone major changes in its estimation methodology over the years. This has consequently affected the CPC, a measure composed of other quality indicators, including the IDD itself.

1.4 Contributions and Delimitations

Value-added measures are useful predictors of quality for educational institutions and can be improved by incorporating further measures of institution-specific characteristics, thereby contributing to more accurate predictions of their impact on test results (Chetty et al., 2014). The VAMs have two major advantages over other (e.g., Status model, Cohort-to-cohort change model, Growth model) models: 1) they provide a fairer estimate of each

[5] The author's translation for: "*assegurar processo nacional de avaliação das instituições de educação superior, dos cursos de graduação e do desempenho acadêmico de seus estudantes*".

institution's contribution to its students' progress, as they follow up on students over time and include their achievements before admission; and 2) they provide a more accurate estimate of this contribution, as they incorporate contextual characteristics of both students and institutions (Kim & Lalancette, 2013). Isolating the school effect on student achievement is known to require at least two achievement test results (in different years), as well as student- and school-related variables (e.g., income, origin, and school excellence) when these change significantly across the different schools (Kim & Lalancette, 2013; Liu, 2011a; Melguizo & Wainer, 2015).

This dissertation aims to add to the debate about the quality of undergraduate programs in Accounting in Brazil by probing into how different VAMs measure the contribution of undergraduate programs to student achievement and identifying their practical implications for policy making. It also aims to contribute to the literature by spelling out the impacts of educational policies on the methodological decisions made by researchers. Its underlying thesis is that by neglecting both the student's personal characteristics and the institutional and learning characteristics that are significant for academic achievement, the IDD proves to be an inaccurate measure of quality and, therefore, should not be used to rate and rank programs.

As mentioned above, Inep has already used four different models to estimate the value added by undergraduate programs in Brazil. The first method was employed from 2006 to 2010 and considered the Enade scores of freshmen and seniors, ratios of students whose parents had higher education, and institutional characteristics. The second method was used in 2011 and 2012, while the third model was applied in 2013; both relied on the Enade scores of seniors and the Enem scores of freshmen in a given year, in addition to institutional characteristics and ratios of students whose parents had higher education. The fourth model, in place since 2014, estimates the value added by an undergraduate program based on the scores its students had attained for admission (the Enem score) and the Enade scores these same students obtained by the end of their undergraduate studies.

The study is limited to undergraduate programs in Accounting in Brazil and based on public databases made available by Inep, namely: Enade microdata, CPC microdata, and IDD microdata. All databases are from year 2015 and correspond to the latest data available before the writing of this dissertation. A total of 30,668 students from 911 undergraduate programs in Accounting were included in the sample, i.e., 46.98% of the total population of accounting students.

1.5 Dissertation Outline

This dissertation is composed of seven chapters, including this Introduction. Chapter 2 reviews the theoretical framework that underpins this research. It addresses the characteristics of higher education in Brazil, as well as the Sinaes and its quality indicators. Emphasis is placed on the historical evolution of the IDD, including its concepts and purposes according to the Technical Bulletins released by Inep. It also discusses the education production function and value-added models, with a focus on research in higher education and criticisms of the model.

Chapter 3 describes the methodological procedures used in the present study, including its research design, sampling, data collection, and data analysis. Chapter 4 reports the data and critically analyzes the results of comparing the IDD models used by Inep. Chapter 5 introduces a new value-added model for estimating the IDD in Brazil and discusses its advantages and disadvantages compared to the current model.

As already mentioned the concept of quality in education is controversial and may involve different perspectives. A discussion about the different visions of quality in higher education seems to be timely as the author of this dissertation herself carried out part of her doctoral studies in the United States (US) and experienced a different learning model and a different quality assessment system from those used in Brazil. As such, chapter 6 is based on the assumption that further dimensions can be added to gain a deeper understanding of educational quality from a value-added perspective.

Chapter 7 concludes this dissertation with final remarks about its key findings and limitations. It also includes suggestions for further research.

2 THEORETICAL FRAMEWORK

2.1 Higher Education in Brazil

Studying higher education in Brazil requires a historical perspective and a deep understanding of its structure and its current conditions of supply and demand. In addition, it requires a discussion of its system for assessment of higher education and how the quality of higher education is comprehended within the national education system. This section addresses such topics, with emphasis on specific aspects of undergraduate programs in Accounting.

Education is a social right provided for in Article 6 of the Federal Constitution. The Federation, Federated States, the Federal District and Municipalities are responsible for legislating on guidelines and frameworks for education (Art. 22) and providing the means for the Brazilian population to have access to education (Art. 23). Chapter III in the Constitution is devoted to the education system and clearly states in its Art. 208 that the government shall provide basic education to all, while higher levels of education should be attained according to the capacity of each individual (Brasil, 1988). In other words, the state is tasked with providing education to all citizens until they complete high school, while higher education is aimed at a specific part of the population.

Education in Brazil, which is also provided by private enterprises, is the duty of both families and the government. It is aimed at "the individual's full development, preparation for exercising citizenship, and training for work"[6] (Brasil, 1988, Art 105). A minimum allocation of resources is constitutionally assured for the maintenance and development of education in Brazil (Art. 212 and 213), with the Federation and other government units allocating 18% and 25%, respectively, of their tax revenues to the education system.

Higher education institutions in Brazil are supposed to follow the principle of inseparability between teaching, research, extension and outreach; they also enjoy teaching, scientific, administrative, financial and managerial autonomy (Brasil, 1988, Art. 207; Brasil, 1996, Art. 52). However, they must comply with general norms of national education defined by the Ministry of Education and submit to authorization protocols and quality assessments

[6] The author's translation for: "*o pleno desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho*".

by the state (Brasil, 1988, Art. 209), which has been the responsibility of the CONAES (Brasil, 2004).

Act No. 9,394, as of 1996, defines the guidelines and frameworks for national education. It sets forth in its Art. 43 the purpose of higher education in Brazil:

> I - spur cultural creation and development of scientific spirit and reflective thinking;
> II - train students in different areas of knowledge to engage in professional fields and in the development of the Brazilian society, as well as collaborate to their continuous education;
> III - encourage scientific research and investigation, aiming at the development of science and technology and the creation and propagation of culture, thereby developing a deeper understanding of both the mankind and the environment;
> IV - promote the dissemination of cultural, scientific and technical knowledge that form the human heritage and pass on knowledge through teaching, publications, or other forms of communication;
> V - arouse the permanent desire for cultural and professional improvement and provide the means for its materialization while continuously integrating the acquired knowledge into an intellectual structure that systematizes the knowledge of each generation;
> VI - stimulate awareness of the problems in the present world, in particular the national and regional ones, provide specialized services to the community, and establish a reciprocal relationship with it;
> VII - promote extension and outreach for engagement of the general population, aiming at disseminating the achievements and benefits derived from the culture and the scientific and technological research produced in the institution.
> VIII - act in favor of improved basic education for all by training professionals, carrying out pedagogical research, and developing extension and outreach activities that bridge both levels of education. (Brasil, 2004, Art. 43)

Considering the Brazilian GDP, higher education has earned approximately 1% of the total public investment. Most of this 1% investment (70%) is allocated for personnel expenses, including payroll taxes (Inep, 2016). Total social spending on both education and culture grew by 158% from 2002 through 2015, driven mostly by investments in higher education. Increased spending at this level of education has been more significant from 2012 onwards, due both to programs 'Science Without Borders' and 'Restructuring and Expansion of Federal Universities' (REUNI – Reestruturação e Expansão das Universidades Federais) and to investments in distance learning. In addition, the government has also provided incentives to expand the private sector through programs 'Financing Higher Education' (Fies) and 'University for All' (Prouni) (BRASIL, 2016).

The main objective of the MEC policy is to provide democratic access to higher education in Brazil, as universities are considered to play a strategic role in the country's economic and social development. Such a policy is expected to produce a positive impact on reducing social and regional inequalities, as well as contribute to "scientific and technological

development, social inclusion, and generation of work and income"[7] (MEC, 2016). However, data from the Higher Education Census (2019) show that increased access to higher education has not been followed by significant changes in social, economic and school admission inequalities. While schooling rates continue to rise among women, the wealthy, white people, and residents in the South and Southeast, the rate of 18-24-year-old Brazilians pursuing further studies is still low, with only 17.6% of this population attending higher education (Inep, 2019).

Higher education is provided by five types of institutions in Brazil, most of which are private organizations (88.2% of total institutions in 2018): universities (7.8% of total institutions in 2018), university centers (9.1%), colleges (81.5%), higher education institutes, and technology education centers (1.6%) (Brasil, 2019). A student can choose from four types of learning programs: 'sequential' programs (which offer short-term technological training), undergraduate programs (which provide bachelor's degrees and teacher education), graduate programs (divided into lato sensu, which include diplomas and MBAs; and stricto sensu, which award master's and doctoral degrees), and extension courses (which meet requirements established by each institution).

This dissertation exclusively addresses undergraduate programs that award a bachelor's degree in Accounting. Most of such programs are private (88.85% in 2018) and located in the South or Southeast (80.2% in 2018). Over the last seven years, only 36% of the Accounting graduates have obtained the minimum score to pass the proficiency exam required to practice the profession in Brazil (CFC, 2018).

Educational data also point to overall mean Enade scores lower than 45% of total points in all years it has been applicated (Inep, 2018) and significant numbers of dropouts amongst accounting students. According to the Higher Education Census (Inep, 2019a), a total of 359,840 Accounting students represented as much as 4.3% of all undergraduate students in Brazil in 2018 – 53,271 of whom graduated, whilst 60,628 suspended registration, 108,494 had their registration cancelled, and 3,753 opted for transfer to a different program (Inep, 2019a). This scenario calls upon studies that seek to contribute to a better achievement amongst higher education students, especially accounting undergraduates.

Researchers have searched both for a better understanding of underachievement of students in Brazil and for explanations for some outperforming others. Such studies have provided different accounts of the situation (see Cornacchione et al., 2010; Fernandes &

[7] The author's translation for: "*o desenvolvimento científico e tecnológico, para a inclusão social e para a geração de trabalho e renda*".

Miranda, 2018; Ferreira, 2015; Miranda, 2011; Miranda et al., 2013; Miranda et al., 2015; Rodrigues et al., 2016, Rodrigues et al., 2017; Santos et al., 2009; Santos, 2012; Souza, 2008). More specifically, accounting undergraduates may underperform for several reasons, including:

1) individual behavior characteristics (e.g., lack of motivation, anxiety, stress) (Miranda et al., 2014);
2) personal academic characteristics (e.g., absenteeism, poor study habits, poor prior knowledge) (Cornacchione Jr. et al., 2010; Fernandes & Miranda, 2018; Miranda et al., 2017);
3) HEI infrastructure (e.g., lack of technological resources, poor classroom infrastructure) (Ferreira, 2015; Fernandes et al., 2018);
4) Pedagogical teaching structure (e.g., large number of students in the classrooms, weak educational plan) (Ferreira, 2015);
5) Faculty characteristics (e.g., low qualifications in the academic, professional and/or pedagogical domains) (Miranda, 2011; Vogt et al., 2016).

All such studies seek to identify the optimal quantity and/or quality of inputs that enable students to perform better (Abdullah, 2011; Fernandes et al., 2018; Harrington et al., 2006; Miranda et al., 2015; Nonis & Hudson, 2010). They suggest three main educational inputs: student's individual characteristics, school characteristics, and faculty characteristics.

Similarly, the Sinaes aims to assess higher education quality in Brazil based on three dimensions of analysis: institutions, undergraduate programs, and student achievement. The next section explains the assessment and its quality indicators.

2.1.1 The National Assessment System for Higher Education (Sinaes)

Because of the significant growth of enrollments and undergraduate programs in Brazil over the last decade, social pressure has increasingly demanded accountability, the disclosure of assessment findings and the measurement of the quality resulting from public investments in education. Educational assessments have been prominent in the country since the 1990s but the current system, Sinaes, was only established in 2004 to assess student achievement, quality of undergraduate programs, and quality of higher education institutions.

The assessment has been based on several educational aspects encompassed by the three dimensions of analysis, including teaching, research, extension and outreach, social responsibility, student achievement, institutional management, faculty, and facilities. A

number of instruments are used to assign rates to each of these dimensions and provide an overview of the quality of the higher education programs and institutions in the country (Brasil, 2004).

The institutional assessment comprises ten sub dimensions, namely: 1) mission and Institutional Development Plan (PDI – Plano de Desenvolvimento Institutional); 2) policies for teaching, research, graduate studies, extension and outreach; 3) social responsibility; 4) communication with society; 5) personnel policies, faculty and staff careers; 6) management organization; 7) physical infrastructure; 8) assessment planning; 9) student service policies; 10) financial sustainability (Inep, 2018). It is based on two types of assessment: 1) self-assessment, which is coordinated by an HEI-run Assessment Committee (CPA – Comissão Própria de Avaliação) and must comply with the CONAES standards; and 2) external assessment, which is carried out by an Inep-appointed commission.

The assessment of undergraduate and graduate programs involves three sub dimensions, namely: pedagogical teaching structure, faculty profile, and physical facilities. All programs in the country undergo three different stages of assessment: 1) one for authorization (before starting a new program), one for recognition (when the first class of the new program is halfway to graduate), and one re-assessment for renewal of recognition (repeated every three years upon first recognition). The assessment of graduate programs has been performed successfully (from a methodological perspective) by the Coordination for the Improvement of Higher Education Personnel (CAPES) (Pedrosa et al., 2013). The graduate studies, however, are not the focus of this dissertation.

The assessment of undergraduate student achievement is performed through the Enade, an exam which "assesses the achievement of undergraduate students against the syllabus, skills and competences acquired in their study track"[8] (Inep, 2018a). The Enade is a mandatory exam, with (non-)attendance noted on the student's transcript of records. Held every three years by area of knowledge, it assesses the students' proficiency in field-specific topics as determined by the National Curriculum Guidelines (DCN – Diretrizes Curriculares Nacionais) (Pedrosa et al., 2013). Table 1 shows the assessment instruments used in the Sinaes.

[8] The author's translation for: "*avalia o rendimento dos alunos concluintes dos cursos de graduação, em relação aos conteúdos programáticos, habilidades e competências adquiridas em sua formação*".

Table 1. Sinaes assessment instruments

Type of Assessment	Purpose and frequency	Agency in charge	Variables used
Self-assessment	- Authorization and recognition: one time only; produces quality indicator Program Quality Level (CC – Conceito de Curso). - Accreditation of institutions and government schools ("escolas de governo" – aimed at training, improving and professionalizing civil servants in policies and management): only once - Re-assessment: every three years or whenever a quality indicator ranks below 3.	CPA (HEI-run Assessment Committee) - Sinaes pool of raters - Pool of raters for government schools - Technical Committee for Assessment Follow-up (CTAA – Comissão Técnica de Acompanhamento da Avaliação)	1. Mission and institutional development plan; 2. Policies for teaching, research, graduate, extension and outreach; 3. Social responsibility; 4. Communication with the society; 5. Personnel policies, faculty and staff careers; 6. Organization and management, representativeness of academic boards, and engagement of the university community in decision-making processes; 7. Physical infrastructure; 8. Process planning and assessment; 9. Student service policies; 10. Financial sustainability.
External assessment			
Undergraduate program assessment		CPA, External raters	1. Pedagogical teaching structure 2. Physical facilities; 3. Faculty profile.
Undergraduate student achievement	Every year for each area of knowledge, in a three-year assessment cycle[1]	Inep	1. Exam involving topics from the undergraduate program syllabus as established in the DCN and topics related to current state of affairs in Brazil and worldwide. 2. Student's socioeconomic profile questionnaire. 3. Questionnaire about the students' perception of the exam and their education process in both the program and the HEI. 4. Program coordinator questionnaire

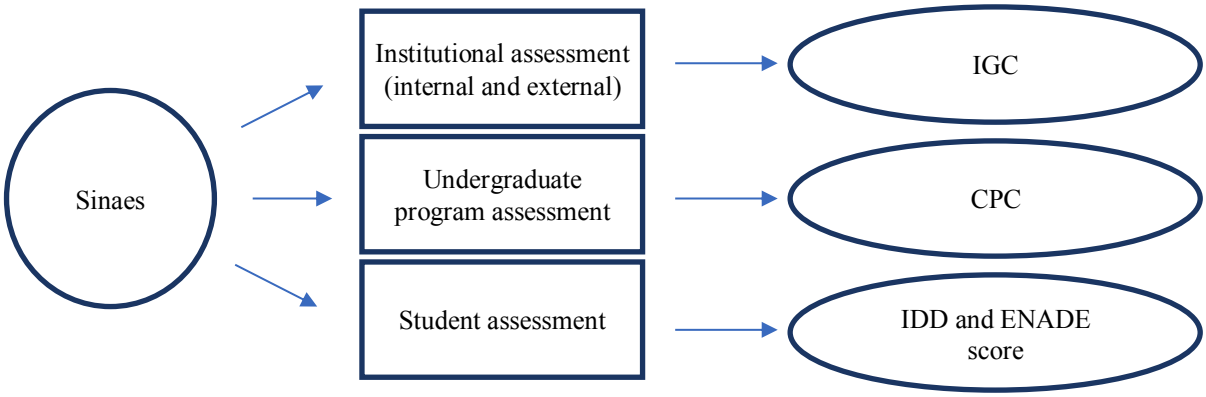
(1) The area of Applied Social Sciences is assessed in year 3, i.e., as of the writing of this doctoral dissertation, the latest assessment of undergraduate programs in Accounting took place in 2018.
Source: based on Technical Bulletin INEP/DAES/CONAES No. 65, as of 9 October 2014 (MEC, 2014); Administrative Rule No. 840, as of 24 August 2018 (MEC, 2018); Act No. 10861, as of 14 April 2014 (Brasil, 2014).

Based on all these assessment instruments, the government releases indicators that 1) measure the quality of the Brazilian undergraduate programs and higher education institutions and 2) support the development of policies for higher education. All data are public and retrievable from the Inep website.

2.1.2 Quality Indicators in the Brazilian Higher Education

Figure 1 shows how the Sinaes instruments are related to the quality indicators used in (re)assessment processes.

Figure 1. Sinaes quality indicators



Source: the author.

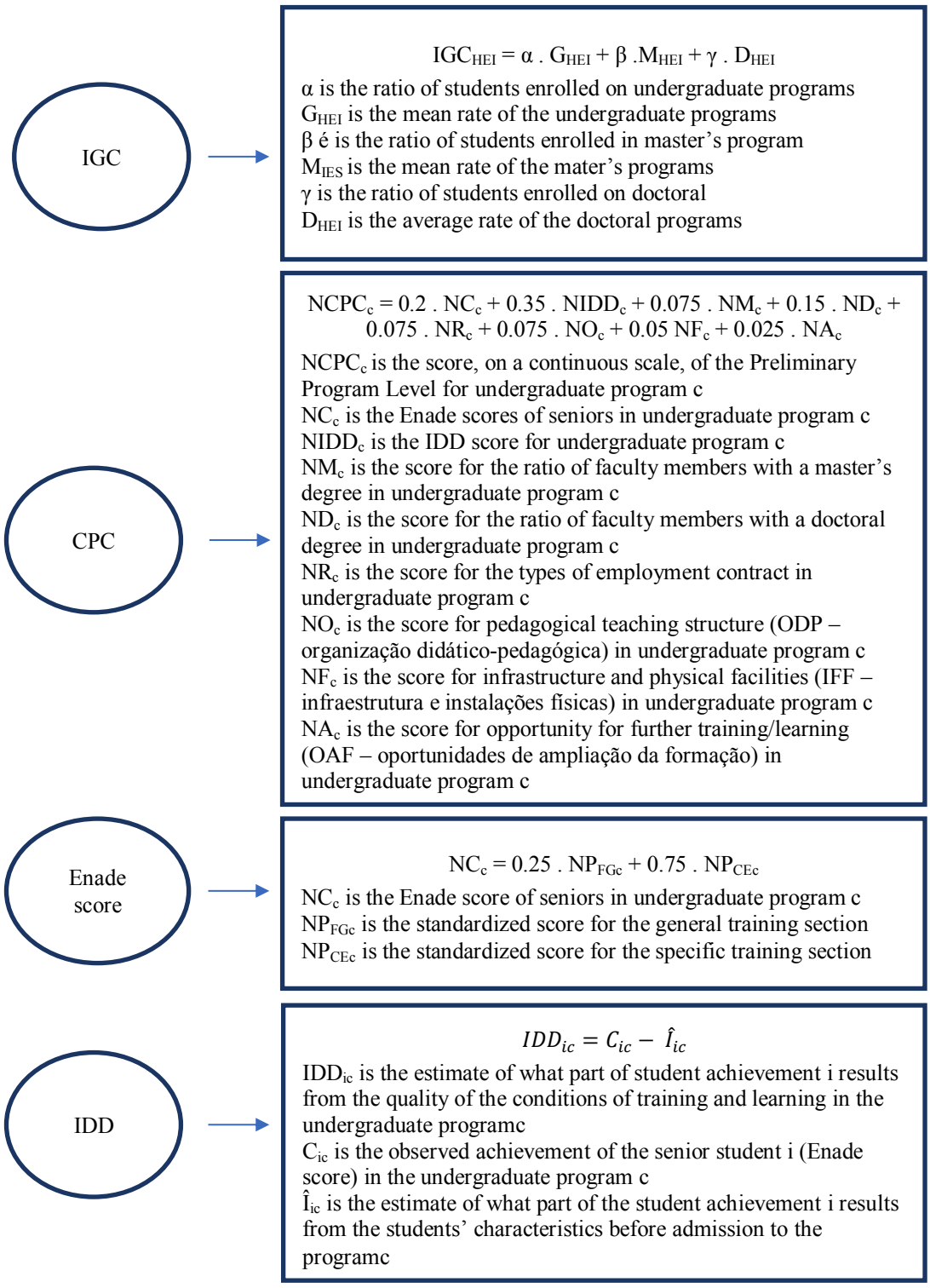
The quality indicators are expressed on a continuous scale from 1 to 5, with levels above 3 indicating satisfactory quality. There are four major indicators:

1) Enade score, regulated by Administrative Rule No. 40, as of 12 December 2007: it assesses the undergraduate programs based on the students' results in the exam considering 1) the number of participating students, 2) the senior's achievement in the 'general training' section, and 3) the senior's achievement in the 'specific training' section;
2) Indicator of Difference between Observed and Expected Achievements (IDD), established by Administrative Rule No. 4, as of 5 August 2008: it seeks to estimate the value added by the undergraduate program to the senior's development building on 1) their achievement on the Enade and 2) their development characteristics before admission to the program;
3) Preliminary Program Quality Level (CPC), established by Administrative Rule No. 4, as of 5 August 2008: it is estimated in the year after the Enade and is based on 1) student achievement, 2) value added by the training process provided by the program, 3) faculty, and 4) students' perception of their training/learning conditions;
4) Higher Education Institution's General Index of Programs (IGC), created by Administrative Rule No. 12, as of 5 September 2008: it is provided annually based on 1) the mean of the CPCs of the HEI's programs in the preceding

three years weighted by the number of students enrolled on each of them, 2) the mean of CAPES rates for the graduate programs in the preceding three years weighted by the number of students enrolled on each of them, and 3) student distribution in the different levels of education, excluding item 2) for those HEIs that do not feature graduate programs.

Figure 2 shows the current methodology (as of 2019) for estimating each quality indicator.

Figure 2. Current methodology for estimating the Sinaes quality indicators (as of 2019)



Source: based on Technical Bulletins No. 16/2018/CGCQES/DAES (Inep, 2018), No. 17/2018 CGCQES/DAES (Inep, 2018a), No. 18/2018/CGCQES/DAES (Inep, 2018b), and No. 19/2018/CGCQES/DAES (Inep, 2018c).

Two of the quality indicators deserve especial attention for the purposes of this doctoral dissertation: the IDD and the CPC. As mentioned in the Introduction, the IDD is "a quality indicator that measures the value that an undergraduate program adds to the

development of its seniors by probing their achievements on the Enade as compared to their developmental characteristics at the beginning of their study track" (Inep, 2017b, p. 1). It is the difference between the students' observed Enade scores and predicted scores based on their admission scores – to estimate the achievements resulting from the students' characteristics before higher education, the estimation includes their scores on the Enem, an exam taken by the end of high school and widely used for admission to the Brazilian undergraduate programs. In other words, this indicator aims to quantify how much each higher education institution adds to their students' achievements during their undergraduate studies. Its relevance for Program Quality Level is borne out through the 35% share that it contributes to the CPC (see Figure 2).

In turn, the CPC is a weighted sum of means related to student achievement (i.e., IDD and Enade score), faculty characteristics (i.e., type of employment contract, and ratio of faculty members with a doctoral degree and a master's degree that work in the program), and the program structure in the students' perception (OAF, IFF, and ODP). The rates for student achievement are based on large-scale knowledge exams, namely the Enade and the Enem. Data for the faculty characteristics are obtained from the Higher Education Census and the Institutional Assessment (Internal and External). The standardized scores for program structure (OAF, IFF, and ODP) are averaged from the students' answers to items related to their perception of their training/learning conditions. This means that 15% of the rate awarded to the undergraduate programs is impacted by the student's perception of these characteristics.

As such, the CPC combines different quality indicators. It is an important rate because it streamlines on-site assessments, which would be impossible to perform in every program every year due to the significant number of undergraduate programs (Fernandes et al., 2009). As programs with a CPC above 2 are deemed to meet the minimum operating conditions, they are all exempt from on-site assessments.

Since its inception in 2006, the Sinaes quality indicators have undergone several changes aimed at improvement, as reported in the MEC Technical Bulletins (Inep, 2009, 2012, 2015). To understand how such changes have taken place, the next section – based on reading and analyzing the Technical Bulletins available on the Inep website – pinpoints the main changes that have been made to estimate both the IDD and the CPC since 2006.

2.1.3 Historical Evolution of the IDD and CPC Indicators

The Inep website (Inep, 2019b) provides all documentation and law regarding quality indicators for higher education that have been produced in Brazil since 2007. Reading all Technical Bulletins addressing the CPC and the IDD allows for identifying all changes in the methodology for estimating both indicators.

Table 2. Variables used to estimate the CPC in 2006-2010

Variables	Description	Estimation
Faculty members with a doctoral degree (Doc)	Ratio of the number of all faculty members with a doctoral degree in program i (DOC) by the total number of faculty members in program i (PROF)	$Doc = \frac{DOC}{PROF}$
Faculty members with a master's degree (Me)	Ratio of the number of all faculty members with a master's degree or above in program i (ME) by the total number of faculty members in program i (ME)	$Me = \frac{ME}{PROF}$
Employment contracts (RT)	Ratio of the faculty members with part-time or full-time employment contracts in program i (PR) divided by the total number of faculty members in program i (PROF)	$RT = \frac{PR}{PROF}$
Infrastructure (IFF)	Ratio between 1) the number of students in program i who answered "Yes, in all of them" or "Yes, in most of them" (SIM1) to question "Practical classes: is the equipment available enough for all students?" and 2) the total number of students who answered this question (NA)	$IFF = \frac{SIM1}{NA}$
Pedagogical teaching structure (ODP)	Ratio between 1) the number of students in program i who answered "Yes, all of them do" or "Yes, most of them do" (SIM1) to question "Do the teaching plan for each course contain all the following features: objectives, teaching and assessment procedures, syllabus, and bibliography?" and 2) the total number of students who answered this question (NA)	$ODP = \frac{SIM2}{NA}$
Seniors' Enade score (NC)	Weighted mean of Enade scores (as detailed in Figure 2) of all Seniors in program i who took the exam (N1)	$NC = \frac{\sum_{n=1}^{N} Enade\ score}{N1}$
Freshmen's Enade score (I)	Mean of Enade scores (as detailed in Figure 2) of all Freshmen in program i who took the exam (N2)	$I = \frac{\sum_{n=1}^{N} Enade\ score}{N2}$
Indicator of Difference between Expected and Observed Achievements (IDD)	Difference between the Enade scores of the seniors in program i (c) and the average score expected for them given the existing data about the profile of the freshmen in said program ($\hat{c}$)	$IDD = C - \hat{C}$ $c = \beta i + \gamma w + \delta z + \varphi Doc + \lambda Me + \eta RT + \rho IFF + \upsilon ODP + e$ $\hat{c} = \hat{\beta} i + \hat{\gamma} w + \hat{\delta} z$

Source: based on Inep (2009a).

The IDD had its first standalone Technical Bulletin released in 2011, as its estimation procedure used to be described in the Technical Bulletin provided for the CPC. The first CPC Technical Bulletin available on the Inep website, as of 17 of December 2009, defines the estimation methodology for both CPC and IDD in 2008. Table 2 shows their variables and equation.

Before estimating the CPC, all variables in Table 2 were standardized and transformed into a new scale. Standardization was performed as follows:

$$AP_x = \frac{x - \bar{x}}{DP_x} \quad (1)$$

where: AP is the standard deviation of quality measure x in program i, x is the quality measure of program i, $\bar{x}$ is the mean of this measure in area j, and SD is the standard deviation of measure x in area j. Transformation into a scale from 0 (poor) to 5 (excellent) was performed as follows:

$$Nx = 5 * \frac{AP_{x+|AP_x inferior|}}{|AP_x superior| + |AP_x inferior|} \quad (2)$$

where: 1) the lower and upper APs in measure x are the standard deviation for area j, rather than for program i, and 2) the lower and upper APs greater than $|3|$ are outliers and excluded from the indicator estimation. Hence, all variables used to estimate the CPC are provided in a continuous scale from 0 to 5.

However, no program can be rated level 5 if any CPC component scores 0.9 or lower. In this case, the program is rated level 4. No changes have been applied to this rule and to the methodological procedures of standardization and scale transformation.

The estimation of the IDD until 2010 included the Enade scores of both freshmen and seniors, as both groups were required to take the exam until that year. Two multiple linear regression equations used to be performed to estimate the IDD. Equation 1 estimated the parameters of each variable:

$$c = \beta i + \gamma w + \delta z + \varphi Doc + \lambda Me + \eta RT + \rho IFF + \upsilon ODP + e \quad (3)$$

where: c is the senior's Enade score in a given year; i is the freshman's Enade score in the same year; w is the ratio of students whose parents had higher education; z is the ratio of seniors and freshmen; *Doc* is the ratio of faculty members with a doctoral degree; *Me* is the ratio of faculty members with a master's degree; *RT* is the type of employment contract; *IFF*

is the rate for pedagogical teaching structure; and *e* is the random error of the equation (Inep, 2009a).

After estimating the regression parameters, the Enade score ($\hat{c}$) was calculated, and the standardized residual was estimated to exclude all those values greater than | 3 |. Once the outliers were excluded, the regression equation was run a second time to find the final parameter values. Such parameter estimates were used for the final calculation of the estimated Enade score ($\hat{c}$), as in equation 2. The IDD was calculated as the average difference between the actual Enade score (c) and the estimated score ($\hat{c}$) of all students of a specific program.

$$\hat{c} = \hat{\beta} i + \hat{\gamma} w + \hat{\delta} z \quad (4)$$

Using student characteristics (w) and program characteristics (z) to estimate student scores was based on the following reasoning:

> It is known that the difference between the Enade achievements of students graduating from two undergraduate programs from two different institutions does not depend only on the differences in quality between such programs. Differences in student profile before admission to both programs also influence differences in their outcomes.[9] (Inep, 2009a, p. 6)

The Brazilian government acknowledged that estimating how much a program contributes to student achievement (which is the value added – the IDD) would require an account of the characteristics of both the students and the institution if it were to avoid an underestimation or overestimation of the value added. Therefore, the justification for including the institutional variables in "*c*" was to reduce the estimation bias of the coefficients associated with the programs, since better educated students would more likely be admitted into better quality programs (Inep, 2009a). After estimating all variables in Table 2, the CPC was computed from 2006 to 2010 by weighting these variables, as shown in Table 3.

Table 3. CPC estimation in years 2006-2010

Variable	Weight	CPC Equation
Doc	0.20	$\text{CPC} = 0{,}2 * \text{Doc} + 0{,}05 * \text{Me} + 0{,}05 * \text{RT} + 0{,}05 * \text{IFF} + 0{,}05 * \text{ODP} + 0{,}30 * \text{IDD} + 0{,}15 * \text{I} + 0{,}15 * \text{C}$ (35
Me	0.05	
RT	0.05	
IFF	0.05	
ODP	0.05	
IDD	0.30	
I	0.15	
C	0.15	

Source: based on Brasil (2009a).

[9] The author's translation to: "*Sabe-se que a diferença entre os desempenhos no Enade de alunos egressos de dois cursos de graduação de duas instituições distintas não depende somente das diferenças de qualidade entre esses cursos. As diferenças em relação ao perfil dos alunos de ambos os cursos ao ingressar no ensino superior também influenciam as diferenças nos resultados*."

Finally, the programs had their continuous rates for CPC (estimated from equation 3 in Table 3) transformed into one of five levels. Level 1 was assigned to programs with CPCs lower than or equal to 0.94; Level 2 to programs with CPCs from 0.95 to 1.94, inclusive; Level 3 to programs with CPCs from 1.95 to 2.94, inclusive; Level 4 to programs with CPCs from 2.95 to 3.94, inclusive; and Level 5 to programs with CPCs above 3.95.

Significant changes occurred in the estimation of CPC from 2011 onwards influenced by Fernandes et al. (2009), who were the former Inep president, director for educational studies, coordinator, and general coordinator for educational instruments and measures, respectively. Their study aimed to analyze the purpose, quality and consistency of quality indicators IDD and CPC while also providing a new, "ideal" estimation methodology that would show the actual contribution of the undergraduate programs to their students' learning (Fernandes et al., 2009, p 5).

In analyzing the former IDD estimation methodology, Fernandes et al. (2009) detected significant shortcomings in this indicator – some related to the use of standardized exams to measure knowledge and skills for professional practice, and others related to the estimation methodology itself. First, the graduates' academic quality could vary across programs, across areas of knowledge and even across Enade editions, since new, different questions are used every year. Second, the IDD varied greatly across the institutions and its estimation as a regression residual was significantly biased against the educational institutions that attract good students, i.e., "if the quality of a program is positively correlated with the scores of its freshmen, the IDD underestimates the quality of those program that attract students with prior good achievements"[10] (Fernandes et al., 2009, p. 7).

To overcome this bias, the authors suggest that estimating the value added (IDD) should include the freshmen's achievement, as well as institution- and student-related variables, as in equation 4:

$$c_{ij} - i_{ij} = q_j + bi_{ji} + x_{ij}\varphi + v_{ij} \quad (6)$$

where: *j* refers to the institution and *i* to the student, *ç* is the achievement score by the end of the study track, *i* is the achievement score before admission, *q* is the set of institutional characteristics, *x* is the set of student characteristics, and *v* is the random error (relative to both student and program).

[10] The author's. translation for: "*se a qualidade do curso está positivamente correlacionada com as notas dos ingressantes, o IDD subestima a qualidade dos cursos que recebem alunos com bom desempenho*".

Introducing i_{ij} onto the right side of the equation means that "adding value depends on the freshmen's achievement"[11] (Fernandes et al., 2009). Thus, if b is positive, it follows that it is more difficult to add value to students with high prior achievement, while the opposite holds true if b is negative. Accordingly, the equation to determine the IDD should be:

$$c_j = (1 + b)i_j + q_j + x_j\varphi + v_j \quad (7)$$

The authors also point to need to use institutional variables (q) that are significant for the quality differences across institutions, such as whether the institution is public or private. Public institutions generally attract better students than the private HEIs in Brazil, as their selection process tend to be more rigorous and their infrastructure tend to be of higher quality. Furthermore, the public programs do not require tuition fees or any other fees from the students and their families.

Zoghbi, Moriconi and Oliva (2010), too, argue that using the Enem score was a more effective and efficient procedure for estimating the IDD than the freshmen's Enade scores. Efficiency, in this case, result from the reduced cost of using the Enem scores and not having the freshmen take the Enade. Effectiveness results from the students' strong incentive to take the Enem (as it was already part of the admission process to most undergraduate programs in the country), which tend to provide a better quality indicator for prior achievement.

As of 2011, the freshmen were dismissed from taking the Enade, and their Enem scores started being used to estimate the CPC. All other variables in the IDD equation remained unchanged, except for the IFF and the ODP, which were slightly adapted (see Table 4).

A requisite for estimating the IDD in 2011 and 2012 was that a program had at least 10 freshmen who took the Enem and 10 seniors who participated in the Enade, all of whom with scores above zero. Another requisite in 2011 was a minimum rate of 20% of the students from a program taking the exam. Because of such restrictions, the IDD was not estimated for some fields tested on the Enade and was replaced by the seniors' scores (NC) for computing the CPC (INEP, 2012, 2013).

[11] The author's translation to: "*agregar valor depende do desempenho dos ingressantes*".

Table 4. Variables with new estimation methodology in years 2011-2012

Variable	Description	Equation
Infrastructure and physical facilities (IFF)	Weighted ratio between 1) the number of students in program i who answered "Yes, all of them" (SIM1), "Yes, most of them" (SIM2), or "Only some" (SIM3) to question "Do the equipment and/or materials available in the rooms for practical classes suffice for the number of students? (If you are a distance learning student, please consider the conditions of the physical learning support center)" and the total number of students who answered the question (NA)	$IFF = \frac{\sum SIM1 * 1}{NA} + \frac{\sum SIM2 * 0{,}5}{NA} + \frac{\sum SIM3 * 0{,}5}{NA}$
Pedagogical teaching structure (ODP)	Weighted ratio between 1) the number of students in program i who answered "Yes, all features" (SIM1), "Yes, most features" (SIM2), or "Only some features" (SIM3) to question "Do the teaching plan for each course contain all the following features: objectives, teaching and assessment procedures, syllabus, and bibliography?" and 2) the total number of students who answered this question (NA)	$ODP = \frac{\sum SIM1 * 1}{NA} + \frac{\sum SIM2 * 0{,}5}{NA} + \frac{\sum SIM3 * 0{,}5}{NA}$
Indicator of Difference between Expected and Observed Achievement (IDD)	The same linear regression equation of previous years, with the freshmen's mean Enade score replaced with the students' mean in the four Enem sections (Natural Sciences, Humanities, Language, and Mathematics).	$IDD = C - \hat{C}$ $c = \beta\overline{Enem} + \gamma w + \delta z + \varphi doc + \lambda me + \eta rt + \rho iff + vodp + \epsilon$ $\hat{c} = \hat{\beta}\overline{Enem} + \hat{\gamma} w + \hat{\delta} z$
Freshmen's Enade score (I)	It has no longer been estimated as the freshmen have been dismissed from Enade as of 2011, according to the 2010 version of Administrative Rule MEC No. 40/2007. This measure was replaced by the students' mean in the four Enem sections.	

Source: based on INEP, 2012, 2013.

Inep also changed the weights of the variables that make up the CPC, as variable I (freshmen's Enade scores) had been excluded from the model. Table 5 shows the final structure of the CPC in 2011 and 2012.

Table 5. CPC estimation in years 2011-2012

Variable	Weight	CPC Equation
Doc	0.15	$\text{CPC} = 0{,}15 * \text{Doc} + 0{,}075 * \text{Me} + 0{,}075 * \text{RT} + 0{,}075 * \text{IFF} + 0{,}075 * \text{ODP} + 0{,}35 * \text{IDD} + 0{,}20 * \text{C}$ (8)
Me	0.075	
RT	0.075	
IFF	0.075	
ODP	0.075	
IDD	0.35	
C	0.20	

Source: based on BRASIL, 2012, 2013.

Variables Me and RT had their weights increased by 2.5% compared to the previous years, while variable Doc had its weight decreased by 2.5%; therefore, the faculty-related variables continued to represent 30% of the CPC. The program-related measures

(ODP and IFF) had their weights increased by 2.5% and represented 15% of the CPC. The variables related to student achievement (IDD and C) had their weights reduced by 5% to 55% of the CPC (IDD, C and I had corresponded to 30%, 15% and 15%, respectively, in previous years).

Inep used to employ few questions from a student questionnaire to estimate variables ODP and IFF, as shown in Tables 2 and 4. Only in 2012 has Inep started to make a more thorough use of the questionnaire, which consists of over 40 questions about the students' perception of their training/learning conditions in the undergraduate program and about their socioeconomic status. In addition, a third variable related to their perceptions was added to the CPC estimation: opportunity for further training/learning (OAF).

Thus, components OAF, IFF and ODP started being estimated by using the mean of a broad set of questionnaire responses. This was the only change to the estimation of the continuous CPC as compared to the previous model (see Table 6). Given the introduction of variable OAF with a weight of 0.025, the weight of variable IFF reduced from 0.075 to 0.05. Since then, the CPC estimation has been performed according to equation 7 (see Table 6).

Table 6. CPC estimation from 2013 to date

Variable	Weight	CPC Equation
Doc	0.15	
Me	0.075	
RT	0.075	
IFF	0.05	$\text{CPC} = 0{,}15 * \text{Doc} + 0{,}075 * \text{Me} + 0{,}075 * \text{RT} + 0{,}05 * \text{IFF} +$ (9)
ODP	0.075	$0{,}075 * \text{ODP} + 0{,}025 * \text{OAF} + 0{,}35 * \text{IDD} + 0{,}20 * \text{C}$
OAF	0.025	
IDD	0.35	
C	0.20	

Source: based on INEP, 2014, 2015, 2016, 2017, 2018, 2019.

Some changes were also applied to the CPC Levels 1-5, now defined as in Table 7.

Table 7. CPCs Level from 2013 to date

CPC (Level)	CPC (Continuous)
1	$0 \leq \text{CPC} < 0.945$
2	$0.945 \leq \text{CPC} < 1.945$
3	$1.945 \leq \text{CPC} < 2.945$
4	$2.945 \leq \text{CPC} < 3.945$
5	$3.945 \leq \text{CPC} \leq 5$

Source: INEP, 2014, p. 20.

The quality indicator IDD also had its estimation changed because of variable OAF. This variable was added to the regression equation "c" (see Table 4), leading to the following IDD estimation in 2013:

$$IDD = C - \hat{C} \tag{10}$$

$$c = \beta\overline{Enem} + \gamma w + \delta z + \varphi doc + \lambda me + \eta rt + \rho iff + \upsilon odp + \mu oaf + \epsilon \tag{11}$$

$$\hat{c} = \hat{\beta}\overline{Enem} + \hat{\gamma} w + \hat{\delta} z \tag{12}$$

The latest change to the CPC estimation took place in 2013. In contrast, the IDD had one final significant change in 2014, influenced by Zoghbi et al. (2010). Consequently, the IDD estimation started to include both the Enade and the Enem scores of the same student, with the retrieval of their data from the Inep databases.

Until 2013, the IDD estimation was based on the seniors' Enade scores in a given year and the freshmen's Enem scores in said year. As the Enem scores had been released since 2009, the seniors in 2014 already had their Enem scores recorded in the Inep database. As a result, the IDD started to "estimate the value added by the undergraduate program to the seniors' development considering their achievement on the Enade and their development characteristics before admission to the program"[12] as measured through their Enem scores (Inep, 2015, p. 6).

The new estimation methodology came to include a multilevel hierarchical regression analysis, and the characteristics of both students and programs were excluded from the new model. The IDD is now estimated as follows:

$$IDD_{ij} = C_{ij} - \hat{I}_{ij} \tag{13}$$

where: the IDD is the estimate of the part of the student achievement resulting from the quality of the training/learning conditions provided by the undergraduate program.

The IDD estimation employs two-level hierarchical linear modeling. One level is that of the student, estimated through:

$$C_{ij} = \beta_{0j} + \beta_{1j} * CN_{ij} + \beta_{2j} * CH_{ij} + \beta_{3j} * LC_{ij} + \beta_{4j} * MT_{ij} + \lambda_{ij} \tag{14}$$

[12] The author's translation for: "*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*".

where: C_{ij} is an achievement estimate for senior student *i* on the Enade as weighted by his/her scores in the specific training section (75%) and in the general training section (25%) for the undergraduate program *j*; CN_{ij} is the measure of achievement in Enem section 'Natural sciences and their technologies' for senior student *i* in undergraduate program *j*; CH_{ij} is the measure of achievement in Enem section 'Humanities and their technologies' for senior student *i* in undergraduate program *j*; LC_{ij} is the measure of achievement in Enem section "Languages, codes, and their technologies" for senior student *i* in undergraduate program *j*; MT_{ij} is the measure of achievement in Enem section "Mathematics and its technologies" for senior student *i* in undergraduate program *j*; λ_{ij} is the random effects associated with senior student *i* in undergraduate program *j*.

The second level of analysis is the program, as estimated through:

$$\beta_{0j} = \beta_{00} + u_{oj} \quad (15)$$

where: β_{00} represents the mean or general intercept, which is constant across the undergraduate programs; and u_{oj} is the random effects associated with undergraduate program *j*.

The multilevel regression model is estimated twice. The first regression extracts the parameters, estimates the standardized residual and excludes those with a modular value higher than 3. The second regression uses the parameter values to produce the estimate Î as in:

$$\hat{I}_{ij} = \hat{\beta}_{0j} + \hat{\beta}_{1j} * CN_{ij} + \hat{\beta}_{2j} * CH_{ij} + \hat{\beta}_{3j} * LC_{ij} + \hat{\beta}_{4j} * MT_{ij} \quad (16)$$

where: $\hat{I}_{ij}$ is the estimate of the part of the Enade achievement of senior student *i* in undergraduate program *j* resulting from the students' characteristics before admission to the program.

A gross IDD_{ij} is estimated for each student i from undergraduate program j, as in equation 11; then, a mean IDD_{ij} is estimated for each program (sum of all IDD_{ij} for program j divided by the number of students from program j). As with the other variables that make up the CPC indicator, the IDD_j score is standardized and transformed into a continuous scale from 1 to 5. The IDD estimation has been disclosed in a specific Technical Bulletin since 2016.

2.1.4 Criticisms of the Sinaes

Student associations and groups of teachers and education specialists have opposed the higher education assessment systems implemented by the federal government since its first version in 1996. The criticisms include: 1) exams applied to students cannot measure university education adequately, 2) students are punished when they do not participate in the mandatory assessment process, 3) assessments have been linked to increasing privatization of higher education in Brazil and are thought to be a form for the government to set the stage to privatize the entire education system, 4) the system reduces or revokes the universities' academic freedom by imposing mandatory teaching content for examination, 5) institutions with little selection potential can be penalized because of the poor academic achievement of students with historically poor educational background (Pedrosa et al., 2013, p. 63).

Such contrary manifestations have resulted in boycotts of the Enade by students, which can affect the overall assessment of both the program and the educational institution. Academic assessments may play a relevant role in the institutional assessment process, but students have seen them with distrust or discomfort (Darwin, 2016). In contrast, private institutions have trained students to take the exam and rewarded faculty members for the students' results. All such situations negatively affect the Sinaes (Pedrosa et al., 2013).

Much as the Brazilian government has made an effort to engage the academic society in its processes for building instruments for higher education assessment, the information produced by the Sinaes reports, with data and characteristics of the educational institutions, have not achieved program coordinators properly enough to improve higher education substantially (Gryboski, 2012). Some of them have been unaware of the existence of online reports (Freitas, 2012). A survey of coordinators of undergraduate programs in Accounting showed that 61% of the sample had read the 2006 Enade report and pointed to a positive association between the use of the report and the program's achievement in the following assessment cycle (Freitas, 2012). A similar result was found in a study on how coordinators of undergraduate programs in Business Administration and Accounting perceived the usefulness of the 2012 Enade reports: the highly rated programs had coordinators with a positive stance toward the quality and usefulness of such reports (Freitas et al., 2015).

Also, the lack of a reference measure has been pointed out as a major reason for not using the reports. Nonetheless, all SINAE quality indicators, estimated on a scale from 1

to 5, rely on a model designed according to government standards, which entails that the rates are relative and dependent on the data of all HEIs, rather than on an expected level of achievement (Pedrosa et al., 2013). In other words, an institution rated 5 cannot be said to have an expected level of excellence, but rather that it is amongst the best institutions when it comes to its undergraduate programs that have been analyzed.

In fact, most criticisms toward the Sinaes are based on 1) its quality assessment of a single program/institution on account of its final rate or 2) the lack of baseline criteria for pinpointing excellence programs/institutions in a given field of studies. However, when it comes to the cumulative data by institution and by program, the Sinaes provides an important contribution to understanding the general state of affairs in higher education in Brazil, indicating areas and sectors that need greater regulation, investment, control and assessment by the government (Pedrosa et al., 2013).

2.1.5 Practical Implications of the Sinaes Quality Indicators

The fact that the data obtained by the Ministry of Education, via Inep, are fully disclosed to the general public has two important consequences for the assessment of higher education in Brazil. One is positive: the process is transparent, which contributes to its credibility. Anyone can freely access the indicators and retrieve the data (Pedrosa et al., 2013). However, such transparency also comes at a price: some HEIs have used their achievement outcomes for advertising purposes, which is actually forbidden, and the media have focused on specific, isolated data, such as the Enade results, to disclose program rankings, which may lead to misunderstandings about the quality of the Brazilian undergraduate programs (Darwin, 2016; Pedrosa et al., 2013).

The quality of higher education programs in Brazil, as provided by the Sinaes quality indicators, is an important topic of interest, and not only because of the practical issues related to student training and future contributions to the society. From a policy making perspective, the quality of undergraduate programs is directly related to their funding. For example, public funding of higher education in private institutions through the Fies is conditional on the quality of the undergraduate programs as assessed through the Sinaes indicators. The Fies was initially regulated through Act No. 10260, as of 12 July 2001, and recently amended through Act No. 13530, as of 7 December 2017. It is “intended for funding undergraduate students from private higher education institutions that obtained positive assessment in processes carried out by the Ministry [of Education] according to its own

regulations"[13] (Brasil, 2017, Art. 1). It is especially targeted at students who have not graduated yet and who have not been awarded any student funding (Art. 1, Paragraph 6).

Article 4 of Administrative Rule No. 209, as of 7 March 2018, regulates the Fies from 2018 onwards and clearly states that eligible programs for public funding are those with at least a satisfactory rating on the Sinaes, i.e., rated 3 or higher in the Program Quality Level (CC), in the Preliminary Program Quality Level (CPC), or on the Enade (Brasil, 2018). The current rule for student funding across different programs is as follows: up to 50% of the places for Level-5 programs, up to 40% for Level 4, up to 30% for Level 3, and up to 25% for programs that have no CC or CPC (Brasil, 2019, Art. 9, Paragraph 4).

Administrative Rule No. 952, as of 2 May 2019, defines the selection process for student financing. The conditions for a candidate to apply for the Fies are: 1) mean Enem score above 450 points, and no zero in written composition; 2) gross monthly household income of up to 3 National Minimum Wage[14] per capita (Fies) or up to 5 National Minimum Wage per capita (P-Fies) (Brasil, 2019). The actual mean income in Brazil was 2,291.00 BRL (corresponding to 2.3 National Minimum Wage) in the first quarter of 2019, as estimated by the Brazilian Institute of Geography and Statistics (IBGE, 2019). The per capita household income was 1,373.00 BRL in 2018 (equivalent to 1.47 National Minimum Wage in the same year). This lays bare that the Fies is relevant for a significant share of the Brazilian population.

Clearly, ranking students on standardized achievement exams, such as the Enem and the Enade, is the cornerstone of public funding for educational institutions, making them strive as much as possible for outstanding student achievement. Yet, national data show that Accounting undergraduates have underperformed in standardized tests, as previously mentioned.

Inep microdata has been used as a source for research based on the Enade score as a response variable and as a proxy for academic achievement, with variables related to the HEIs, undergraduate programs and students being used as explanatory variables of said achievement. However, Brazilian studies have focused on the student's final score (Enade) and neglected the knowledge added over a certain period (e.g., higher education). Monk (1989) argues against focusing on parametric estimates. He states that instead, we should look at broad relationships between key constructs to anticipate results from different choices of

---

[13] The author's translation to: "*destinado à concessão de financiamento a estudantes de cursos superiores não gratuitos e com avaliação positiva nos processos conduzidos pelo Ministério [da Educação], de acordo com regulamentação própria*".

[14] In 2019, the National Minimum Wage in Brazil remained fixed at 998 BRL (198,56 USD) per month.

resource allocation. By contrast, Hanushek (2003) has done a lot of statistical analysis to find the right relationship among key educational variables in the education production function and the output of the education enterprise. Hanushek (2003) clearly stated that the great advantage of using economic theory in education lies in quantifying not the determinants themselves but rather the value added. Therefore, this dissertation opens up a new avenue for investigating education in Brazil, one which is more focused on the output (i.e., academic achievement), rather than the inputs, in the education production function. The next section discusses the economic theory of education, addressing the education production function and value-added models.

2.2 The Economic Theory of Education

Economics, as a Social Science, seeks to understand how firms make decisions about the inputs needed for production, since the society has scarce, limited resources. In other words, firms need to face tradeoffs to produce goods or provide services. In education, for instance, one has to decide between more teachers or more computers, more libraries or more classrooms, more public schools or more scholarships for private schools. As such, economics is closely related to politics, as these choices are dependent on the type of investment that should be made.

Mitchell and Mitchell (2003) argue that much effort is devoted to determining inputs that drives school effectiveness (the desired transformation of inputs into outputs), while little is discussed about the role of school. Clarity about the purpose of education can help identify the most appropriate choice of inputs in the education production function. Table 8 shows this relationship.

Table 8. A framework for analyzing political economy establishing the relative value of education

	What Aims for Education?	
Who benefits?	***Education as Technical**: Training in Skills of Practical Value Having Economic Value*	***Education as Cultural**: Awakening of Identity and Character Having Political Value*
A private good: Distributed results accruing to individuals as education is obtained.	Durable product: Durable skills and knowledge with workplace value that persists over time (lasting benefits) (skilled work vs. business subsidies)	Direct service: Safe, nurturing, sensitive, caring child rearing and decent working conditions for teachers (child rearing vs. public safety)
A public good: Cumulative benefits for everyone; expected to accrue interest over time	Capital investment: System capacity building with some risk of not being realized by enough individuals to be worth cost (human capital vs. infrastructure construction)	Cultural legacy: Establishment of civic value that determines status and may lead or lag society (civic socialization vs. civic culture development)

Source: Mitchell and Mitchell (2003, p. 130,134).

The authors call into question what kind of economic good is education. For them, education in modern society can be seen as 1) a service industry, 2) a manufacturer of durable goods, 3) an investment system in training human capital, or 4) a driver of cultural legacies that pass from one generation to another. The choice of educational inputs and the related tradeoffs will vary depending on how one perceives the role of education.

Considering the scarcity of resources, if the state conceives of the school as responsible for training skilled professionals (education as technical), the policy maker may face the tradeoff between investing in educational institutions or letting the companies themselves train their professionals by offering them subsidies. In contrast, if the government understands that the school lends itself to cultural training, then the trade-off may be between investing in childhood or providing public security, for example. At the aggregate level, the investment option could be between producing human capital or building roads (i.e., education as technical), or between investing in civic education or building parks and theaters (education as cultural).

Mitchell and Mitchell (2003) contend that there may be overlaps, i.e., it is possible to embrace both roles of education (technical and cultural), but in economic policy the choices need to be made considering what is priority. Therefore, policy makers need to have a clear understanding of the role of education if they are to decide when and where to invest.

2.2.1 The Education Production Function

In economics, transforming inputs into a final product is supported by the production process, and the production function is the formula that describes this process (Monk, 1989). It may be expressed as y = f (x), where y refers to the quantity of products and x refers to the quantity of inputs employed to produce y. By the same token, like firms need inputs (e.g., rubber) for production (e.g., tire); HEIs also need faculty, classrooms, library etc. to produce educated students.

In education, the production function is given by:

$$O = \gamma H + \beta X + \varepsilon \tag{17}$$

where: O is the result, H is the human capital, X is the vector of other determinants for O, and ε is the stochastic error (Hanushek & Woessmann, 2011). As the function is provided from the point of view of microeconomics, data are determined at the individual level, with result

O related to individual gains and *X* related to market experience, gender, health, conditions of training and learning, etc.

From an economic perspective, the greatest empirical question is how to estimate human capital (H). In general, level of education is the most preferred option. However, for education to really equate to human capital, strong assumptions are needed to sustain this measure (Hanushek & Woessmann, 2011). It should include personal characteristics of individuals and their families, as well as features and qualities of school inputs and teachers. This is where the education production function (EPF) is brought to bear, as in:

$$H = \gamma F + \varphi(qS) + nA + \alpha Z + v \quad (18)$$

where: (F) stands for the factors related to families, (qS) for the quality and quantity of school inputs, (A) for individual skills, and (Z) for other relevant factors. As such, level of education is represented by S and its quality by q (Hanushek & Woessmman, 2011).

Given the scarcity of resources, not only the number of inputs matters but also their quality (Monk, 1989). Therefore, qualified teachers, for example, are needed in the classroom to maximize the students' learning (HANUSHEK, 1989). However, there is no systematic answer to what maximizes student achievement. Even so, the EPF is useful in identifying what is associated with higher student achievement in a given context, as the relationship between educational inputs and output varies across different social groups, economies, countries, and so on (MONK, 1989).

The relevance of studies on educational output and determinants of academic achievement is in showing how difficult it is to empirically apply some economic models and in having findings that account for or point to a wide range of problems. In other words, such studies provide results beyond the school itself, including, for example, issues related to salaries, status, funding, and impacts of education quality (Hanushek, 1979).

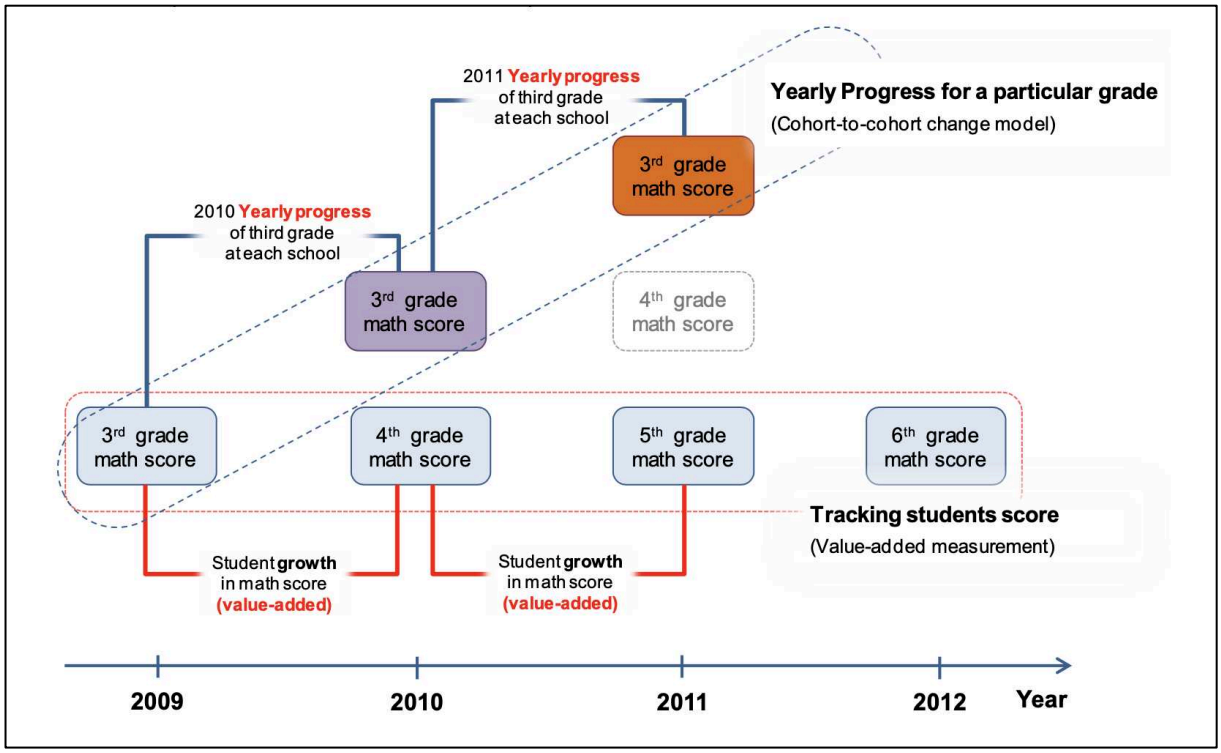
Basically, every country has looked for models to assess school effectiveness and support public policies that truly improve the economic and social perspectives of their population, as limited resources need to be productively allocated to meet the society's demands (OECD, 2016).

2.2.2 Assessment Models for School Effectiveness

Different assessment methodologies based on achievement tests can be used for decision making in education, including: status models, cohort-to-cohort change models,

growth models, and value-added models. Figure 3 depicts how academic progress is determined in each of these models.

Figure 3. Comparison between models for estimating academic progress



Source: Kim and Lalancette (2013, p. 5).

The status model serves to determine student achievement at a given time and compare it to a given goal. This type model answers questions such as "What percentage of students are performing well this year?" or "Has school x achieved the country's proficiency goal?" (See National Research Council, 2010, p. 4).

The cohort-to-cohort change model measures the variation in academic test scores for a given teacher, school, or state, and compares their statuses at two different times (but not based on the same students). Thus, yearly proficiency can be compared in two consecutive years for the same teacher or school, for example. This model tends to answer the following question: "Are this year's students in grade x performing better/worse compared to students in the same grade last year?" (National Research Council, 2010, p. 4).

The growth model measures student achievement by tracking their scores year by year to determine how well they have progressed. It attempts to answer the following question: "What was the achievement gain of student x from grade 1 to grade 2?" This model can also introduce goals to be achieved. It is focused on measuring the gain rather than the

factors responsible for student achievement growth; as such, it does not control for school or student factors (National Research Council, 2010, p. 4).

Value-added models, which are the focus of this dissertation, are usually complex statistical models that attribute a part of student achievement growth to particular schools, teachers or programs. It seeks to answer questions such as "How does the contribution of school x (or teacher x) to student improvement compare to that of the average school (or teacher)?" or "How much of the change in student achievement can be attributed to students' attending a given school (or a given teacher's class) instead of another?" (See National Research Council, 2010, p. 4).

The underlying assumption of each method for educational assessment has implications for policy decisions, which must take into account the purpose of assessment. For example, the status model is ideal if policy makers are interested in the accountability system of education and in having students achieve a certain level of proficiency. In this model, schools with high academic achievement and low value added will be rewarded, while schools with poor academic achievement but high value added will be punished. In fact, schools should be rewarded for their effectiveness as shown through their students' above average progress, even though several of them do not reach the required level of proficiency (National Research Council, 2010).

If the purpose is to assess which teachers or schools have been most effective compared to their peers in promoting the academic growth of their students throughout the school year, then the value-added model would be the best assessment system to determine rewards. Still, there are several ways to estimate this growth (by average gains in test results, by proportional change in students meeting or exceeding predetermined standards, by difference between current and expected average growth), and they will impact determining both the scope of the objectives and the tradeoff of the assessment process (National Research Council, 2010).

Academic achievement is a cumulative process resulting from the performance of previous teachers, peer interaction, actions of school leaders, and factors such as out-of-school experiences, including those in the family and in the community (Harris & Sass, 2005). Because of this, schools and teachers have argued that assessment methods (other than the value-added models) are not fair as they ignore factors over which the institutions have no control. The next section provides the concept, use, and methods for estimating value added.

2.2.3 Value added in the Education Production Function

Over the 1990s, elementary schools were compelled to prove their operational results, as the decision makers' focus shifted from the educational inputs (e.g., teacher-student ratio, expense per student) to the results (i.e., achievement measures) (Liu, 2011b). The need to measure the effect of school on student achievement has grown ever since, perhaps especially because of, political demands in countries of the Organization for Economic Co-operation and Development (OECD) (Melguizo & Wainer, 2015; OECD, 2008).

In the US, for example, standardized tests assessing student learning were widely incorporated into the education system upon the launch of the No Child Left Behind Act (NCLB) in 2002. This has allowed for following up student academic achievement over the years and feed into a growing body of research estimating value added. As a result, value-added measures have been used as metrics of school/teacher effectiveness in the country (Kim & Lalancette, 2013; Steedle, 2012).

The mixed model approach (with fixed and random effects), as developed by William Sanders (TVAAS – Tennessee Value-Added Assessment System), stands out amongst the most cited approaches to basic education. It is also used to implement public policies for teacher salary and allocation of resources to educational institutions in the US and OECD countries (Kim & Lalancette, 2013).

In higher education, not only is the need to measure school contribution to achievement more recent, but large-scale standardized test data are scarce, and early studies date back to the late 2000s (Klein et al., 2007; Klein et al., 2009; Liu, 2008, 2011a, 2011b; Steedle, 2009, 2010, 2011; Steedle et al., 2010). Incentives for estimating the value added of HEIs have stemmed from the increased costs of this level of education for both families and governments, as well as high rates of dropouts and concerns with the knowledge acquired by students during their undergraduate studies, which has seemed to fall short of what should be expected for the labor market (Melguizo et al., 2017). In addition, as tuition and other fees have increased, different stakeholders have been interested in how public investment has been used and what can be done to improve institutional effectiveness (Liu, 2011b).

Value-added measures may be used to: 1) develop exploratory research on institutional and/or faculty characteristics that positively impact academic achievement (Clotfelter et al., 2007; Easton, 2008); 2) improve the effectiveness of the school (or

teacher)[15] by identifying what stands out in the most or least effective ones so as to improve the distribution of resources across institutions (Sanders & Horn, 1998); 3) assess institutional programs that are contributing to the academic achievement of students in different schools (Xu et al., 2007); 4) support accountability systems for schools and teachers, with value-added estimates providing citizens (i.e., taxpayers) with information about the (in)efficient use of public money (National Research Council, 2010); 5) determine public policies with direct impacts on teachers/schools; and 6) identify incentives and outcomes within schools or across teachers (Koedel & Betts, 2009).

The education production function seeks to identify the human capital accrued per student in a given period to estimate input contribution in this period. To measure how much a teacher/school affects student achievement based on a theoretical value-added model that includes the fixed effects of a teacher/school (Koedel et al., 2015; Hanushek & Rivkin, 2010), the education production function is represented as follows:

$$A_g = \theta A_{g-1} + \tau_j + \varphi S + \gamma X + \varepsilon \quad (19)$$

where: A_g stands for the achievement of student *i* at time g, A_{g-1} for the student's achievement at time g-1, S for a vector of school factors, X for a vector of family background factors, and letters θ, φ and γ for unknown parameters. The stochastic error is represented by the letter ε. The teacher's fixed effects are represented by τ_j, which provides the measure of the value added of teacher or school j (Hanushek & Rivkin, 2010).

However, this estimation methodology is not a consensus amongst researchers and decision makers, varying across studies and countries. As different methods do not produce similar results, caution is needed when comparing studies. From a public policy perspective, care must be taken in deciding which control variables to include and the optimal sampling necessary to accurately estimate school effectiveness; as such decisions can affect the value-added estimate (Center for Education Policy Research, 2011).

2.2.4 Value-Added Models (VAM) for Higher Education

The concept of value added in education is related to student achievement as growth of skills, proficiency, or other attribute that students have achieved as a result of their

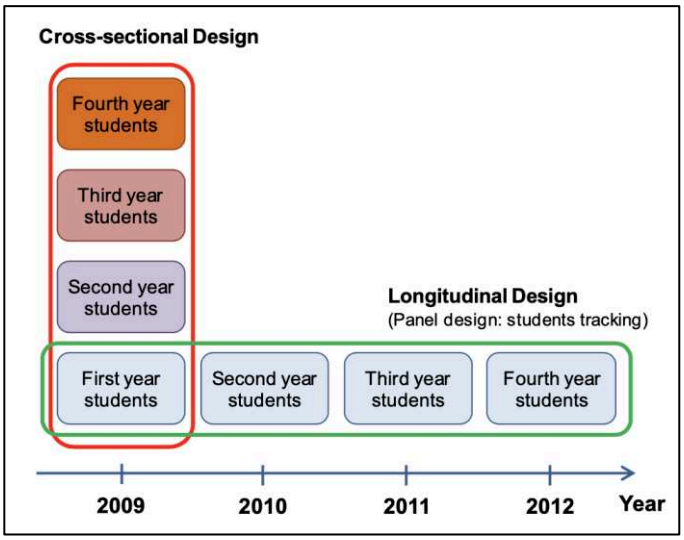
[15] In K-12 education, Alexander, Jang, and Kankane (2017) as well as Alexander & Jang (2019) found that including VAMs in teacher evaluation models were not a very effective way of improving student achievement in the US. It had negligible improvements in reading and no significant improvement in math. Also, it did not result in reduced disparities among key student groups.

experience within an education system in a given period. A VAM is a statistical model that uses student data over time to measure their learning gain and answer questions such as: "What is the proportion of variance in student achievement that can be ascribed to schools?", "How effective is a school in achieving results?", or "What institutional features or practices are associated with effective schools?" (See Kim & Lalancette, 2013, p. 5).

Studies using the VAM-based education production function represent a shift in research from an approach to the relationship between specific school characteristics and academic achievement to a less parametric approach that identifies the institution's and/or teacher's overall contribution to student learning (Hanushek & Rivkin, 2010) precisely because it places focus on the educational outcome. In other words, VAMs are focused on the link between institutional characteristics and academic achievement, rather than on the relationship between observable variables and academic achievement. However, value-added studies addressing higher education are incipient (Liu, 2011), especially in Brazil.

In general, research on higher education has used a cross-sectional design (by working with test scores applied to different student groups within a given HEI: freshmen and seniors) because of the inherent difficulty in following up a given student along his/her academic studies. In longitudinal research, the students' academic growth is followed up over the years, as shown in Figure 4.

Figure 4. Cross-sectional and longitudinal design in higher education research



Source: Kim and Lalancette (2013, p. 8).

Despite the practical difficulty of developing longitudinal surveys, Liu (2011a) argues that these should be prioritized, as correspondence between freshmen and senior may

not be optimal, given the high rates of transfers to other programs or institutions, dropouts, and suspended registrations. Four major distinct approaches may be used to address value added in higher education, namely: 1) estimating value-added by comparing predicted achievement to actual achievement based on standardized test results at the beginning and at the end of undergraduate studies, 2) comparing results between tests with open questions applied in the first and in the last year of studies, 3) comparing student commitment in the first and in the last year, and 4) collecting feedback from employers about the skills of graduates hired by them (Coates, 2009).

In general, research has resorted to the first two approaches, because of data availability. Studies, for instance, are recent and difficult to compare in the US, where no standard assessment system is in place with extensive use of standardized tests. In contrast, standardized tests are applied at the end of high school and at the end of undergraduate studies in Brazil and Colombia, where it turns out to be possible to perform value-added analyses using approach 1 (Coates, 2009).

Awareness of the contrasts across the different methodologies for estimating value added is relevant to choose the ideal model in light of the context and purpose of the education policy (Kim & Lalancette, 2013). In general, the academic achievement of educational institutions is measured by using the results of knowledge tests applied to their students. Achievement in such cases can be measured as mean scores, as percentage of students who attain a minimum score, or as growth from one time to another. Table 9 shows the most common VAMs in higher education research.

The different VAMs share the fact that they all use knowledge tests as a measure of prior achievement. They also fit variables related to student characteristics or to school context, but studies have no consensus on which variables to include. All models eventually show that some schools are significantly better or worse than the mean. Since models differ in how they use data (years, assumptions, missing data, and variable fitting), their results are not the same (National Research Council, 2010).

Table 9. Value-added models (VAM) for Higher Education

Model	3.5 Determinants of achievement	Value added Estimation
Ordinary Least Square Linear Regression Models [1]	$y_{ij(2)} = \beta_0 + \beta_1 y_{ij(1)} + \beta_2 X_{ij} + \beta_3 X_j + \varepsilon_{ij}$	$VA_j = ave(y_{ij(2)} - \hat{y}_{ij(2)})$
Random effects models using hierarchical linear model [2]	$y_{ij(2)} = \beta_{0j} + \beta_{1j}(y_{ij(1)} - \bar{y}_{j(1)}) + \beta_{2j}(X_{ij} - \bar{X}_j) + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{0s} W_{sj} + u_{0j}$	$VA_j = u_{0j}$
Difference in residuals model: OLS linear regression-based approach [3]	$\bar{y}_j = \beta_0 + \beta_1 \overline{SAT}_j + \varepsilon_j$	$VA_j = [\bar{y}_{j,se} - \bar{y}_{j,fr}] - [E(\bar{y}_{se}) - E(\bar{y}_{fr})]$
Difference in Residuals Model: HLM based [4]	$y_{ij} = \beta_{0j} + \beta_{1j}(SAT_{ij,se} - \overline{SAT}_j) + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + \gamma_{0s} W_{sj} + u_{0j}$	$VA_j = [\bar{y}_{j,se} - \bar{y}_{j,fr}] - [E(\bar{y}_{se}) - E(\bar{y}_{fr})]$
HLM-based residual analysis model [5]	$y_{ij,se} = \beta_{0j} + \beta_{1j}(SAT_{ij,se} - \overline{SAT}_{j,se}) + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{SAT}_{j,se} + \gamma_{02}\bar{y}_{j,fr} + u_{0j}$	$VA_j = u_{0j}$

Notes: [1] $y_{ij(2)}$: test score of student i in school j in time 2; $y_{ij(1)}$: test score of student i in school j in time 1; β_0: intercept (mean score of all students when all independent variables equals 0); $\beta_1, \beta_2, \beta_3$: slope of the line for the independent variables; X_{ij}: student characteristics; X_j: school characteristics; ε_{ij}: error term, assumed to be normally distributed and covariate-independent; [2] $\bar{y}_{j(1)}$: mean of prior test scores in school j; $\bar{X}_j$: mean of each student characteristic in school j; β_{0j}: intercept of school j; W_{sj}: school characteristics (s stands for the number of characteristics in the model); γ_{00}: level-2 intercept; γ_{0s}: slope of the level-2 line; ε_{ij}: level-1 residual (normally distributed and level-1 covariate-independent); u_{0j}: level-2 residual (normally distributed and level-2 covariate-independent); [3] $\bar{y}_j$: mean test score of a student at school j; $\overline{SAT}$: mean score for academic knowledge before admission to school j; $\bar{y}_{j,se}$: seniors' mean scores in school j; $\bar{y}_{j,fr}$: freshmen's mean scores in school j; $E(\bar{y}_{se})$: seniors' expected mean score; $E(\bar{y}_{fr})$: freshmen's expected mean score; [4] $SAT_{ij,se}$: academic ability score of student i in school j; [5] $y_{ij,se}$: test score of senior student i in school j; $\bar{y}_{i,fr}$: mean score of the freshmen's admission test in school j.
Source: adapted from Kim and Lalancette (2013).

Table 9 shows that value-added models are entirely based on standardized tests that measure student learning scores at two points: at the beginning and at the end of the undergraduate studies. The unknown variable *y* stands for student achievement at the end of the graduate studies, while *SAT* corresponds to student achievement at the end of high school. In Brazil, while the Enade measures student knowledge in the last year of the undergraduate studies, the Enem tests student knowledge at the end of high school.

In Colombia, students, too, take standardized tests to measure their knowledge and skills at the end of high school (SABER11) and at the end of the undergraduate studies (SABER PRO), which allows for estimating the contribution of a given institution to their students' training. Like in Brazil, the learning test named SABER PRO is mandatory for students who have completed at least 75% of their undergraduate studies and is composed of a specific training section and a general training section (Melguizo et al., 2017).

In Australia (Coates, 2009), Germany (Zlatkin-Troitschanskaia et al., 2016), OECD countries (Tremblay et al., 2012) and the US (VSA, 2019), different methodologies are aimed to detect the students' learning outcomes at the end of their undergraduate studies

in order to estimate value-added measures. In the US, for example, the Voluntary System of Accountability (VSA) was created in 2006 to assess the main higher education outcomes and improve the society's understanding of how the HEIs operate (Liu, 2011a). The VSA is a program that allows participants to access a large number of measures of their institution and other participating HEIs in order to compare them (VSA, 2019). These measures include students' learning test scores at the end of the undergraduate studies, namely: Student Achievement Measure (SAM), Measure of Academic Proficiency and Progress (MAPP), Collegiate Assessment of Academic Proficiency (CAAP), and Collegiate Learning Assessment (CLA). They also include the results of knowledge tests applied to students at the end of high school, such as the Scholastic Aptitude Test (SAT) and the American College Testing (ACT).

Based on 2006 and 2008 MAPP data from a sample of 6,196 students from 23 different U.S. institutions, Liu (2011a) estimated value added as the difference in achievement between freshmen and seniors upon controlling the admission scores (as measured by the SAT or the ACT), i.e., by applying an ordinary least squares (OLS) based residual analysis model. The analysis was at the institutional level, and the response variable was the mean MAPP. To estimate value added, each regression model (for freshmen and seniors) produced an error, which is the difference between the observed MAPP and the expected MAPP. This error was standardized (through the ratio between the error and its standard deviation), and the difference between the freshmen's and seniors' standard errors was the value-added score of each HEI (LIU, 2011a).

In Liu's study (2011a), the students' prior knowledge, as measured through the SAT, accounted for 70% of their achievement variance as measured by the MAPP. The remaining 30% of variance was accounted for by institutional and student factors such as income, maturity, interest, and family support. Liu (2011a) admits that increased scores on knowledge tests is important for revealing institutional effectiveness, but other factors such as rate of graduates against total number of original places, time to graduation, employability rates, student satisfaction, and rates of admission to graduate studies are also relevant for discussing the extent of an institution's contribution to student achievement. However, the author acknowledges a limitation in not having used other explanatory variables (related both to the student and the institution) for the MAPP achievement.

In a later study, Liu (2011b) used the same sample to compare two VAMs: OLS-based residual analysis model and HLM-based residual analysis model. In addition, the author introduced two control variables: selectivity (as measured by the percentage of students

admitted from all candidates), and degree awarded (undergraduate or graduate degree). Selectivity proved to be a significant variable to explain academic achievement, which suggests that institutional factors should be included in estimations of value added. Considering that the students were grouped by school, the HLM proved to provide a multilevel structure of data more adequately, estimating how much of the variance in academic achievement was accounted for by differences between and within institutions (Liu, 2011a). In contrast, the OLS model is based on the assumption that the observations are independent; however, school-level observations (starting and ending scores) are not independent when estimating value added as they are affected by institutional characteristics (Liu, 2011).

Melguizo et al. (2017) used one single database to compare three VAMs: 1) fixed effects, 2) random effects, and 3) aggregated residuals effect. Building on data from Colombia, the authors employed as response variables the graduation rate, the employment rate of graduates, and the SABER PRO score. The ranking of the HEIs changed for each response variable, which suggests that different achievement variables should be explored to assess educational quality. The study showed empirical evidence to support that the VAM based on fixed effects of the HEIs was the best alternative to address student selection bias. According to Kim and Lalancette (2013, p. 14), fixed effects models assume that each school has its own fixed effects on student achievement in order to "reduce bias caused by not reflecting unmodeled differences between schools".

2.2.5 Criticisms of Value-Added Models

The very existence of different methodologies of value-added estimation evinces unresolved issues in the literature. Three controversial dimensions stand out in debates amongst researchers and practitioners, namely: 1) use of the VAM and its possible consequences in higher education, 2) VAM measurement methods, and 3) VAM statistics. Criticisms in the first dimension can be reduced by: 1) making explicit the methodological choices and the purpose of estimation, which eventually allows for comparisons and identification of tradeoffs; 2) using other achievement metrics to make the model reliable and valid for important policy decisions; and 3) predicting the consequences of the potential incentives that the VAM may awaken in the stakeholders in the educational process (National Research Council, 2010).

Criticisms in the second dimension may be summarized as follows:

1) the use of knowledge tests is insufficient to measure academic achievement – if the test is not broad enough to cover most of the expected student knowledge or, else, if it is too rigorous, it is unlikely to provide accurate information about the value added by teachers/schools;
2) all tests are susceptible to measurement errors at both the individual and aggregate levels – the greater the error, the lower a study's ability to measure value added by employing a knowledge test;
3) the tests must be reported on equal interval scales, i.e., the difference between 10 and 20 points must be equal to the difference between 80 and 90 points;
4) some VAMs allow for comparing different series along a common scale; it is important to figure out how the properties of vertical scales affect value-added estimates (National Research Council, 2010).

Green et al. (2012) agree that the prediction errors are a technical problem of the VAM. For instance, a 2010 survey by the U.S. Department of Education found that the rate of Type 1 errors (the likelihood of a "good teacher" being falsely identified as a "bad teacher") and Type 2 errors (the probability of a "bad teacher" being falsely identified as a "good teacher") is 25% in a 3-year-long database and 35% for a 1-year-long database.

Another criticism is toward the fact that the VAMs neglect the knowledge acquired by a student in interaction with peers (Center for Education Policy Research, 2011). In fact, students do share knowledge in groups to perform academic activities. In higher education, students engage not only in junior research, extension, teacher assistance and tutorial education, but also in sports and cultural activities through which they share knowledge with students from different programs.

Four major criticisms stand out toward the statistics used in the VAM. The first one is bias-related: to address the non-randomness of students in schools, the VAM uses their prior knowledge and sometimes the students' and the schools' characteristics as control variables; as a result, the school effect may be underestimated or overestimated depending on the predictor variables included in the model. Another criticism involves sample size: surveys based on a small number of students of a given teacher/school do not accurately estimate value-added, as few underperforming students can significantly reduce teacher/school effectiveness and, therefore, reduce the actual value added. Another issue is data quality: missing data may affect the accuracy and stability of the value-added estimate and produce bias. Finally, more complex methods tend to feature better technical quality, but they are

more difficult for the general public to understand; therefore, it is necessary to choose models that are both complex and transparent (National Research Council, 2010).

Green et al. (2012) discuss the drawbacks in estimating teacher effectiveness using value-added models. They show that the correlation between a teacher's value added from one year to the other is low, around 0.2 to 0.3, and cannot be explained either by student characteristics or by variables of the educational context. Such low correlation, the authors argue, may be indicative of the low significance of using the model to implement teacher policies. Alexander, Jang and Kankane (2017), examining k-12 students performance among states from 2007 to 2013, showed that including student achievement in teacher evaluation models does not have significant positive effect in improving student performance. Also, they found out that policies based exclusively on student scores didn't get rid of wide gaps in achievement between students groups (Alexander, Jang, & Kankane, 2017). Therefore, choosing a methodology requires awareness of its limitations in order to ensure research quality.

As this dissertation aims to show the impacts that changing the methodology for estimating the IDD may have on the ranking of undergraduate programs in Accounting in Brazil, the next chapter describes the methodological steps pursued to attain this goal.

3 METHODOLOGY

This chapter details the methodological aspects that guided the study of the object of research (the IDD quality indicator), including the research approach and the method and procedures for data collection and analysis.

Within the economics of education, researchers face some methodological challenges, especially to carry out empirical research, as in this case. One of the explanations for the lack of consensus on the relationship between inputs and output in the education production function stems from three recurring problems: 1) the lack of 'good' data to estimate the function, 2) theoretical challenges in adapting the microeconomic theory of firm to the economics of education, and 3) methodological challenges to statistical analysis (Brito, 2012).

Defining the research methodology in education is also challenging because this is a complex area that has multiple determinants. Still, such definition is expected to be linked to both the research problem and the theoretical framework used to approach it. The next section describes the research approach used in this dissertation.

3.1 Quantitative Research

A quantitative research design is defined as a logical set of procedures for data collection, analysis, and dissemination that allows for answering the research question and testing hypotheses about specific variables. Quantitative studies are used when the research is aimed to test the effect of a given process or describe trends and relationships between variables (Clark & Creswell, 2015). This design was considered as the most suitable for the present purpose of investigating the impacts of changing the methodology for estimating the IDD on the ranking of undergraduate programs in Accounting in Brazil.

The statistical method was chosen as the method of procedure to ensure greater research objectivity and accuracy (Gil, 1999). Such method is recommended when it is impossible to gain a deeper understanding of the phenomena and their relationships without quantifying them, because of the myriad and complexity of these phenomena (Martins & Theóphilo, 2017).

3.2 Database and Sample

The analysis of secondary data is aimed to describe trends in the variables of a given population by using a large, previously collected database and resorting to statistics

(Clark & Creswell, 2015; Johnston, 2014; Smith, 2011). The major advantages of this type of design include its cost-effectiveness and convenience of using available data.

However, like any other type of research, one which uses secondary data requires some methodological precautions regarding the sample and its validity, the measuring instruments, and the validity of both instruments and sample. The researcher is expected to avoid bias in data selection, use a representative sample of the population, observe if variability is sufficient to tap into key concepts, be careful with missing data, and sort out subgroups from large data sets (Clarke & Cossette, 2000). Large databases collected by government agencies tend to use a high-quality data set and be representative of the population, thus supporting the validity and generalization of the findings (Johnston, 2014).

The researcher is also expected to notice possible flaws in the original instruments and misconceptions in proxy measures. Also, time has some effect on the research, as some original data collection may have taken place long before data analysis and mismatches are likely to occur between the initial measurement conditions and the researcher's own proposal (Clarke & Cossette, 2000).

Johnston (2014) suggests a step-by-step procedure to use secondary databases, namely: 1) develop the research question, 2) identify the database, and 3) assess the database to ensure its usability for the research purposes. Clarke and Cossette (2000) also recommend that after being granted access to the data, researchers should 1) prepare the database for proper use in their own research, which includes spotting missing data and making the necessary arrangements, 2) use statistics for descriptive analysis and other types of analysis, and 3) report the results properly.

Based on the concerns above, the first step in this research was to identify which databases would provide the necessary material to attain the purpose of this dissertation. Data from the different levels of education in Brazil are public, free, and easily retrievable from the federal government's website. The following databases were obtained over the second half of 2018: 1) 2015 Enade score, 2) 2015 Preliminary Program Quality Level (CPC), and 3) 2015 Indicator of Difference between Observed and Expected Achievements (IDD).

As for the 2015 Enade score, data were obtained from all students who took the Enade in 2015, as well as their answers to the socioeconomic questionnaire applied before the test. To be included in the sample, students needed to have a score above zero and valid answers to the questionnaire. The following data were collected from this database: Enade score per student, student socioeconomic characteristics (age, gender, ethnicity, nationality,

parental education, family income, type of high school, weekly study hours, yearly number of books read, employment status, and student funding or scholarship for high school).

As for the 2015 CPC, data were extracted from all undergraduate programs in Accounting that had a CPC available for year 2015. More specifically, the following data were collected: program code, learning modality, federated state, seniors registered to and participating in the 2015 Enade, mean Enade scores (general training, specific training, and overall score), number of participants with Enem score, program score, standardized score for pedagogical teaching structure (ODP), standardized IDD score, standardized score for infrastructure and physical facilities (IFF), standardized score for opportunity for further training/learning (OAF), number of faculty members, standardized score for faculty members with a master's degree (Me), standardized score for faculty members with a doctoral degree (Doc), standardized score for employment contract (RT), continuous CPC, and CPC Level. These variables were chosen because they are the ones disclosed by Brazilian government when they talk about quality in higher education. All of them (ODP, IDD, IFF, OAF, Me, Doc and RT) are variables used to describe institutional characteristics of specific IES.

As for the 2015 IDD, data extraction targeted the student's prior performance as measured through their Enem scores. The scores were collected for the four knowledge areas assessed on Enem, namely: languages (LT), mathematics (MT), natural sciences (CN), and humanities (CH). Finally, as for the 2015 IGC, data collection targeted all data related to the educational institutions that provide undergraduate programs in Accounting. This included: type of institution and percentage of undergraduate and graduate students.

After concatenating all databases and considering that 65,283 Accounting undergraduates had signed up for the 2015 Enade (Brasil, 2015), the sampling resulted in a total of 30,668 students from 911 programs, representing 46.98% of the population.

3.3 The Correlational Research

Correlational research, aimed to describe to what extent the predictor variables relate to the response variables, was the statistical research of choice to estimate the magnitude and direction of association between the variables (Clark & Creswell, 2015), as the value-added measures establish a relationship linking institutional and academic variables to academic achievement.

Correlational research design is used to describe the relationship between variables. The procedure generally follows this logic: multiple variables are measured for

each participant, and statistics is used to estimate the magnitude and direction of the association between such variables (Clark & Creswell, 2015, p. 196)

Inep's value-added models are all based on regression equations where student and program variables are predictive of academic achievement. Thus, to meet the first specific objective of this study (i.e., compare the value-added models used in the Sinaes since its implementation), the undergraduate programs' IDDs were re-estimated for the sample following both the correlational design and Inep Technical Bulletins released since 2006.

Table 10 shows the periods of use of each IDD model and their estimation formulas, as well as the changes made due to data availability and the names ascribed to the models for analysis purposes. Appendix 1 provides a table with all the variables used in this dissertation, including name, code, type, and description.

Table 10. Value-added models tested in this dissertation

Inep models		**Models tested in the dissertation**	
Period	Inep formula	Changes to Inep formula	Formula used
2006-2010 (Model 1)	$IDD = C - \hat{C}$ $c = \beta_1 i + \beta_2 w + \beta_3 z + \beta_4 Doc + \beta_5 Me + \beta_6 RT + \beta_7 IFF + \beta_8 ODP + \varepsilon$ $\hat{c} = \hat{\beta}_1 i + \hat{\beta}_2 w + \hat{\beta}_3 z$	- The Enade scores of seniors (c) and freshmen (i) were replaced by the Enade and Enem scores of the same student, respectively. - Variable z (ratio between the number of seniors and the number of freshmen in program i) was excluded from the model because the same student is used in both equations.	$IDD = C - \hat{C}$ $c = \beta_1 CN + \beta_2 CH + \beta_3 LC + \beta_4 MT + \beta_5 w + \beta_6 Doc + \beta_7 Me + \beta_8 RT + \beta_9 IFF + \beta_{10} ODP + \varepsilon$ $\hat{c} = \hat{\beta}_1 CN + \hat{\beta}_2 CH + \hat{\beta}_3 LC + \hat{\beta}_4 MT + \hat{\beta}_5 w$
2011-2012 (Model 2)	$IDD = C - \hat{C}$ $c = \beta_1 \overline{Enem} + \beta_2 w + \beta_3 z + \beta_4 Doc + \beta_5 Me + \beta_6 RT + \beta_7 IFF + \beta_8 ODP + \varepsilon$ $\hat{c} = \hat{\beta}_1 \overline{Enem} + \hat{\beta}_2 w + \hat{\beta}_3 z$	- Instead of using Enem scores and Enade scores from different freshmen and seniors (c), the model used the respective scores of the same students. - Variable z (ratio between the number of seniors and the number of freshmen in program i) was excluded from the model because the same student is used in both equations.	$IDD = C - \hat{C}$ $c = \beta_1 \overline{Enem} + \beta_2 w + \beta_3 Doc + \beta_4 Me + \beta_5 RT + \beta_6 IFF + \beta_7 ODP + \varepsilon$ $\hat{c} = \hat{\beta}_1 \overline{Enem} + \hat{\beta}_2 w$
2013 (Model 3)	$IDD = C - \hat{I}$ $C = \beta_1 \overline{Enem} + \beta_2 w + \beta_3 z + \beta_4 Doc + \beta_5 Me + \beta_6 RT + \beta_7 IFF + \beta_8 ODP + \beta_9 OAF + \varepsilon$ $\hat{I} = \hat{\beta}_1 \overline{Enem} + \hat{\beta}_2 w + \hat{\beta}_3 z$	- Instead of using Enem scores and Enade scores from different freshmen and seniors (c), the model used the respective scores of the same students. - Variable z (ratio between the number of seniors and the number of freshmen in program i) was excluded from the model because the same student is used in both equations.	$IDD = C - \hat{I}$ $C = \beta_1 \overline{Enem} + \beta_2 w + \beta_3 Doc + \beta_4 Me + \beta_5 RT + \beta_6 IFF + \beta_7 ODP + \beta_8 OAF + \varepsilon$ $\hat{I} = \hat{\beta}_1 \overline{Enem} + \hat{\beta}_2 w$
2014 – to date (Model 4)	$IDD = C - \hat{I}$ $C = \beta_0 + \beta_1 CN + \beta_2 CH + \beta_3 LC + \beta_4 MT + \varepsilon$ $\hat{I} = \hat{\beta}_0 + \hat{\beta}_1 CN + \hat{\beta}_2 CH + \hat{\beta}_3 LC + \hat{\beta}_4 MT$	- No estimation changes	$IDD = C - \hat{I}$ $C = \beta_0 + \beta_1 CN + \beta_2 CH + \beta_3 LC + \beta_4 MT + \varepsilon$ $\hat{I} = \hat{\beta}_0 + \hat{\beta}_1 CN + \hat{\beta}_2 CH + \hat{\beta}_3 LC + \hat{\beta}_4 MT$

Source: based on Inep (2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019).

The variables were used following the current Inep description for all models tested in this research (Models 1-4). This stage aimed to identify the methodological and practical implications in estimating the value added by a program to student achievement following the different Inep Technical Bulletins. Statistical differences between the models were analyzed using the freeware R (R Core Team, 2017) and software Excel®.

Following Inep Technical Bulletins, OLS multiple linear regression was the statistical technique used to estimate the IDD in Models 1-3, while HLM regression, also known as multilevel regression analysis, was used to estimate the IDD in Model 4.

3.4 Program Quality Levels According to the Value-Added Methodology

To accomplish the second specific objective (i.e., rate undergraduate programs in Accounting based on different VAMs), the quality indicator CPC was re-estimated for each of the models tested. To this end, the Enade score had to be re-estimated following the 2015 Inep Technical Bulletin, so that it could fit to the research sample data. The variables related to the student's perception of the training/learning process (infrastructure 'IFF', pedagogical teaching structure 'ODP', and opportunity for further training/learning 'OAF') and to the faculty characteristics (faculty members with a doctoral degree 'Doc', faculty members with a master's degree 'Me', and employment contract 'RT') were extracted from the 2015 CPC database. Once the CPC was re-estimated, the quality of the programs based on this indicator was compared across the different models by testing the means and analyzing the variance of the continuous CPCs and the number of programs in each CPC Level, with a view to identifying if the CPC of a given program would change depending on the IDD estimation methodology.

A further measure of program rating was a measure of school effectiveness and was based on the standard error associated with the value-added estimate of each model. It was developed by the Tennessee Value-Added Assessment System and applied to higher education by Liu (2011a). Measuring this standard error helps identify whether each institution's value-added estimate is significantly different from the growth pattern of the other institutions.

The IDD is used to compare effectiveness across schools in the sample. It rates institutions as in Table 11.

Table 11. Program rating by school effectiveness

Program ranking	Program's value-added estimate compared to the default value	Index	Interpretation
Level 5: Most effective	At least 2 standard errors above	Higher than or equal to 2.00	Significant evidence that students exceeded the Growth Standard.
Level 4: Above-average effectiveness	Between 1 and 2 standard errors above	Between 0.99 and 1.99	Moderate evidence that students exceeded the Growth Standard.
Level 3: Average effectiveness	Between 1 standard error above and 1 standard error below	Between -1.00 and 1.00	Evidence that students met the Growth Standard.
Level 2: Approaching Average effectiveness	Between 1 and 2 standard errors below	Between -1.99 and -0.99	Moderate evidence that students did not meet the Growth Standard.
Level 1: Least effective	More than 2 standard errors below	Less than or equal to -2.00	Significant evidence that students did not meet the Growth Standard.

Source: based on USA (2019, p. 35).

After rating the programs according to their comparative effectiveness, an analysis was carried out to identify whether their effectiveness is upgraded or downgraded (i.e., level change) by changing the IDD estimation.

3.5 Determinants of Achievement

To identify the determinants of academic achievement for undergraduate programs in Accounting in all editions of Enade (third specific objective), a database composed of Inep microdata (CPC and Enade) for years 2006, 2009, 2012 and 2015 was collected from the Inep website in April 2018 (Inep, 2018). The CPC database provided data about the characteristics of the HEIs, while the Enade database provided sociodemographic data about students.

The statistical technique of choice was regression analysis, one of the most traditional techniques for establishing a relationship between variables (Hair Jr. et al.). As data had been collected with students grouped into undergraduate programs in Accounting, a multilevel regression analysis was used to tackle the independence of observations.

The multilevel analysis was performed by following the steps in Laros and Marciano's (2008) recommendation manual: 1) model analysis without any explanatory variable, which is a null model, 2) model analysis with all lower-level fixed explanatory variables, 3) introduction of second-level explanatory variables, 4) assessment of regression coefficients, and 5) introduction of between-level interactions between explanatory variables (Laros & Marciano, 2008).

The proxy for achievement was overall Enade score, used as a dependent variable in the regression model. The independent variables (socioeconomic and school variables) were chosen based on previous studies on the determinants of academic achievement amongst Accounting undergraduates in Brazil (see Ferreira, 2015; Miranda et al., 2014; Miranda et al., 2015; Rodrigues et al., 2016; Rodrigues et al., 2017; Santos, 2012):

1) student-related: gender, age, marital status, ethnicity, income, parental education, number of books read, allowance to continue studies, gainful occupation, funding, hours of study, affirmative policy, school, and type of high school;
2) HEI-related: administrative category, type of academic organization, regional location, standardized program score for faculty members with a master's degree, faculty member with a doctoral degree, employment contract, infrastructure and physical facilities, pedagogical teaching structure, and opportunity for further training/learning.

Before performing the regression analysis, multicollinearity between the explanatory variables was assessed based on the variance inflation factor (VIF) – those with a VIF above 10 were excluded (Hair et al., 2009). Variable removal was backward, i.e., non-significant variables were removed one by one from the complete model (Draper & Smith, 1998). After extracting the regression parameters for the explanatory variables in the final model, the following tests were performed to assess the residual assumptions: normality (Shapiro-Wilk), independence (Durbin-Watson), and homogeneity of variance (Bartlett).

3.6 Proposal of a New IDD Estimate

Recent studies have estimated the value added of higher education institutions in different countries by using regression equations that include independent variables related to students and institutions (i.e., variables that were not linked to their policy, but rather to uncontrollable factors) (Bogoya & Bogoya, 2013; Cunha & Miller, 2014; Kim & Lalancette; 2013; Liu, 2011a, 2011b; Melguizo et al., 2017; Milla et al., 2016; Pike, 2016; Shavelson et al., 2016; Steedle, 2012).

Based on the assumptions of the value-added approach, the conceptual framework underling model design is focused on variables that usually escape institutional control but are predictive of academic achievement. Consequently, the VAMs include demographic and contextual factors, which are less susceptible to institutional control. In contrast, they omit

variables that represent policies subject to institutional action, because such policies are directly related to the institution's own effectiveness. Because of these policy influences, a fixed effect model with a new approach is suggested to calculate the value added. The fixed models "estimate school effects on student achievement as a fixed parameter", assuming that "each school has its own fixed effects on student achievement" (Kim & Lalancette, 2013, p. 14).

The proposal of a new IDD estimation model (Model IDD-VDCF) for undergraduate programs in Accounting in Brazil was dependent of identifying the determinants of students' academic achievement in all Enade editions. This fourth specific objective of this dissertation is aimed to investigate whether including control variables (at student and institution level) changes the quality rating of undergraduate programs. To this end, a correlational study was developed following Laros and Mariano's (2008) guidelines. The statistical procedures used for Model IDD-VDCF were: 1) multicollinearity analysis of the explanatory variables, 2) backward removal of significant variables, 3) variance analysis of the final model, and 4) testing of residual assumptions.

3.7 Dimensions of Quality for Education in Accounting

Up to this point, the discussion in this dissertation was focused on quality as conceived of by the Brazilian government. Regardless of the undergraduate program, quality in Brazilian higher education is measured through the quality indicators produced by the National Assessment System for Higher Education (Sinaes). However, the author's experience in her doctoral stay in the United State showed that the quality of undergraduate studies is not necessarily determined by a government-driven national assessment system. When it comes to higher education in Accounting, for instance, there is a strong presence of professional boards that work in partnership with educational institutions and determine the quality of their programs.

Thus, the fifth specific objective of this dissertation (i.e., discuss quality dimensions in Accounting training) was pursued by taking a basic qualitative research-based approach. This type of research aimed to explore multiple perspectives (Clark & Creswell, 2015) and was deemed suitable for the purpose of tapping into the different perspectives on what determines the quality of undergraduate programs in Accounting in Brazil and in the US.

Documental research was used as the data collection technique. Public documents included: 1) reports of professional boards – Association to Advance Collegiate Schools of Business (AACSB), American Institute of Certified Public Accountant (AICPA), Brazilian Federal Board of Accountants (CFC); and 2) Inep reports, documents, and laws. All public documents were retrieved from official websites in January 2019. Private documents included scientific articles published in national and international journals that address the determinants of academic achievement in Accounting education. Handwritten notes about these documents were used to collect and organize information. Content analysis was used to categorize the quality dimensions proposed by Harvey and Green (1993).

4 DATA ANALYSIS AND DISCUSSION

This dissertation aims to show the impact of changing the IDD estimation methodology on the rates of undergraduate programs in Accounting in Brazil. As described in the previous chapter, a large database was arranged (with Enade, CPC, and IDD data) to meet the research purposes. In this chapter, the results are divided into four sections. Section 4.1 describes data at the student level and the program level. Section 4.2 compares the IDD across Models 1-4. Section 4.3 compares the Program Quality Levels based on the CPC and school effectiveness across the four models. Section 4.4 pinpoints the determinants of academic achievement in all Enade editions.

4.1 Descriptive Analysis

The research database consists of 30,668 Accounting students from 911 undergraduate programs who took the Enade in 2015, answered the student questionnaire and had a valid Enem score before admission. These students were men (36.52%) and women (63.48%) in the last year of their undergraduate program in Accounting, with ages ranging from 18 to 79 years ($\mu = 26.6$; $\sigma = 6.5$). Most were single (69.28%), white or Asian descent (53.35%), and Brazilian nationals (98.83%). Most had fathers and mothers who achieved some level of basic education (90.13% and 87.18% respectively), and most were the first family member to achieve higher education (56.13%).

Most had a family income of up to 3,258.00 BRL in 2015, paid their own bills (56.74%) and worked (80.73%). Some had access to student funding (47.18%), and some received an allowance for daily expenses (4.19%) or a scholarship (8.41%) from the institution. Some (23.52%) were admitted to higher education using affirmative policies. Over a half of the sample studied up to three hours a week (59.93%) and read up to three books a year in addition to those required at school (51.72%).

Their achievements on the 2015 Enade test were as follows: overall mean score of 41.26 points ($\sigma = 12.61$), mean score in the general training section (FG) of 52.91 points ($\sigma = 14.26$), and mean score in the specific training section (EC) of 37.36 points ($\sigma = 14.47$). Most of them (93.33%) took the 2009, 2010 or 2011 Enem, and obtained an overall mean score of 517.69 points (out of 1,000). Table 12 provides the descriptive data for academic achievement.

Table 12. Descriptive Statistics of Accounting Students on the 2015 Enade

Variable	Mean	Median	Standard Deviation
Overall Enade score	41.26	40.50	12.61
Enade score – FG	52.91	53.20	14.26
Enade score – CE	37.36	36.30	14.47
Enem score – CN	486.50	488.00	71.32
Enem score – CH	514.57	515.40	79.14
Enem score – LT	527.75	534.70	65.18
Enem score – MT	541.93	545.20	101.17

Source: research data.

The 911 programs in the sample represented 96% of the undergraduate programs in Accounting that had their Preliminary Program Quality Level released in 2015. Most programs were based on in-presence learning (98.24%), rather than on distance learning (n = 16; i.e., 1.8%). Most were provided by private institutions (86.28%), in university centers, federal institutes or universities (55%), as well as in colleges (45%).

The descriptive statistics shows that the scores varied significantly between public and private institutions. Previous studies (Camargo et al., 2016; Fernandes et al., 2009; Fernandes et al., 2018; Ferreira, 2015) have already reported a higher achievement of Accounting students from public HEIs than those from private institutions. Therefore, program-related variables had their analysis segmented by administrative category (i.e., public or private).

Table 13 presents the descriptive statistics for the institutional variables, by administrative category. The variables related to physical features (IFF) and pedagogical features (ODP and OAF) pointed to higher standardized scores in the private HEIs than in the public institutions. In contrast, faculty-related variables (Me, Doc, and RT) had higher standardized scores in the public HEIs than in the private institutions.

Table 13. Descriptive statistics of undergraduate programs in Accounting on the 2015 Enade

Variable	Admin. Category	Mean	Median	Standard Deviation
IDD	Public	2.52	2.48	0.64
	Private	2.54	2.49	0.79
ODP	Public	2.42	2.27	1.00
	Private	3.32	3.32	1.05
IFF	Public	2.46	2.42	1.09
	Private	3.43	3.54	1.06
OAF	Public	2.62	2.53	0.91
	Private	3.20	3.23	1.10
Me	Public	3.47	3.67	1.15
	Private	3.38	3.52	1.17
Doc	Public	1.63	1.28	1.31
	Private	1.49	1.32	1.11
RT	Public	4.60	5.00	1.00
	Private	3.53	3.76	1.22
Enade (Continuous)	Public	2.93	3.00	0.91
	Private	2.27	2.22	0.76
CPC (Continuous)	Public	2.69	2.66	0.54
	Private	2.58	2.58	0.60

Source: research data.

In the end, the continuous Enade score was higher for the public programs than for the private ones, which is consistent with prior Brazilian studies, including for other programs, as previously mentioned. The Program Quality Level, the CPC, too was higher for the public HEIs, which probably results from the fact that the Enade score accounts for 20% and the faculty scores account for 30% of this indicator.

4.2 Comparative Analysis of IDD Estimates across Value-Added Models

To perform a comparative analysis across the VAMs, the 2015 IDD was re-estimated based on the Inep methodology in force between 2006 and 2010 (Model 1), 2011 and 2012 (Model 2), 2013 (Model 3) and 2014-to date (Model 4), as described in Table 10. The procedures for estimating the IDD in all models were as follows: 1) calculate regression equations according to the period-specific Technical Bulletins (INEP, 2009; 2010; 2013; 2016), 2) extract parameters to estimate standardized residuals, 3) remove values with standardized residuals greater than | 3 |, 4) re-estimate regression equations to compute parameters without outliers, 5) compute the Enade estimate, and 6) estimate the IDD (actual Enade score minus Enade estimate).

In Models 1- 3, the IDD was estimated based on a least-square multiple linear regression equation, where the best fit was to minimize the sum of squared differences

between estimates and actual data. In other words, the ordinary least squares (OLS) minimize the sum of the squared residuals in the regression, in order to maximize the model's degree of fit to the data set.

Models 1-3 are characterized as difference in residuals models. Their main characteristic is that they capture whether the students' academic growth in a given program is close to or above that observed amongst institutions that admit students with similar academic skills (Kim & Lalancette, 2013). To estimate the expected scores on the achievement test (Enade), Models 1-3 use regressions of actual Enade scores for freshmen and seniors. As the freshmen's scores were unavailable, this study assumed the value-added measure (IDD) to be the difference between the observed values and the expected values based on the admission scores (Enem) and the Accounting seniors' scores (Enade).

The difference between Model 1 and Models 2-3 is that the independent variable is the score in each Enem section (Natural Sciences – CN; Human Sciences – CH; Literature and Languages – LT; Mathematics – MT) in Model 1, but the mean scores of all Enem sections in Models 2-3. Model 3 is different from Models 1-2 in that the student questionnaire has been a broader version since 2013, and a new variable was captured based on their answers: the "opportunity for further training/learning" (OAF); in addition, variables "pedagogical teaching structure" (ODP) and "infrastructure" (IFF) were reformulated based on the questionnaire.

According to Inep Technical Bulletins (INEP, 2009; 2010; 2013; 2016), data are expressed around the mean of each program. Therefore, the IDD estimation for Models 1-3 was based on data at the program level, rather than at the student level. Table 14 shows the regression statistics for Model 1 after excluding outliers. Variables Doc and Me were non-significant in the model, while OPD and w were significant at 10%.

Table 14. Regression statistics – Model 1

	Coefficients	**Standard Error**	**t-statistics**	**p-value**
Intercept	-32.46	2.59	-12.52	0.00
CN	0.02	0.01	2.56	0.01
CH	0.02	0.01	2.65	0.01
LT	0.04	0.01	5.03	0.00
MT	0.05	0.00	10.71	0.00
w	-2.76	1.59	-1.73	0.08
Doc	0.02	0.15	0.13	0.90
Me	0.18	0.15	1.19	0.23
RT	0.39	0.11	3.53	0.00
IFF	0.78	0.30	2.57	0.01
ODP	0.53	0.31	1.71	0.09

Source: research data.

A comparison between Tables 14 and 15 shows that Models 1 and 2 did not produce much difference in the estimates for parameters related to each variable under scrutiny. This is probably because both models only diverge in how they employ the student's prior knowledge variable(s) (CN, CH, LT, MT, Enem) to explain the Enade results: Model 1 uses the scores of four Enem sections, while Model 2 uses their mean.

Table 15. Regression statistics – Model 2

	Coefficients	**Standard Error**	**t-statistics**	**p-value**
Intercept	-31.92	2.26	-14.10	0.00
Enem	0.13	0.00	30.65	0.00
w	-2.45	1.61	-1.53	0.13
Doc	0.18	0.15	1.15	0.25
Me	0.06	0.15	0.39	0.69
RT	0.32	0.11	2.91	0.00
IFF	1.07	0.30	3.56	0.00
ODP	0.31	0.31	0.99	0.32

Source: research data.

The results shown in Tables 14, 15 and 16 show that faculty characteristics (Doc and Me) produce no significant contribution to the models tested, and the same holds true for variable w, which means that parent's education does not account for student achievement. In Models 1, 2 and 3 the only significant institutional characteristic was infrastructure.

Table 16. Regression statistics – Model 3

	Coefficients	**Standard Error**	**t-statistics**	**p-value**
Intercept	-32.06	2.28	-14.05	0.00
Enem	0.13	0.00	30.54	0.00
w	-2.40	1.61	-1.49	0.14
Doc	0.06	0.15	0.43	0.67
Me	0.18	0.15	1.14	0.25
RT	0.32	0.11	2.95	0.00
IFF	1.17	0.36	3.29	0.00
ODP	0.37	0.34	1.11	0.27
OAF	-0.17	0.33	-0.51	0.61

Source: research data.

To the best of the author's knowledge, no study has used such a model at the institution or program level for higher education. In general, studies using OSL regression equations for higher education arrange their database at the level of freshmen and seniors (Kim & Lalancette, 2013; Liu, 2011b, Steedle, 2012). The value added is estimated by averaging the difference between the students' actual and estimated scores in a given

institution. In Brazil, as the variables were already based on the mean, the value-added estimate was already at the institutional level.

Unlike Models 1-3, Model 4 estimates the IDD using an HLM regression equation, with the main database arranged at the student level and the analysis performed at two levels: that of the student, and that of the institution (in this case, the undergraduate program in Accounting). Much as one of the assumptions in OLS models is that the observations are independent, academic achievement is known to be the result of student experience in the institution; thus, the test scores of students in a given program cannot be considered independent of the institution (Kim & Lalancette, 2013).

Table 17. Correlation between standardized Enem and Enade scores

	Student Level			Institution Level		
	Enade	**General Training**	**Specific Training**	**Enade**	**General Training**	**Specific Training**
Enem	0.54	0.37	0.51	0.71	0.61	0.68
CN	0.44	0.31	0.41	0.63	0.56	0.60
CH	0.43	0.31	0.40	0.60	0.55	0.56
LT	0.45	0.33	0.41	0.66	0.57	0.62
MT	0.44	0.26	0.43	0.69	0.53	0.67

Source: research data.

Table 17 shows, for instance, that the correlation between the Enade scores and the Enem scores is higher at the institution level than at the student level. As such, HLM models are more appropriate, and Model 4 is a better predictor in that it differentiates the variance of student scores within and between programs. Table 18 provides the results for the regression statistics in Model 4.

Table 18. Regression statistics – Model 4

Maximum likelihood-adjusted, linear mixed effects model					
Random Effects Variable: CUR_1					
	Intercept	Residual			
Standard Deviation	3.31	9.74			
Fixed Effects Variables					
	Coefficient	Standard Deviation	Degrees of Freedom	t-statistics	p-value
(Intercept)	-8.09	0.56	29634	-14.51	0.00
CN	0.02	0.00	29634	20.13	0.00
CH	0.02	0.00	29634	16.82	0.00
LT	0.03	0.00	29634	27.19	0.00
MT	0.02	0.00	29634	35.13	0.00

Source: research data.

The assumption of variance homogeneity was tested for Model 4, i.e., if level-1 (student) variance is equal across all institutions, it follows that the institutions do not contribute with additional variance to student achievement and, therefore, the method of choice should be the classical regression model estimated by OLS. The intraclass correlation coefficient was estimated to indicate the percentage of total variance that was explained by the programs. The result was a coefficient of 0.17, i.e., 17.04% of the variance in student achievement on Enade is explained by the difference between the programs, which is consistent with other studies focused on higher education (Ferreira, 2015; Liu, 2011b; Rodrigues et al., 2016).

The correlation between the IDDs was estimated to assess the relationship between the models, as suggested by Melguizo et al. (2017). Table 19 points to a high correlation between the fitted models. The values found are close to those reported by Melguizo et al. (2017), who compared three VAMs using data from Colombia. They found a correlation of 0.74 when comparing the cumulative residual model (similar to Models 1-3 in this dissertation) with the fixed effects model (with selectivity as a control variable), and a correlation of ca. 0.9 between fixed effects model and random effects model (with or without control variables).

Table 19. Correlation between value-added estimates (IDD)

	M1	**M2**	**M3**
M1	1.00		
M2	0.98	1.00	
M3	0.98	1.00	1.00
M4	0.76	0.78	0.78

Source: research data.

Models 2-3 are perfectly correlated (1). Table 20 confirms this through the similarity of the parameters across the different models. Model 4 has less similar parameters because its intercept varies at both the student and the institution levels (see Table 18).

Table 20. Parameter estimation in Models 1-4

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	-32.46*	-31.92*	-32.06*	-8.09*
CN	0.02*	-		0.02*
CH	0.02*	-		0.02*
LT	0.04*	-		0.03*
MT	0.05*	-		0.02*
Enem	-	0.13*	0.13*	
w	-2.76	-2.45	-2.40	-
Doc	0.02	0.06	0.06	-
Me	0.18	0.18	0.18	-
RT	0.39*	0.32*	0.32*	-
IFF	0.78*	1.07*	1.17*	-
ODP	0.53	0.31	0.37	-
OAF	-	-	-0.17	-

* significant at 95% confidence level.
Source: research data.

Some considerations are required when comparing Model 4 against Models 1-3. First, Models 1-3 ignore all student-level information, as they estimate the IDD at the institution level, while Model 4 uses a student-level database. Results tend to be more reliable at the student level, since the number of students is much larger than that of institutions (Liu, 2011b; Klein et al., 2008). In addition, while Models 1-3 uses characteristics of both students (parental education) and institution (faculty and program), Model 4 is based on standardized test scores but it doesn't include other variables such as parental education or infrastructure. However, research on Brazilian undergraduate programs in Accounting (Araújo et al., 2009; Beck & Rausch; 2014; Miranda, 2011; Ferreira, 2015; Rodrigues et al., 2015; Rodrigues et al., 2016) has underscored that several individual and institutional aspects should be factored in to determine academic achievement.

Finally, Table 21 provides the results for the analysis of IDD variance across the different models, which point to a significant difference between one or more IDD means. In order to find out which specific groups'means (compared with each other) are different, a Tukey's HSD teste was run. The results (Table 22) elucidate that the IDD difference in group-by-group mean tests is significant only between Model 4 and the other models.

Table 21. Analysis of IDD variance across the models

Source of Variation	SS	df	Variance	F	p-value	F crit
Between Groups	30.43	3	10.14	23.52	0.00	2.61
Within Groups	1569.50	3640	0.43			
Total	1599.926	3643				

Source: research data.

Table 22. Tukey HSD test across the models

Groups	Undergraduate Programs	Sum	Mean	Variance
IDD_M1	911	2023.79	2.22a	0.48
IDD_M2	911	1952.97	2.14a	0.52
IDD_M3	911	1943.91	2.13a	0.51
IDD_M4	911	2152.04	2.36b	0.21

Source: research data.

After estimating the IDD (difference between observed and Expected Achievement), the programs' CPC and the IDD standard deviation were estimated to rank the programs according to their effectiveness compared to the others (see section 3.4).

4.3 IDD-based Ranking of Undergraduate Programs

Following the Sinaes methodology, undergraduate Program Quality Level is measured by the CPC indicator, a pool of indicators (each with its specific weight) established as follows: Enade (20%), IDD (35%), percentage of faculty members with a master's degree (7.5%), faculty members with a doctoral degree (15%), percentage of faculty employment contract of over 40 hours/week (7.5%), program infrastructure (5%), pedagogical teaching structure (7.5%), and opportunity for further training/learning (2.5%).

Table 23. CPC Levels[16]

CPC (Continuous)	CPC (Categorical)
$0 \leq CPC < 0.945$	1
$0.945 \leq CPC < 1.945$	2
$1.945 \leq CPC < 2.945$	3
$2.945 \leq CPC < 3.945$	4
$3.945 \leq CPC \leq 5$	5

Source: INEP, 2018.

The program's continuous CPCs are rounded and transformed into Levels from 1 to 5, as shown in Table 23. To assess whether the program rankings differed through re-estimating the IDD, the CPCs were estimated in the four models for all 911 programs in the sample. Table 24 shows the count of programs in each CPC Level. It is interesting to note that

[16] The rounding of CPCs to fit one of five levels is questionable to compare the difference between programs: in the current terms, a program with a continuous rate of 2.95 has the same quality level as a program with 3.90, i.e., they are both in Level 4 (CPC = 4); meanwhile, a program with 2.94 is rated as 3 (CPC = 3), even though its continuous rate is much close to 2.95. In practice, the programs with continuous CPCs of 2.95 and 2.94 are more likely to be similar to each other and much different from those with continuous CPCs of 3.90. Had the rounding process adopt a different scale or a larger number of Levels, there could arguably be greater similarity between the programs ascribed with equal Levels.

the worst and the best (considering that the same institutions account for these numbers) are consistent across the models. It is also interesting that Model 4 results in more institutions considered to have achieved acceptable standards.

Table 24. Program ranking in CPC Levels

CPC Levels	Model 1	Model 2	Model 3	Model 4
Level 1	4	5	5	5
Level 2	145	160	161	101
Level 3	593	586	586	639
Level 4	165	156	155	161
Level 5	4	4	4	5

Source: research data.

From Model 1 to Model 2, 34 programs had their CPC Levels changed, 30 of which upgraded one level and four of which downgraded one level. The largest and smallest differences between the continuous CPCs of the 30 upgraded programs were 0.18 and 0.02, respectively, which implies that their upgrade was merely an issue of transforming continuous scores into categorical levels. In turn, the largest and smallest differences between the continuous CPCs of the four downgraded programs were respectively 0.1 and 0.03. From Model 2 to Model 3, two programs had their CPC Levels upgraded, with differences of 0.005 and 0.012 between their continuous CPCs. From Model 3 to Model 4, 112 programs had their CPC Levels changed, out of which 90 upgraded one level and 22 downgraded one level. In this case, differences between continuous CPCs were greater, ranging from 0.04 to 0.71.

In any case, such changes also occurred due to the change in value-added estimation. The control variable of student's prior knowledge was their individual scores in the four Enem sections in Model 1, but the mean of such scores in Models 2 and 3. Model 3 also included a further explanatory variable at the institution level – opportunity for further training/learning (OAF) – and extensively used the student questionnaire to measure the conditions of training and learning from the students' perspective. Model 4 excluded the institution-level explanatory variables and used the scores in the four Enem sections as the only explanatory variables in a multilevel regression model. This explains why a greatest change was found in comparing Models 3 and 4.

Amongst the programs that had their CPC Levels changed from Model 3 to Model 4, 62 upgraded from Level 2 to 3. Their mean IDD was 1.3 in Model 3 and 2.0 in Model 4. The 26 programs that upgraded from Level 3 (Model 3) to 4 (Model 4) had their mean IDD at

1.8 and 2.4 in Models 3 and 4, respectively. The two programs that upgraded from Level 4 (Model 3) to 5 (Model 4) had their mean IDD at 3.8 in Model 3 and 2.4 in Model 4.

Using HLM-based residual analysis (Model 4) has the advantage of increasing the reliability of school effects compared to OLS-residual difference models (Model 3). Thus, the reliability and consistency of the year-by-year estimate of value added for each institution are higher when applying the multilevel model (Kim & Lalancette, 2013). As such, the value-added estimate can be used by school leaders to determine how effective an institution is in contributing to student achievement over the years.

In general, value-added models lend themselves to comparing effectiveness across institutions (Horn et al., 2017; Kim & Lalancette, 2013; Liu, 2011a; 2011b; Melguizo et al., 2017; National Research Concil, 2010), which in turn requires defining measurement scales to compare the models. In addition to ranking the programs based on their CPCs, school effectiveness was defined using the standard deviations of the value-added scores in Table 11 (see section 3.4, page 64). Table 25 provides the ranking results.

Table 25. Program rating by school effectiveness across Models

Effectiveness Level	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Level 1 – Least Effective	14	15	16	78
Level 2 – Above average effectiveness	68	59	57	110
Level 3 – Average Effectiveness	578	587	587	532
Level 4 – Above Average Effectiveness	192	189	189	102
Level 5 – Most Effective	59	61	62	89

Source: research data.

For example, Table 25 shows that in Model 3, 16 programs are less effective compared to the others (Level 1); 57 are below average (Level 2); 587 have average effectiveness (Level 3); 189 are above average (Level 4); and 62 are the most effective programs in the sample (Level 5). The worst and the best (presuming that the same institutions account for these numbers) are consistent across the models. It is also interesting that Model 4 results in more institutions considered to have achieved acceptable standards.

Tables 24 and 25 reveal that the program rankings were similar across Models 1-3, supporting the results of the regression statistics. Considering Model 4-based CPC ranking (Table 24) compared to the others, programs tend to concentrate in Level 3, which assumedly stands for programs with satisfactory quality in the Sinaes classification. However, in comparing effectiveness across programs, the central values (Level 3) are reduced while

discrepancy increases across the programs, which were re-ranked in the most extreme levels (see Table 25).

When comparing Model 1 to Models 2 and 3, 37 programs had their comparative effectiveness reduced in one level and 44 had an increase of one level, even though the absolute numbers are close in Table 25. This shows that using mean Enem scores instead of separate scores for each section (CN, CH, LT, and MT) also impacts comparison across programs. When comparing Models 2 and 3, one program had its effectiveness level reduced, and three programs had their effectiveness increased. This supports previous data showing the similarity between these two models.

The greatest difference is found from Model 3 to Model 4: 70 programs had their effectiveness upgraded in one level, two programs upgraded two levels (both from "average" to "more effective" programs), 35 downgraded two levels, and 214 downgraded one level. Most programs (590) did not experience any change in their comparative effectiveness, but a meaningful percentage (i.e., 40% of these programs) did have their effectiveness level changed, which calls for a more fine-grained analysis.

Table 26 shows how the means for student and program variables changes with changing the estimation methodology (from Model 3 to Model 4). Column 1 describes the observed variables, column 2 provides the means of these variables for programs downgraded in two levels of effectiveness (from Model 3 to Model 4), column 3 shows the means for the programs upgraded in two levels (from Model 3 to Model 4), column 4 includes the p-value on the Student's t-test for the difference of means between the two groups, and column 5 introduces an interpretation for the change in mean from one model to the other.

Table 26. Comparison between means (Model 3 vs. Model 4)

	2-level down-grade	**2-levels upgrade**	**t test (p-value)**	**Interpretation (The most effective programs have…)**
Enade mean	42.71	50.96	0.00	Highest overall score on the Enade
Mean Enade score for FG	52.91	58.53	0.00	Highest Enade score for FG
Mean Enade score for CE	39.28	48.41	0.00	Highest Enade score for CE
Mean books read/year	0.43	0.55	0.05	Highest % of people reading above 3 books
Mean hours of study	0.28	0.51	0.00	Highest % of people studying above 3h
Mean Enem score for CN	490.83	501.31	0.05	Highest Enem score for Natural Science
Mean Enem score for CN	509.80	540.18	0.00	Highest Enem score for History
Mean Enem score for LT	533.45	546.44	0.00	Highest Enem score for Literature
Mean Enem score for LT	556.93	574.22	0.00	Highest Enem score for Maths
Mean IGC (Continuous)	2.52	3.00	0.00	Highest General Index
Mean ODP	2.26	4.29	0.00	Highest Pedagogical teaching
Mean IFF	2.27	4.48	0.00	Highest Infrastruture
Mean OAF	2.26	4.20	0.00	Highest Opportunity for further training
Mean Me	2.87	3.99	0.00	Highest % of faculty with a master's degree
Mean Doc	1.16	1.84	0.04	Highest % of faculty with a doctoral degree

Source: research data.

In programs with 2-level effectiveness upgrade, lower values are found for number of single students, number of white students or students of Asian descent, student income, and number of students with scholarships; while higher values are found for students' average age, ratio of parents with only some level of basic education, and number of individuals with student funding. Although the means were different between models, they were non-significant for the variables mentioned, and therefore, are not shown in Table 26 However, the t-test was significant for the difference in means for two student-related variables: number of books read, and hours of study. The group of programs with upgraded effectiveness has a higher percentage of students who study over three hours a week and those who read over three books a year.

Not surprisingly, such 2-level-upgraded programs also have higher Enem and Enade. They also had higher means for institutional variables IFF, ODP, OAF, Me, and DOC than the 2-level-downgraded programs. Particularly, the variable that measures the institutions' continuous IGC was also significantly higher. This means that the most effective programs are those with better infrastructure, better pedagogical teaching structure, and more qualified faculty members, as well as those within universities, university centers or federal institutes.

It follows from such a finding that: 1) the Tennessee ranking system (US, 2019), whereby school effectiveness is measured from the deviations of the mean is a good way to compare program effectiveness within and between different value-added models, and 2) multilevel regression models, which include data grouped by institution and data per student, provide more accurate measures of school effectiveness, especially when comparing institutions.

Finally, the historical evolution of the IDD estimation methodology shows that Inep has improved its system of measuring a program's contribution (value added) to student achievement. In the beginning, estimation of value added was based on the scores of both freshmen and seniors in a given institution in a given year and on control variables related to students and programs. Then, the scores of the same students were considered in two moments – before admission and at graduation – eliminating other variables.

Such changes have had an impact on the ranking of undergraduate programs in Accounting, with some institutions having their quality levels upgraded and others having their levels downgraded. The current model is the most appropriate from a statistical and practical standpoint, since pooling students within institutions reduces bias in estimating the value-added of a program and allows for comparability across programs. However, excluding

control variables related to the characteristics of students and institutions may be a problematic choice, as value-added measures provide more accurate estimates of a program's contribution to student achievement because they do incorporate a set of contextual characteristics of both students and institutions (Kim & Lalancette, 2013).

4.4 Determinants of Academic Achievement in Accounting

In several countries, institutional effectiveness is assessed primarily by student achievement measures, usually provided as mean scores in standardized tests (Milla & Martín; Bellegem, 2016), percentage of graduates (Bailey & Xu, 2012; Yunker, 2005) or achievement growth using value-added measures (Coates, 20009; Melguizo et al., 2017; Rodgers, 2005; Shavelson et al., 2016; Zhang, 2009).

The challenges to modeling value-added estimates lie both in gaining access to a valid, reliable instrument that measures the knowledge acquired by students in their undergraduate programs and in accounting for the bias of students' choosing one institution over the others, with motivation and prior academic preparation influencing learning (Melguizo et al., 2017). In Brazil, such challenges are partly overcome because the Enade measures students' knowledge at the end of their undergraduate studies, while the Enem measures their knowledge at the end of high school. In addition, the National Census of Higher Education and the Student Questionnaire provide relevant data both on the conditions of training and learning and on the students' socioeconomic characteristics. Therefore, it is possible to isolate the contribution of an HEI from the students' personal efforts by using the proper modeling.

After defining the achievement measure to be used for estimating value added, it is important to decide how the model will account for differences in students' quality before admission and in academic contexts (Shavelson et al., 2016). As such, the explanatory variables of achievement should be carefully chosen.

After carrying out a comparative analysis between the value-added models used by the government and review of the Brazilian and international literature on value-added models, an investigation was pursued to explore which student and institution variables are significantly related to the overall Enade score, which could eventually add to the debate on the most appropriate value-added model for the Brazilian context.

Enade data were used from the 2006, 2009, 2012 and 2015 editions as the student-level response variable for each year, and student and program variables were tested through

multilevel regression, as described in section 3.5. A regression equation with all model variables was tested for each year, with non-significant variables being removed one by one backwardly.

First, a study of multicollinearity among the explanatory variables was carried out to ensure that a high correlation between them did not produce negative effects on the goodness of fit. The method of choice was the variance inflation factor (VIF), which measures how much the variance of a coefficient is inflated by its collinearity. Values above 10 were excluded: gainful occupation (year 2006), and school shift (year 2012). Subsequently, a null model was fitted for each year to identify the most appropriate choice of regression model: linear or multilevel. The HLM was put forward for all years.

The intraclass correlation coefficient (ICC) was estimated to indicate the percentage of total variance explained by the programs. In 2016, Enade students' ICC was 0.13, which suggests that 13% of the Enade score variability is explained by the variability between the undergraduate programs in Accounting. The subsequent ICC figures were as follows: 0.15 in 2009, 0.11 in 2012, and 0.14 in 2015. Such percentages of school effect in null models are consistent with previous studies focusing on Brazil ((Ferreira, 2015; Liu, 2011a; Miranda et al., 2015; Rodrigues et al., 2017). Besides, the null model showed good reliability in all years (above 0.85). Reliability for level-1 intercept is measured through the ratio of parameter variance to total variance and is the mean reliability of all level-2 units in the model. These results suggest that the hierarchical model should be used to isolate the effects within and between Accounting programs in Brazil.

The final model was determined in two stages. First, the student variables were introduced, except for those that were non-significant in each year. Table 27 shows the regression results. Among the independent variables, age is the only continuous one; all the others covariates are categorical.

Table 27. Two-level fitted model introducing student variables

Explanatory Variables of the Model – Student Level		2006	2009	2012	2015
Fixed Effect			Estimates		
Intercept		34.76	40.09	35.47	49.42
Age		-0.03	-0.13	-	-0.07
Gender		-1.86	-2.75	-1.67	-2.82
Marital status		-1.12	-1.07	-1.08	-0.69
Ethnicity		-	0.5	1.09	1.02
Income		-2.14	-3.15	-2.61	-3.74
Mother's level of education		-	-0.86	-0.61	-
Father's level of education		0.38	-	-	-
Gainful occupation		1.49	1.12	-	0.96
Student funding		-	2.01	-	1.79
Allowance to continue studies		-	-2.58	2.47	-1.01
High school institution		-0.39	-	-0.95	-1.6
Type of high school		-	-0.47	-0.33	-0.74
Number of books read		0.87	0.55	0.93	0.71
Hours of study		1.29	2.27	2.6	1.8
Affirmative policy		-	-	-	0.42
Academic activity		-0.49	-0.98	-0.89	0.87
Random effects					
Intercept variance		12.08	24.17	17.81	21.69
Residual variance		84.88	148.79	148.59	138.06
Intraclass correlation		0.12	0.14	0.11	0.14
Number of observations (students)		16898	21708	44203	54860
Number of groups (undergraduate programs)		515	610	739	817
Goodness of fit	AIC	123687	115767	347719	427281
	BIC	123826	115926	347893	427728

Source: research data.

Finally, a complete model was produced by including explanatory variables related to programs and institutions. After fitting the model by excluding non-significant variables, the final model was obtained for Enade application year (see Table 28). Blank cells stand for non-significant variables in the final model.

Table 28. Two-level fitted model, with explanatory variables at both levels

Explanatory Variables of the Model – Student Level		2006	2009	2012	2015
Fixed Effect			Estimates		
Intercept		34.67	35.66	23.56	48.32
Age		-0.03	-0.07	-	-0.06
Gender		-1.88	-2.26	-1.72	-2.8
Marital status		-1.1	-0.96	-1.09	-0.71
Ethnicity		-	-	0.97	1.01
Income		-2.15	-3.17	-2.52	-3.67
Mother's level of education		-	-0.73	-0.6	-
Father's level of education		-	-	-	-
Gainful occupation		1.39	1.08	-	0.95
Student funding		-	1.55	-	1.86
Allowance to continue studies		-	-2.56	-2.38	-1.1
High School Institution		-0.48	-	-0.95	-1.6
Type of high school		-	-0.47	-0.33	-0.74
Number of books read		0.86	0.55	0.93	0.71
Hours of study		1.29	2.27	2.6	1.8
Affirmative policy		-	-	-	0.42
Academic activity		-0.48	-0.88	-0.75	0.77
Administrative category		-4.23	-6.12	-	-4.83
Academic organization		1.5	1.42	1.33	1.05
Region of the undergraduate program		2.3	2.99	1.38	1.48
Program infrastructure		0.78	0.58	0.03	1.59
Pedagogical teaching structure		-0.85	0.97	0.09	-1.68
Opportunity for further training/learning		0.66	-	-	0.91
% of faculty members with a master's degree		-	0.54	0.02	0.6
% of faculty members with a doctoral degree		-	-	-	0.52
Employment contract of faculty		-	0.34	-	-0.27
Random effects					
Intercept variance		7.98	12.04	14.33	16.12
Residual variance		84.71	148.61	148.56	137.93
Intraclass correlation		0.09	0.07	0.08	0.1
Number of observations		16898	21708	44203	54860
Number of groups		515	610	739	817
Goodness of fit	AIC	123858	171249	347843	427698
	BIC	123958	171376	347965	426868

Source: research data.

To complete the statistical analysis of variance as suggested by Laros and Marciano (2008), residual normality was tested to check if the errors were normally distributed and the assumptions were met, implying that the residuals tend to normal distribution. After meeting the statistical requirements, one specific objective remains to be answered, namely: what are the determinants of achievement amongst Accounting students in Brazil? Table 29 shows which variables were statistically significant at α=.05 in all Enade application years.

Table 29. Interpretation of the model results

Variable	Estimate	Result Interpretation
Gender	Negative	Lower female achievement compared to male achievement.
Marital status	Negative	Lower achievement of single students.
Income	Negative	Lower achievement of students from families with income of up to 6 National Minimum Wage.
Number of books	Positive	Higher achievement of students who read more.
Hours of study	Positive	Higher achievement of students who study more hours by themselves
Academic activity	Negative	Lower achievement of students who engage in research, extension, outreach or teacher assistance.
Academic organization	Positive	Higher achievement of students who attend institutions classified as superior to colleges (e.g., universities)
Region of the undergraduate program	Positive	Higher achievement of students who attend higher education institutions located in the South or Southeast of Brazil.
Infrastructure	Positive	Higher achievement of students who attend higher education institutions with the highest rates for infrastructure
Pedagogical teaching structure	Positive	Higher achievement of studies who attend higher education institutions with the highest rates for pedagogical teaching structure.

Source: research data.

It is interesting to note that students who participated in academic activity such as Extension Projects, Reserching Projects or Advisoring Activities showed a lower achievement than those who didn't participate in any of these activities. This result was consistent throughout 2006, 2009 and 2012, as shown in Table 28. Ferreira (2015) aimed to identify significant variables in explaining Enade 2012 results. The author suggested that students who participated in research and teacher assistance activities got worse score than those who didn't, while students who participated in extension activities got better grades.

In 2015, students were not asked if there were enrolled in these kind of activities or not, instead the questionnaire they answered asked if they got scholarship practicing any of the activities mentioned. Table 28 showed that this variable (Academic Activity) had a positive effect meaning that students who got scholarship based on involvement in a certain extra-curricular activity got better Enade score.

To analyze Enade as a measure of the quality of undergraduate programs in Business Administration, Primi et al. (2018) assessed data from over 71,000 freshmen and 58,000 seniors who took the 2006 Enade. They found that the differences between scores are significant before admission (prior knowledge) and that the greatest variability of Enade scores is within the programs themselves and is explained by the students' socioeconomic profile.

In fact, the students' individual characteristics have been consistently significant in explaining their final achievement (i.e., income, reading, hours of study, engagement in extracurricular activities) over the years, and these should be considered in any statistical model aimed at accounting for student achievement. Because the CPC ranking is based only on student achievement scores (Enade and IDD) and variables related to the programs' conditions of training and learning, it is actually reinforcing the difference that exists between HEIs in Brazil, perpetuating the unequal opportunities of access to higher education.

The IDD, which in theory should measure a program's contribution to the seniors' academic achievement, does not include the significant variables that do predict achievement. Moreover, because the IDD is included as an integral part of the CPC, its weight in assessing school effectiveness reduces substantially, to the point of being non-significant to the final ranking of undergraduate programs in the country, as shown in this dissertation.

While still appreciating the efforts made within the Sinaes to identify and measure contribution of the undergraduate programs to student achievement through the IDD, this dissertation adds to debates about estimating the programs' value added by not only focusing on determinants of achievement but also introducing them into a new proposal of IDD estimation. In addition to this theoretical contribution, this dissertation also aims to add to practical actions of undergraduate programs in Accounting by pinpointing which variables related to the IDD that program coordinators can use to influence their students in the search for better professional qualification and personal training.

The next chapter introduces this new IDD model proposal. It also discusses its impacts on ranking school effectiveness.

5 PROPOSAL OF MODEL IDD-VDCF[17]

The proposal of a new IDD estimation model (Model IDD-VDCF) for undergraduate programs in Accounting in Brazil follows the identification of the determinants of academic achievement on the Enade. Bailey and Xu (2012) contend that factors unrelated to achievement should be investigated if one is to provide comprehensive, unmistaken data on institutional effectiveness, especially when those factors are beyond institutional control. As students are admitted to undergraduate programs with varying degrees of academic skills and aspirations, these personal characteristics can affect their achievement, including their likelihood of completing their studies (Bailey & Xu, 2012).

Liu (2011a, 2011b) argues that in addition to accounting for student-related variables, it is necessary to control institutional variables, such as the selection process and the provision of graduate programs, which may affect student's final achievement. Steedle (2012) refutes this claim: in his study, the control variables (administrative category, historically black people-oriented HEIs, existence of graduate programs, selectivity, full-time enrolment, percent enrolled white, full-time retention rate, overall graduation rate, and student/faculty ratio) accounted for 10% of the variance in the achievement scores, but none of them was significant in the model after introducing the SAT variable at the institution level. The SAT compares to the Enem in Brazil, as both provide the students' score at high school: SAT exam is often taken at various points in high school and may also be taken multiple times while Enem is a more uniformly administered exam only taken at the end of high school.

5.1 Defining the Value-Added Model IDD-VDCF

The procedure included choosing and estimating the variables for each sample participant (Accounting students) as well as defining the adequate statistics to determine the magnitude and direction of the association between the variables. As the value-added measures available for use in the Brazilian context are based on student achievement measures, the Enade score was used as a proxy for achievement.

The model included the significant determinants of achievement in all Enade editions as explanatory variables, as shown in Table 27. However, the variables related to institutional policies, such as conditions of training and learning (infrastructure and

[17] The name of the proposed model refers to the existing Sinaes indicator (IDD) and to the name of the researcher who is proposing it (Vivian Duarte Couto Fernandes-VDCF).

pedagogical teaching structure), were not included because they were deemed as an integral part of the very program management policy. Hence, including such variables would undermine the actual value added of the programs. Similarly, the variable related to students' engagement in teacher assistance, research and extension was not included because the existence of such activities was also considered to be part of the institutional policy of the undergraduate program. Table 30 shows the variables included in each construct.

Table 30. Description of Variables in Model IDD-VDCF

Constructs	Variable Name (Acronym)	Variable Description
Student achievement	Enade (Enade)	Overall Enade score
Past achievement	Enem Score – Natural Sciences (CN)	Enem score in Natural Sciences
	Enem Score – Humanities (CH)	Enem score in Humanities
	Enem Score – Literature (LT)	Enem score in Literature
	Enem Score – Mathematics (MT)	Enem score in Mathematics
Students' sociodemographic characteristics	Gender (gen)	1 female, 0 male
	Marital status (stat)	1 single, 0 married
	Household income (inco)	1 up to 3,258.00 BRL, 0 above 3,258 BRL
	Number of books read in a year (book)	1 up to 3 books, 0 above 3
	Number of extra-class study hours (hour)	1 up to 3 hours a week, 0 above 3
Conditions of training and learning	Type of academic organization (type)	1 colleges, 0 universities and university centers
	Learning modality (moda)	1 in-person learning, 0 distance learning
	Region of undergraduate program (reg)	1 South or Southeast, 0 others

Source: research data.

Model IDD-VDCF is a value-added model that can be used to estimate how much higher education institutions contribute to the student's final achievement, considering their sociodemographic characteristics, their prior achievement, and the specific characteristics of their HEI. This is a multilevel mixed effects model (random intercept for the program, and fixed effects of the explanatory variables), where value added is the mean difference between observed scores and estimated scores for all students in a given program.

The hierarchical model tested for this purpose has two levels of analysis: Accounting students, and programs. At level 1, the students' prior achievement scores (Enem) and sociodemographic characteristics were used as explanations for the Enade score. At level 2, the conditions of training and learning in the Accounting programs were used as predictors of the level-1 intercept.

Thus, the equation that describes Model IDD-VDCF for undergraduate programs in Accounting in Brazil is given by:

Level 1: $$Enade_{ij} = \beta_{0j} + \beta_1 CN_i + \beta_2 CH_i + \beta_3 LT_i + \beta_4 MT_i + \beta_5 gen_i + \beta_6 sta_i + \beta_7 inc_i + \beta_8 book_i + \beta_9 hour_i + r_{ij} \quad (18)$$

Level 2: $$\beta_{0j} = \gamma_{00} + \gamma_{01} type + \gamma_{02} moda_j + \gamma_{03} reg_j + u_{0j} \quad (19)$$

where: $Enade_{ij}$ is the score of student i from the undergraduate program in Accounting j on the 2015 Enade; β_{0j} is mean Enade score for all students in program j; $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$ are the level-1 regression parameters; *CN, CH, LT* and *MT* are continuous variables for previous achievement (Enem scores); *gen* is the dummy variable for gender; *sta* is the dummy variable for marital status; *inc* is the dummy variable for income; *book* is the dummy variable for number of books students read a year; *hour* is the dummy variable for hours used for extra-class studies; r_{ij} is the residual score of student i in program j; γ_{00} is the overall mean Enade score of all undergraduate programs in Accounting in Brazil; $\gamma_{01}$is the difference in the mean Enade scores between programs in universities/university centers and programs in colleges; γ_{02} is the difference in the mean Enade scores between the in-person learning programs and distance learning programs; γ_{03} is the difference in the mean Enade scores between programs in the South and Southeast and programs in the other regions of Brazil; u_{0j} is the residual score of program j, pointing to a difference between program j and the overall mean Enade score.

5.2 Results of Model IDD-VDCF

Following Laros and Mariano (2008), the first decision was to test whether the multilevel model (HLM) would be the most appropriate to assess value added in Brazilian undergraduate programs in Accounting. To this end, a null model was tested, i.e., without explanatory variables (see equations 20 and 21).

$$\text{Nível 1: } Enade_{ij} = \beta_{0j} + r_{ij} \quad (20)$$

$$\text{Nível 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (21)$$

The results of the null model supported the use of the multilevel model, since the ICC was 0.17, i.e., 17% of the total variance in the Enade scores is explained by differences between the programs. The multilevel analysis is preferred when a significant amount of the variance in achievement (above 10%) is explained by school factors (Laros & Marciano, 2008).

The second stage was to assess the multicollinearity between the explanatory variables to ensure the model is accurate in only estimating parameters that are not highly correlated. This included assessing the Variance Inflation Factor (VIF): the higher the coefficient of determination, the higher the VIF indicating high collinearity between variables. Variables with VIF above 10 were excluded from the model (Hair et al., 2008). Table 31 provides the test result, showing that no variables should be removed a priori.

Table 31. Multicollinearity analysis of Model IDD-VDCF

Level	Variables	VIF	Level	Variables	VIF
Student	Gender	1.07	Program	Academic organization	1.60
	Marital status	1.14		Learning modality	1.68
	Family income	1.06		Regional location	1.14
	Books read a year	1.06			
	Extra-class study	1.06			
	Enem score – CN	2.12			
	Enem score – CH	2.12			
	Enem score – LT	1.95			
	Enem score – MT	1.60			

Source: research data.

Regression analysis was performed twice. The first regression served to estimate the parameters, observe the predicted values and estimate the standardized residuals. The second regression served to exclude outliers, i.e., those values with standardized residuals above | 3 |. The data reported in this dissertation were obtained in the model after excluding outliers.

The first step was to include the level-1 (student) variables. Table 32 summarizes the results of the fixed effects model with random intercept for the programs. At the student level, prior knowledge, as measured through ENEM scores (CN, CH, LT, and MT), as well as all sociodemographic variables tested in the model, were significant predictors of Enade achievement.

Table 32. IDD-VDCF model estimates with level-1 variables

	Estimate	Standard Error	Degrees of Freedom	t-value	p-value
Fixed Effects (Level 1)					
Intercept	-4.17	0.60	29630	-7.01	0.00
Enem CN	0.02	0.00	29630	18.94	0.00
Enem CH	0.02	0.00	29630	15.82	0.00
Enem LT	0.04	0.00	29630	29.21	0.00
Enem MT	0.02	0.00	29630	32.28	0.00
Gender	-1.64	0.12	29630	-13.61	0.00
Marital status	-0.43	0.13	29630	-3.34	0.00
Income	-1.09	0.13	29630	-8.50	0.00
Books	-0.28	0.12	29630	-2.40	0.02
Hours	-1.25	0.12	29630	-10.55	0.00

Source: research data.

The coefficient of determination (R^2), which measures the fit of the model, was 37.90. The intraclass correlation was 10.09%, which resulted from including the Enem-related variables. Liu (2011b) says that the ICC reduced from 15 to 10% on average when prior knowledge is controlled in the model. The Bayesian Information Criterion (BIC) was 226621, which is lower than the value in the null model, i.e., indicates a better fitted model.

The findings consistently show that the percentage of variance explained by the programs was reduced (ICC = 0.09) after introducing student- and program-level explanatory variables into the equations. The main reason is that the Enem scores accounted for a substantial amount of variation in the Enade achievement because of the high correlation between both Enade and Enem scores ($\rho = 0.55$). The adjusted determination coefficient of the final model was 38.05.

Table 33. Statistics of final Model IDD-VDCF

Fixed Effect		Estimate	Standard Deviation	t-value	p-value
Intercept		-9.82	1.08	-9.06	0.00
Enem CN		0.02	0.00	18.79	0.00
Enem CH		0.02	0.00	15.80	0.00
Enem LT		0.03	0.00	29.05	0.00
Enem MT		0.02	0.00	31.98	0.00
Gender		-1.67	0.12	-13.84	0.00
Marital status		-0.48	0.13	-3.71	0.00
Income		-1.05	0.13	-8.14	0.00
Number of books		-0.29	0.12	-2.48	0.01
Hours of study		-1.26	0.12	-10.58	0.00
Academic organization		1.01	0.26	3.82	0.00
Learning modality		5.02	0.92	5.48	0.00
Region of the undergraduate program		0.92	0.27	3.42	0.00
Random effects					
Intercept variance		9.52	3.08		
Residual variance		93.67	9.68		
Intraclass correlation		0.09			
Number of observations		30668			
Number of groups		911			
Goodness of fit	AIC	226475		Adjusted R^2	38.05
	BIC	226599			

Source: research data.

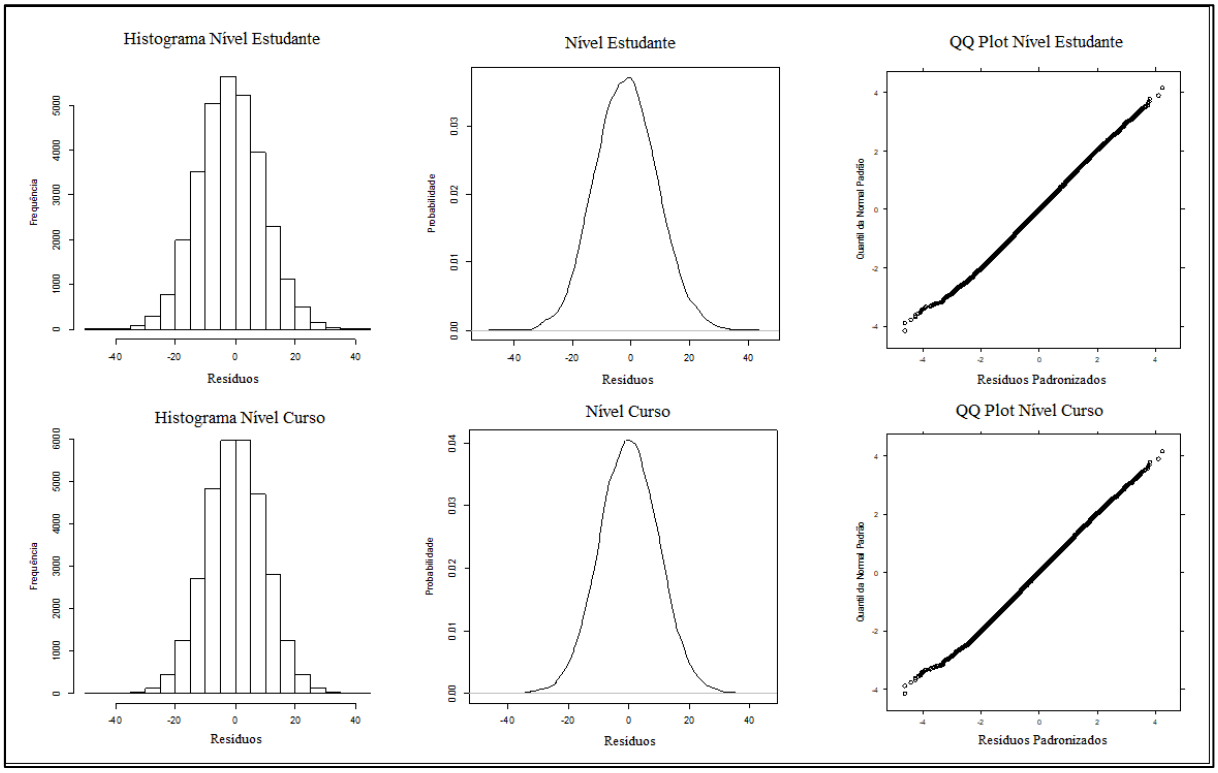
Table 33 shows that students who read up to three books a year have on average achievements 0.29 lower on the Enade than those who read above three books. Students who study up to three hours a week have on average 1.26 lower achievements on the Enade than those studied for three or more hours. Women perform lower than men (1.67 on average), and singles also score lower (0.48 on average). Students with family income of up to 3,258.00 BRL have an Enade score on average 1.05 lower than those with incomes that equaled or exceeded that amount.

Regarding institutional characteristics, students enrolled in programs in universities, university centers or federal institutes have mean Enade scores 1.01 higher than those peers in colleges. In-person learning programs students have Enade scores on average 5.02 higher than those students in distance learning programs. Students attending programs in the South and Southeast of the country have mean Enade scores 0.92 higher than the others.

To assess reliability, residual analysis was performed through graphs and statistical tests. The residuals have a normal distribution (see Figure 5), which was confirmed by a Kolmogorov-Smirnov normality test with p-value of 0.08 (i.e., above the significance

level of 0.05). In other words, the statistical test does not reject the null hypothesis that the residuals have normal distribution. The F test was also performed and confirmed homogeneity in the variance of the model residuals (p-value = 0.44).

Figure 5. Graphic analysis of residuals in Model IDD-VDCF



Source: research data.

Once the statistical requirements were checked for fitting the model, the results were analyzed for their practical implications. After identifying the best fit regression equation for the context of Accounting programs in Brazil, the IDC in Model IDD-VDCF was analyzed by estimating the difference between the students' actual achievements and expected achievements, considering their admission characteristics (Enem scores), their personal characteristics (gender, marital status, income, reading and study habits) and the conditions of training and learning (type of academic organization, type of education, and regional location of the program).

5.3 Comparison Between the Proposed Model and the Actual Inep Model

The correlation between the value added estimates as provided by the Inep model and Model IDD-VDCF is 0.94, which shows a high correspondence between the IDDs in both

models. Both models have significantly different means (α=5%), even though their correlation is high (see Table 34).

Table 34. T-test for dependent samples (Model 4 and Model IDD-VDCF)

	IDD_M4	IDD_VDCF
Mean	2.36	2.54
Variance	0.21	0.20
Pearson Correlation	0.94	
t-statistics	-35.94	
p-value	0.00	

Source: based on the research data.

Statistically, Model IDD-VDCF provides a better fit of the data compared to the Inep model, as it presents lower values of AIC and BIC, as shown in the analysis of variance between both models (see Table 35). In addition, the Likelihood Ratio Test (LRT) was performed, in what:

H_0: the simplest model (Model 4) fits as well as the Model IDD-VDCF

H_1: Model IDD-VDCF fits significantly better than Model 4.

Through the LRT (α=5%), the null hypothesis is rejected, that is, the IDD-VDCF Model presented the best fit. Furthermore, the coefficient of determination of the proposed model (R2 = 38.05) is higher than the coefficient of Model 4 (R2 = 37.05). Another indication that, statistically, the proposed model is more adequate to the analyzed database is the mean square of the error (QME), in which in the IDD-VDCF Model it was 91.9, while in the Inep model it was 97.6.

Table 35. Variance Analysis Between the Inep and IDD-VDCF Models

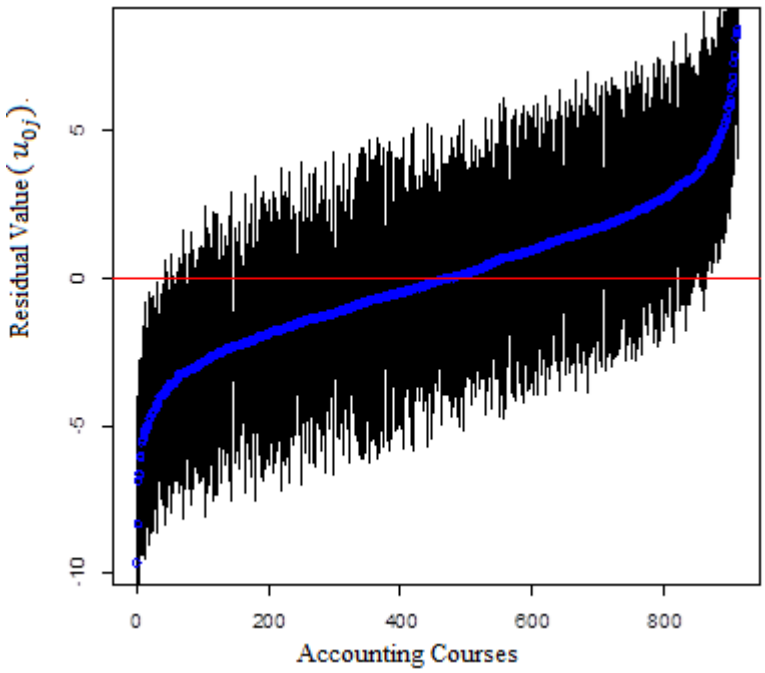
Model	DF	AIC	BIC	logLik	p-value
VDCF	15	228647.4	228772.3	-114309	
M4	7	229096.5	229154.8	-114541	<.0001

Source: research data.

Model IDD-VDCF is different from the current Inep model (Model 4) in that it includes explanatory variables for achievement at both level 1 (student) and level 2 (programs). At level 1, the terms $\beta_5(gen_i)$, $\beta_6(sta_i)$, $\beta_7(inco_i)$, $\beta_8(book\ _i)$ and $\beta_9(hour\ _i)$ were introduced in the model as predictors of achievement. At level 2, the terms $\gamma_{01}(type_j)$, $\gamma_{02}(moda_j)$ and $\gamma_{03}(reg_j)$ were also included in the equation.

At the program level, part of the variance between academic achievements can be explained by variables academic organization, learning modality and regional location, while another part of the variance remains unexplained as captured through the residuals (u_{0j}). Figure 6 shows the correspondence between the residual and the institution's ranking position based on the program value-added estimated in Model IDD-VDCF.

Figure 6. Effect of the undergraduate program on academic achievement in Model IDD-VDCF



Source: research data.

The findings draw attention to the social structure of undergraduate programs in Brazil. Even though such findings are representative only of the undergraduate programs in Accounting, they may be similar for programs in several other fields. Even though a model was developed that controls for the students' socioeconomic conditions, this research reveals that inequality persists in higher education. Despite the public policies of social inclusion headed by the Worker's Party government between 2003 and 2018, the most privileged social class has still dominated the best programs in the country.

Programs below the red line in Figure 6 contributed negatively to the Enade results, i.e., added less than expected in light of the students' Enem scores; programs with residuals above the red line contributed more than expected to Enade results in light of the students' Enem scores. As such, the higher a program's ranking, the better its institutional effectiveness (i.e., the higher the value added to student achievement). Yeh (2020) would consider this the demoralizing effect while Merton (1968, p. 3) refers to this as the Mathew

effect: “For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath”, i.e. those with more get more.

Yeh (2010), when studying K-12 education, identified that small differences between students' previous achievement (measured by standardized tests) are amplified by the structure of schools, not only perpetuating, but widening the achievement gap that exists between poor students and their wealthier colleagues. In this sense, Yeh (2020) argues that the application of standardized tests, classification and comparison of groups based on them serves to demoralize underperforming students, thus justifying the permanence of the achievement gap between low-income minority students and their more affluent peers. For the author, students should be assessed at their own levels, with specific assessment objectives allowing them to know whether they are advancing to higher levels of success. Thus, students would not lose their persistence and motivation to continue with their studies (Yeh, 2020).

Table 36 shows that the programs in the last decile of the ranking score better on the criteria of training and learning (infrastructure, IFF; pedagogical teaching structure, ODP; opportunity for further training/learning, OAF; faculty members with a master’s degree, Me; faculty members with a doctoral degree, Doc; employment contract, RT) and have students from more privileged socioeconomic backgrounds.

Table 36. Differences between first and last deciles in the ranking of undergraduate programs

Student-related Variables	1st Decile (Mean)	10th Decile (Mean)	Program-related Variables	1st Decile (Mean)	10th Decile (Mean)
Age	25.07	24.54	Enade (Continuous)	1.60	3.51
Enade	38.67	51.33	IDD released	1.7	3.43
Enade (FG)	51.47	57.87	Pedagogical teaching ODP	2.5	3.6
Enade (CE)	34.38	49.13	Infrastructure IFF	2.6	3.8
Percentage of singles	0.79	0.83	Opportunity for further training OAF	2.6	3.6
Percentage of white people	0.52	0.58			
Percentage of fathers with some level basic education	0.88	0.87	Master teachers Me	3.0	3.7
Percentage of mothers with some level of basic education	0.85	0.84	Doctoral teachers Doc	1.3	1.8
Percentage of incomes up to 3,258.00 BRL	0,70	0.61	work regimen RT	3.3	4.0
Percentage of first family members to reach higher education	0.86	0.58			
Books read a year	0.49	0.50			
Hours of study	0.36	0.45			
Enem (CN)	495.99	505.40			
Enem (CH)	527.40	535.49			
Enem (LT)	539.89	544.94			
Enem (MT)	555.71	571.27			

Source: research data.

Analyses should not focus only on the students' final achievement, as revealed in their Enade sores, or on a single quality indicator such as the CPC, which combines indicators and is difficult to interpret. Assuming that the purpose of programs, faculty, organization leaders and policy makers is to provide quality education that adds to the students' academic development, greater importance should be placed on analyzing a program's value added alongside student achievement. In other words, it is necessary to find out which variables are related to a program's value added (IDD), so that the program coordinators can provide better training to their students.

To this end, a correlation test was performed between the student- and program-related variables and the different IDD estimation models (Models 1-4 and IDD-VDCF), assuming that an understanding of the determinants of academic gain will support organization leaders in making practical decisions to improve the institutional effectiveness of their programs.

Table 37. Correlation between IDD and student characteristics

		IDD_M1	IDD_M2	IDD_M3	IDD_M4	IDD_VDCF
Marital status	ρ_{pb}	0.20*	0.20*	0.20*	0.04*	-0.01
Ethnicity	ρ_{pb}	0.05*	0.06*	0.06*	0.04*	0.01
Nationality	ρ_{pb}	0.02*	0.02*	0.02*	0.01*	0.00
Level of education (father's)	ρ_{pb}	-0.02*	-0.02*	-0.01*	0.00	-0.00
Level of education (mother's)	ρ_{pb}	-0.01*	-0.01*	-0.01	0.00	0.00
Income	ρ_{pb}	0.06*	0.06*	0.06*	0.03*	0.01
Work	ρ_{pb}	-0.00	0.00	0.00	0.02*	0.01
Funding	ρ_{pb}	-0.17*	-0.19*	-0.19*	0.01	0.00
Affirmative policy	ρ_{pb}	-0.04*	-0.04*	-0.04*	-0.00	0.00
High school institution	ρ_{pb}	0.01	0.01*	0.01*	0.01	-0.01
Type of high school	ρ_{pb}	0.12*	0.12*	0.12*	0.02*	0.01
First in the family in an HEI	ρ_{pb}	-0.02*	-0.02*	-0.02*	-0.01	-0.01
No. of books read	ρ_{pb}	-0.02*	-0.02*	-0.02*	-0.02*	.000
Hours of study	ρ_{pb}	-0.05*	-0.04*	-0.05*	-0.05*	-0.00
Enem (CN)	ρ	0.03*	0.02*	0.02*	0.04*	0.01
Enem (CH)	ρ	0.03*	0.01	0.01	0.03*	0.00
Enem (LT)	ρ	0.04*	0.04*	0.04*	0.04*	0.00
Enem (MT)	ρ	0.03*	0.05*	0.04*	0.06*	0.01*

ρ_{pb}. Point Biserial Correlation Coefficient; ρ Pearson Correlation Coefficient
* Correlation is significant at 0.05 (two-tailed).
Source: research data.

In Table 37, the IDD is significantly correlated with individual student characteristics in all Inep models (Model 1, Model 2, Model 3, Model 4) (i.e., marital status, ethnicity, parental education, income, type of high school, number of books read, hours of study). As a program's value added is correlated with the students' sociodemographic characteristics regardless of institutional efforts, it follows that the public policies of social inclusion are essential in Brazil. If the IDD conceptually is an indicator that measures "the

value a program adds to student achievement"[18] (INEP, 2019), it should not be related to the students' personal characteristics, as in Model IDD-VDCF. The proposed model has only one significant correlation with individual characteristic (Enem MT), suggesting that it does a good job (and better than the other models) of capturing programmatic differences not individual characteristics. Controlling for academic characteristics is crucial if the goal is to measure a program's contribution.

Program coordinators have little control over such sociodemographic variables, since public policies for social inclusion are in general at the level of the HEIs or the federal government. Therefore, attention should be drawn to the fact that the number of books read and hours of study are related to academic achievement. As programs with students who read more books and study longer outperforms the others, practical measures that encourage such activities can improve institutional effectiveness.

Correlation was also tested between the IDD in all models and the program-specific characteristics. The results are in Table 38. Not surprisingly, the program-related variables, such as the characteristics of faculty members' (doctoral degree, master's degree, and employment contract), are significantly related to the programs' contribution to their students' academic achievement. Since these are mutable factors, these correlations are fair and can offer insight on how to make programmatic improvements.

Table 38. Correlation between IDD and characteristics of the undergraduate programs in Accounting

		IDD_M1	IDD_M2	IDD_M3	IDD_M4	IDD_VDCF
Academic organization	ρ_{pb}	0.02	0.02	0.02	0.11*	0.03
No. of programs in the HEI	ρ	0.14*	0.16*	0.16*	0.24*	0.18*
IGC (continuous)	ρ	0.35*	0.34*	0.34*	0.42*	0.40
Learning modality	ρ_{pb}	0.11*	0.04	0.04	0.05	-0.03
ODP	ρ	0.35*	0.35*	0.35*	0.26	0.25
IFF	ρ	0.36*	0.38*	0.38*	0.31*	0.29*
OAF	ρ	0.35*	0.36*	0.36*	0.30*	0.27*
Faculty with a master's degree	ρ	0.12*	0.13*	0.13*	0.18*	0.16*
Faculty with a doctoral degree	ρ	0.07*	0.08*	0.07*	0.11*	0.10*
Employment contract	ρ	0.15*	0.13*	0.13*	0.16*	0.16*

ρ_{pb}: Point Biserial Correlation Coefficient; ρ: Pearson Correlation Coefficient
* Correlation is significant at 0.05 (two-tailed).
Source: research data.

In both Tables 37 and 38, the correlation coefficients in Model 4 are higher than those in Model IDD-VDCF because including control variables in Model IDD-VDCF (regional location, learning modality, and type of institution) reduced the effect of

[18] The author's translation to: "*o valor agregado pelo curso ao desempenho do estudante*".

institutional characteristics on academic achievement. This is another advantage of Model IDD-VDCF over Model 4; after all, policy makers want to identify which institutions are producing higher performers but reporting only differences in scores may be insufficient because of the significance of program-related variables.

Miranda (2011) assessed the relationship between Accounting students' academic achievement and teacher qualification. The author used the 2009 Enade scores as a proxy of achievement and extracted explanatory variables related to three factors (academic qualifications, professional qualifications, and pedagogical qualifications) from a questionnaire applied to undergraduate program coordinators of 218 HEIs in Brazil. Only the factor of academic qualification provided a significant regression coefficient for the Enade scores. However, the level of academic qualification was positively and statistically related to pedagogical qualification, and the variable related to professional qualification was correlated with the Enade scores.

The present data also show that pedagogical teaching structure (ODP), opportunity for further training/learning (OAF) and program infrastructure (IFF) are variables significantly related to a program's value added, i.e., the higher the rates in those variables, the greater a program's contribution to student achievement on the Enade. Cruz, Corrar and Slomski (2008) analyzed the relationship between student achievement and some pedagogical variables of undergraduate programs in Accounting. Based on data from 2002, they found that student achievement was higher when 1) the faculty members had more updated qualification for their respective courses, 2) the teachers used innovative teaching techniques, 3) the students carried out research as a learning strategy, 4) the courses were based on textbooks instead of handouts and summaries, and 5) the students had access to computer laboratories.

Thus, the present research data show that in order to add more value to student achievement, undergraduate programs in Accounting need to strive to provide quality infrastructure, well-designed pedagogical teaching structure, and offer scholarship for students to participate in research, extension and help them to manage their career.

5.4 Practical Implications of Model IDD-VDCF

Probably, the most appealing result for coordinators of Accounting programs, organization leaders and policy makers is the ranking of the Brazilian institutions' value added. So what is the impact of introducing control variables into the IDD model (students' sociodemographic variables, and variables of conditions of training and learning)? What

changes does it entail for discussions over Program Quality Level? What is the difference between the current Inep model (Model 4) and the model herein proposed (IDD-VDCF)?

The CPC is the major indicator used by the federal government to allocate public resources in Brazil, but it is not exclusive. Decree No. 7233, as of 19 July 2010, defines in Art. 4 the parameters to be considered for resource allocation: number of enrollments and number of admissions and graduations, diversity of programs, scientific production, number of patents, relationship between number of students and faculty members, number of graduate programs, extension projects, and results on the Sinaes assessments (Inep, 2010). As the CPC is the Sinaes indicator, its result does matter to organization leaders.

In describing the higher education financing system in Brazil, Calziani et al. (2018) report that public funding of public HEIs has not been based on the institutions' effectiveness or quality, but rather on their size, which generally relates to the magnitude of personnel spending. It comes with no surprise that Secretary of Higher Education Arnaldo Lima Junior told Brazilian newspaper *Folha de São Paulo* in October 2019 that the government is studying new ways of allocating resources to public institutions primarily based on their performance, rather than their size. According to the Secretary, currently 90% of the budget is dependent on the size of the HEIs while only 10% is related to their quality (Fernandes & Tomazelli, 2019).

In the private system, the government provides public resources through 1) student financing programs, such as Fies; 2) tax waiver programs, e.g., by discounting tuition fees from income taxes of individuals, and providing legal entities with tax exemptions in exchange for scholarships (Prouni), and 3) research funding through tax incentives for technological innovation (Calziani et al., 2018). In these cases, funding is granted to programs with a CPC above 3, while the remaining programs are subject to on-site audits and may be unauthorized to operate if the rate persists.

Table 39 shows the program ranking by CPC Levels in both the current Inep model (Model 4) and the proposed model (Model IDD-VDCF): three programs rated 1, 89 rated 2, 610 rated 3, 206 rated 4, and three rated 5. Although they seem to be similar models, they may produce rating differences that are significant for a particular institution.

Table 39. CPC comparison between Model 4 and Model IDD-VDCF

CPC	Model 4	Model IDD-VDCF
1	5	3
2	101	89
3	639	610
4	161	206
5	5	3

Source: research data.

When the quality indicator CPC is estimated, changing from Model 4 to Model IDD-VDCF produced no effect in the Level of 842 programs, while reducing the Level of six programs by 1 and increasing the Level of 63 programs by 1. More specifically, 16 programs would upgrade from Level 2 to 3, which would fend off on-site audits by the Ministry of Education, while 2 programs would have their CPC downgraded to Level 2, which would result in on-site audits and reduced public resources.

The two programs upgraded from Level 1 to 2 have mostly female students (68%), with income of up to 3,258.00 BRL, provide night classes, in-person learning, and are held in colleges with no graduate programs; 1 of them is in the Midwest region and the other one in the Southeast. Their IGC is 1.0 and 1.1. One has scores for ODP (2.3), IFF (0.5) and OAF (0.9) much lower than the other (3.1, 3.0, and 2.9, respectively), but higher scores for faculty (1.0 vs 0.4 for master's degree, 0.8 vs 0.0 for doctoral degree, and 0.8 vs. 0.0 for employment contract).

Students in the 16 programs upgraded from Level 2 to 3 are usually female students (63%), single (78%), and white or Asian descent (50%), and at an average age of 25; 93% and 89% of their fathers and mothers have up to secondary level of education, respectively; 78% have a family income of up to 3,258.00 BRL; 5% had a scholarship, and 18% were admitted through affirmative policies; 70% study up to three hours a week, and 54% read up to three books a year. All programs provide in-person learning at night, and only one is at a university providing graduate studies. The IGC of these institutions ranges from 1.6 to 2.4. The faculty score is low: on average 2.3 for members with a master's degree (SD 1.1), 0.7 for members with a doctoral degree (SD 0.5), and 2.8 for employment contract (SD 1.2). On average, their pedagogical teaching structure is 2.6 (SD 0.9), their infrastructure is 2.7, and their opportunity for further training/learning is 2.5 (SD 1.1).

Most of the 45 programs upgraded from Level 3 to 4 are private (37), and their IGC ranges from 2.2 to 3.8. Two of them are full-time programs, and the remainders are night classes programs. On average, their scores are 4.2 (SD 0.7) for faculty members with a master's degree, 2.4 (SD 1.0) for those with a doctoral degree, and 4.3 (SD 0.9) for

employment contract. These scores are surprising since private programs generally have less formally educated faculty (Table 13, page 71) and it emphasizes that faculty seems to be the key contributor to student achievement for this context. Their educational process scored on average as follows: 3.3 for ODP (SD 1.1), 3.6 for IFF (SD 0.9), and 3.3 for OAF (SD 1.0), which correspond to the averages of private HEI.

Two of the six CPC-downgraded programs moved from Level 3 to 2. Both are held in private colleges that provide nocturnal in-person learning. One is in a higher education institution with IGC of 2.2 and 98% of its students in undergraduate studies. The other is in an institution with IGC 2.3 and 100% of its students in undergraduate studies. The former has higher scores in the variables related to pedagogical teaching structure, infrastructure and opportunity for further training/learning (2.9, 3.2, and 3.3, respectively), but lower scores for faculty (1.3 for Me, 1.4 for Doc, and 2.6 for RT). Again, data shows the importance of faculty characteristics in measuring courses quality by indicators as the CPC. Most students are single (85%) and have parents with some secondary level of education at most (83% for mothers, 85% for fathers). Their average age is 28, and most read up to three books a year (56%) and study up to three hours a week (63%)

Comparing the impacts produced on the CPC by estimating value added through Model IDD-VDCF and Inep Models (Table 39) shows a similar behavior between the program-related variables for programs in equivalent CPC Levels. The same happened when comparing Inep Models 3 and 4: even though the IDD was significantly different between them, this difference did not produce significant statistical impacts on the CPC-based ranking of the programs, which may have been due to the weights assigned to the components of this indicator.

The underlying assumptions for the CPC weights are based on Fernandes et al. (2009), who used freshmen's and seniors' Enade scores in 2005, 2006 and 2007, years when the students had not taken the Enem. As the estimates of seniors' achievement scores were not controlled for prior achievement, Fernandes et al. (2009) considered the necessity of including other variables such as the conditions of training and learning (infrastructure, pedagogical teaching structure, opportunity for further training/learning, and faculty characteristics) to produce a more 'ideal' measure of Program Quality Level.

Besides, the educational context was different. The number of enrollments in higher education expanded 73.6% from 4,626,740 in 2005 to 8,033,574 in 2019 (INEP, 2016a; INEP, 2019). Increased spending at this level of education has been more significant from 2012 onwards, due to the expansion of programs in Science Without Borders and

Restructuring and Expansion of Federal Universities (REUNI – acronym in Portuguese) and the investment in distance learning. In addition, the government has also provided incentives to expand the private sector through programs Higher Education Financing Program (Fies) and University for All (Prouni) (Brasil, 2016).

Therefore, reanalyzing the weights of the variables that make up the CPC is necessary both for higher education in general and for undergraduate programs in Accounting in particular. Building on the data analyzed in this dissertation, it seems that the quality indicator CPC is in fact reinforcing the already existing difference between the programs as it uses the IDD and the conditions of the training and learning (ODP, IFF, OAF, Me, Doc, and RT) to estimate a single measurement of quality. As shown in Tables 37 and 38, the IDD is positively related to these characteristics. In other words, the higher the IDD, the higher the Program Quality Levels for faculty and infrastructure. Therefore, a suggestion is that Sinaes uses the IDD as a separate quality indicator from CPC, so that a program's institutional effectiveness can be compared with its conditions of training and learning, and this information helps organization leaders prioritize educational policies that do produce higher achievement amongst the students.

According to Jones (2018), producing a single achievement measure from various dimensions results in complexity in interpreting this measure since appropriate weights are necessary to combine the previously separate measures. Determining equal weights would be the simplest solution but combining different weights can be necessary depending on the purpose of that indicator (Jones, 2018).

In general, the weights are arbitrary for institution ranking (Kehn, 2014), causing discrepancies for a given institution or program across different systems (Usher & Medow, 2009), as reported by Jones (2018) for the different U.S. ranking systems. For this reason, Marginson (2014) contends that combining several indicators into a single measure may be questionable. In most countries, where no national assessment system is in place, different indicators attest to the quality of an undergraduate program, which comes back to the question of values and purpose of education discussed in Mitchell and Mitchell (2003).

In the U.S. undergraduate programs in Accounting, for example, quality can be gauged based on: the institution's accreditation by external members; its student admission system; its faculty credentials, including professional and academic certifications; the quantity and quality of its publications; the employability of its graduates; and the type of institution, since US private institutions tend to have fewer classrooms, which favors learning (Nagle et al., 2018). Regardless of the undergraduate program, quality in Brazilian higher education is

measured through the quality indicators produced by the National Assessment System for Higher Education (Sinaes).

The last specific objective of this dissertation is to discuss about the concept of quality in undergraduate programs in Accounting by drawing on how quality is perceived and assessed in the US. The choice of the US for analysis was based on the fact that its conditions of training/learning and history of Accounting education share similar characteristics with those in Brazil. Besides, this author had the opportunity to spend one year as a doctoral visitor at the University of Minnesota, US, and could collect data directly in the country. The next chapter is aimed to attain this objective.

6 DISCUSSING THE DIMENSIONS OF QUALITY IN ACCOUNTING

The growth of higher education in Brazil and worldwide has caused social unrest about what the so called " quality" of this level of education. The very educational institutions realized the urgency of quality to respond to a competitive environment, to meet the stakeholders' needs, and keep up with technological advances and the knowledge economy (Kundu, 2016). This discussion was driven particularly by events in the 1960s and 1970s, which led to increased spending on education and mandatory years of schooling, as well as the emergence of economic theories like the Theory of Human Capital. The quantitative boom in education spending in that period was followed by discussions about the "quality" of this investment in the following decade (Bertolini, 2009).

Some concepts are defined for the term in 1980s: quality is determined by the extent to which objectives are achieved (Groot, 1983), quality is fit for purpose (Ball, 1985 *apud* Elton, 1986), and quality is a measure of value added (Mcclain et al., 1989). At the beginning of the following decade, Harvey and Green (1993) published *Defining Quality for Assessment and Assessment in Higher Education*. The authors summarized the major aspects of the discussion and presented a categorization for the different ways of perceiving quality, which will be reported in section 6.1.

Recent studies also consider the quality of higher education based on the principles of total quality management (TQM). Mustafa and Chiang (2006) suggested an assessment framework that involves the following areas of higher education: faculty, staff, infrastructure, academic life, job management policies for undergraduates, curriculum, pedagogical teaching structure, and admission process. Also using the TQM as a parameter, Viswanadhan and Rao (2005) used the following parameters to measure the quality of distance learning programs: management, customer-orientation, communication, access to the program, learning environment, assessment and continuous improvement.

However, Srikanthan and Dalrymple (2002, 2003) criticize the TQM approach in the education system. They argue that TQM models are inadequate for educational services and that distinguishing the teaching-learning functions from the institution service functions is necessary when it comes to higher education. After all, quality certification and accreditation agencies are concerned with student learning in higher education.

The quality of educational programs at the various levels of education is the subject of debate and research worldwide (Freitas, 2012). In basic education, for example, the Programme for International Student Assessment (Pisa), coordinated by the OECD, aims to

produce indicators that contribute to discussing quality of education in over 65 participating countries. Since 1995, the international comparative study *Trends in International Mathematics and Science Studies* (TIMSS), coordinated by the *International Association for the Assessment of Educational Achievement* (IEA), has assessed the math and science knowledge of participating students from 58 countries. All these programs aim to interpret the differences between education systems to improve the teaching-learning process and to assist in decision-making on education in the participating countries.

In higher education, too, quality has been measured by accreditation systems or program assessment systems (Freitas, 2012). The higher education assessment process has gone beyond self-assessment reports by institutions or on-site assessment by a panel of experts; it includes the use of standardized student knowledge tests to assess undergraduate programs and their host institutions (Pedrosa et al., 2013). In 2008, the OECD launched a feasibility study, named Assessment of Higher Education Learning Outcomes (AHELO), to assess the possibility of developing international measures for higher education learning (Tremblay et al., 2012).

The goal was to create a comprehensive multidimensional, interdisciplinary and intercultural system that could assess whether students were really learning valuable knowledge and skills. Having a single system for different countries was aimed at comparability of student learning in higher education across the OECD member countries. However, only one pilot project was implemented given methodological difficulties and lack of political support and funding for such an audacious goal (Melguizo et al. 2017).

In the US, the program Collegiate Learning Assessment (CLA) measures the higher education students' problem-solving ability, analytical thinking, critical thinking, and writing skills. The aim is to provide two types of information to participating institutions: the difference in achievement between freshmen and senior, and whether this difference is larger or smaller than expected given the progress of other institutions (Klein et al., 2008). More recent proposals, such as the Voluntary System of Accountability (VSA), aims to measure educational outcomes using standardized tests based on the concept of value added, and therefore, respond to the public cry for evidence of student learning (Liu, 2010).

When it comes to undergraduate programs in Accounting, non-governmental organizations (e.g., Association to Advance Collegiate Schools of Business – AACSB, Association of International Certified Public Accountants – AICPA, and Brazilian Federal Board of Accountants – CFC) have developed specific procedures to measure Program Quality Level through assessment and certification processes, as well as exams to certify

professional quality. Like other fields of knowledge, Accounting education is driven to improve teaching and learning quality as a means to respond to the new dynamics of organizations and enable professionals to be ethical and reduce possible professional failures (FREITAS, 2012).

Nonetheless, the term "quality" in education is multifaceted and multidimensional in its conceptualization, assessment, and measurement, which means it is difficult to analyze from a single perspective (KRAUSE, 2012; KUNDU, 2016). To improve educational quality, the current level of this quality is a necessary measure, and higher education institutions have developed assessment and measurement tools to check their quality and improve their educational achievement. However, an HEI is not an industry per se, a company with well-defined and delimited processes, but a rather complex organization (VROEIJENSTIJN, 2003).

6.1 Dimensions of Quality

Austin (1980) distinguishes five different views on quality in education, namely:

1) mystical view: quality cannot be defined or measured given the complexity and ambiguity of higher education;
2) reputation-oriented view: based on agreement or consensus on the quality of a particular institution, i.e., if people agree that a particular institution is of high quality, then it is of high quality;
3) resources-oriented vision: based on the institution's physical, personal, and financial resources, i.e., the better the resources, the higher the quality of a given institution;
4) results-oriented view: based on the production approach, i.e., quality as a measure of an institution's products: students, publications, research funding, etc.;
5) value-added view: based on economic principles, i.e., quality as a measure of the benefits provided to students.

Harvey and Green (1994) claim that the different conceptions of quality in higher education can be interrelated and provide five dimensions for thinking about quality: quality as exceptional, quality as consistency or perfection, quality as fitness for purpose, quality as value-for-money, and quality as transformation.

Quality as exceptional "takes as axiomatic that quality is something special". This dimension can be found in three variations: 1) the traditional notion of quality as something distinct, which requires no measurement (i.e., its own existence comes with the idea of distinction), 2) the notion of excellence, whereby quality is related to high standards (e.g., if a university has the best tech lab, it is a top-notch university), 3) quality as confirmation of standards, which is an attenuated notion of excellence, with quality being present if some standards are exceeded, i.e., quality control is overcome (Harvey & Green, 1993, p.3).

In the dimension of quality as consistency or perfection, quality is no longer something unique, special, but rather something that can be achieved by everyone. In this dimension, quality is based on a set of specifications that must be met in products, processes or results. It is a concept that is also linked to the organizational culture, producing the idea that quality is the responsibility of all stakeholders (Harvey & Green, 1993).

Quality as fitness for purpose is an idea that is only possible if the conditions of product manufacturing or service provision are stipulated before delivery. In general, the purposes for the existence of something should be stated by the client; after all, there is no point in a perfect product if it does not meet the consumer's purposes (Harvey & Green, 1993). Hence, quality is measured in terms of the extent to which it achieves the proposed objectives.

Barnett (1992) argues that higher education is an educational process that requires an understanding of what quality is wanted for prior to conceptualizing it. According to the author, the conceptions of product and service are distinct, and therefore, it is not possible to establish generalizable standards in education. His fundamental questions are: "What are we trying to offer students in higher education?", "How would we know whether what we are offering is of high quality?", "How might we improve the quality of what is offered?" As such, the author resumes the discussion about the level of the purpose in higher education, which was later discussed economically by Mitchell and Mitchell (2003).

Elton (1986, p. 84) contends that the idea that quality is fitness for purpose is a simplistic view with the intention of using quality as a planning tool, i.e., to "establish priorities, allocate resources, open and close courses, and close departments and institutions". However, the author questions whether quality is an appropriate tool for such goals and points out that how people's work is assessed and measured impacts their attitude toward their work. Thus, the danger of higher education assessment is to use the results for non-work-driven purposes. It could also lead to a viscous cycle where formative evaluation is not helpful but reinforcing or harmful to the process.

In contrast, Harvey and Green (1993) make it clear that educational quality must be a concept concerned with the stakeholder, with students being the major stakeholders in education. From this perspective, when looking at quality as fit for purpose, one is thinking about the result of education for the end consumer, i.e., the student. Galiffa and Batalle (2010) contend that the quality of education is related to processes or outcomes (i.e., academic achievement).

Quality as value-for-money brings along the economic conception that quality equals its value. The key idea in this discussion is accountability over the investments. Thus, economic effectiveness is the imperative upon which return on investment needs to be measured. In higher education, effectiveness is seen in terms of control mechanisms (quality audit), quantitative results (achievement indicators), on-site assessment (of faculty members), and assessment of academic research (Harvey & Green, 1993).

Finally, quality as transformation is built on the assumption that there is a qualitative change, a change of form. As education does not provide a product but a service, such transformation is not doing anything for the consumer, but rather transforming the very consumer, empowering him/her. Empowerment involves providing the individual with enough strength to influence his/her own transformation. In education, empowering means: engaging the individuals in the decisions that affect their transformation, and the transformation process itself provides an opportunity for self-empowerment with consequent impact on the decision-making processes that affect these individuals (Harvey & Green, 1993). There are four ways to empower students: via student assessment; providing students control over their own learning; developing critical capacity; and providing minimum standards for them to track their own achievements (Harvey & Green, 1993).

Bertolin (2009, p. 133) say that three 21st-century ideas about the mission of higher education have influenced different trends in conceiving of higher education quality, namely: 1) economic competitiveness and market growth, 2) sustainable socio-cultural and economic development, and 3) social cohesion and equity (see Table 40).

Table 40. Perspectives of quality in higher education

Perspective of quality	Associated terms	Stakeholders	Purposes of Higher Education
Economist perspective	Employability and Efficiency	Private sector, OECD, and government	Emphasis on enhancing aspects of economic growth and employability.
Pluralistic perspective	Differentiation, Importance, and Relevance	UNESCO, European Union, and education sector	Diversity of relevant aspects (economy, socio-culture, democracy, etc.), with emphasis on the emergence of local specificities.
Perspective of equity	Equity	UNESCO and education sector	Emphasis on aspects contributing to social cohesion.

Source: Bertolin (2009, p. 134).

The economic perspective is consistent with the private enterprise, international multilateral organizations (e.g., the World Bank and the OECD), and neoliberal governments, and emphasizes the idea that investment in education is directly related to economic growth and indicators of employability. The pluralistic vision is led by international bodies (e.g., UNESCO – United Nations Educational, Scientific and Cultural Organisation) and the academia; it draws on a broader view of higher education that recognizes it not only as a driver of economic growth but also as a facilitator of human development, given its role in contributing to social cohesion, reducing the inequalities produced by the capitalist economic system (Bertolin, 2009).

As quality is a multifaceted term that should be interpreted in terms of purpose and context (Bertolin, 2009; Harvey & Green, 1993; Krause, 2012; Kundu, 2016), the Brazilian and American contexts of higher education in the field of Accounting were analyzed to understand how quality is perceived by government and professional boards.

6.2 U.S. Higher Education Programs in Accounting

Enrollments in Accounting programs have been growing in the US since 1993, peaking 253,082 students in 2014-2015 (see Table 41). Only from 1995 to 2000 is there a significant decrease (23%) in the enrollments.

Table 41. Enrollment trend in U.S. undergraduate programs in Accounting

Year	Bachelor's	Master's in Accounting	MBA in Accounting	Master's in Taxation	PhD	Total
1993-1994	189,863	10,114	3,398	5,942	1,002	210,319
1994-1995	192.870	10,010	3,510	6,510	870	213,770
1995-1996	174,010	9,430	3,360	4,760	770	192,330
1998-1999	134,050	8,620	3,770	2,600	770	149,810
1999-2000	127,960	9,455	2,445	3,130	680	143,670
2000-2001	134,775	10,375	4,000	2,395	800	152,345
2001-2002	133,435	12,565	4,065	3,555	890	154,510
2002-2003	141,175	17,540	5,270	3,550	1,085	168,620
2003-2004	142,735	18,795	4,030	3,595	955	170,110
2006-2007	173,299	21,253	4,482	3,239	1,095	203,368
2007-2008	181,075	24,065	4,166	2,305	1,224	212,835
2009-2010	187,534	29,464	5,100	2,822	1,188	226,108
2011-2012	201,570	29,648	5,132	2,840	1,189	240,379
2014-2015	207,071	39,641	3,107	2,068	1,195	253,082
2015-2016	216,482	29,428	3,014	3,177	946	253,047
2017-2018	207,806	27,482	3,208	2,645	732	241,873

Source: AICPA (2019, p. 6).

According to an annual AICPA report, one of the main factors related to the reduced number of enrollment from 1995 to 2000 was the image of the undergraduate program in Accounting amongst high school students, as they did not know the profession (Taylor Research And Consulting Group, 2000). This image was related to the traditional Accounting course provided in high school, which was oriented to bookkeeping and did not show the real role of an accountant in the society. This diagnosis led to the creation of the Accounting Pilot and Bridge Project to develop a high school curriculum in the US (Deines et al., 2012). For Simons et al. (2003), other factors for the decreased numbers of enrollments included the media coverage of scandals in the Accounting market (failures in audits, questionable Accounting practices, lack of internal control) and the technological changes in the business world since that time.

In the 2018 AICPA annual report, 51% of those enrolled in the undergraduate and master's programs in Accounting are male, 56% are white, and only 10% are international students. In 2018, 76,542 students graduated: 54,947 from undergraduate studies, and 21,595 from master's studies (AICPA, 2019). In 2014-2016, the number of enrollments in graduate programs dropped because of the recent economic recession in the country accompanied by higher tuition fees (NAGLE et al., 2018). However, unemployment is lower amongst certified professionals compared to other professionals, which has encouraged an increasing number of people seeking this type of advantage in the labor market (AICPA, 2019).

Table 42 shows that the number of graduates from both master's programs and undergraduate programs increased in the field of Accounting in the first decade of the 21st

century. However, it slightly decreased after 2013-2014, which led the AICPA, in partnership with the AACSB, to invest in projects to support students and newly graduated professionals. Both organizations have also invested in partnerships with companies to hire certified professionals. In 2018, 19,498 bachelors and 11,405 masters were hired by CPA firms, an increase of 22% and 275%, respectively, when compared to year 2000 (AICPA, 2019).

Table 42. Trend in the number of graduates from US undergraduate programs in Accounting

Year	Bachelor's Graduates	Master's Graduates	Total	Year	Bachelor's Graduates	Master's Graduates	Total
1993-1994	53,450	7,170	60,620	2003-2004	40,420	13,340	53,760
1994-1995	53,360	7,860	61,220	2006-2007	47,662	16,559	64,221
1995-1996	52,030	7,630	59,660	2007-2008	48,968	17,491	66,459
1998-1999	41,170	6,725	47,895	2009-2010	51,036	17,603	68,639
1999-2000	37,115	7,980	45,095	2011-2012	61,334	20,843	82,177
2000-2001	37,855	8,700	46,555	2013-2014	54,423	27,359	81,782
2001-2002	34,995	9,700	44,695	2015-2016	57,119	22,405	79,524
2002-2003	37,010	12,665	49,675	2017-2018	54,947	21,595	76,542

Source: AICPA (2019, p. 13).

In the US, professional boards have a close relationship with the academia and the marketplace. The quality of business programs is directly related to the external certification earned by the higher education institutions (Barilla et al., 2008; Nagle et al., 2018). Regardless of other certification programs, such as the Accreditation Council for Business Schools and Programs (ACBSP) and the International Assembly for Collegiate Business Education (IACBE), the Association to Advance Collegiate Schools of Business (AACSB) is the main association responsible for this certification in the US. Its team of raters awards the U.S. institution the status of members or non-members of the association (Barilla et al., 2008; Nagle et al., 2018; Nelson et al., 2006). Therefore, in the U.S. the quality of Accounting programs is defined by a set of parameters and specifications, with quality programs being those that meet these standards. This is what Harvey and Green (1994) called quality as consistency or perfection.

6.3 The Role of Professional Boards in Determining Program Quality Level

Two U.S. organizations deserve attention in discussing the quality dimension of the Accounting programs in the country: the AACSB and the AICPA. The former is the association of business schools while the latter is the association of certified Accounting professionals.

The AACSB is a nonprofit organization that was founded in 1916 with members from 17 U.S. universities, including Harvard, Ohio, Chicago, and Texas. It has been committed to the quality of business education and provides the necessary requirements for association: basic curriculum elements; balance between courses in the fields of business, law, and engineering; program management and coordination; cooperative projects between business schools and high school. In 1932, the AACSB specified five levels of education in the business area for HEIs: (i) four-year undergraduate program; (ii) professional program for over four years leading to first degree in business; (iii) professional course of one or two years based on earlier business degree; (iv) research courses of three years or more, usually the doctorate program; and (v) evening courses of fewer than four years beyond high school (AACSB, 2019c).

In 1951 the first AACSB statistical report on higher education in business was released by then-president Richard Kozella. In 1954 the first conference was held that discussed teacher quality in business. In 1968 the first university outside the US was accredited: the University of Alberta. In 1974 students from accredited schools were required to take classes in international business. In 1980 specific accreditation was adopted for Accounting (AACSB, 2019c). Since then, certification standards have been revised to ensure quality and continuous improvements in business education in the US.

To obtain and maintain accreditation, a business school must meet 15 criteria involving: mission and strategic management support for students, faculty and staff; teaching-learning process; academic and professional commitment of students and faculty members. Specific accreditation criteria were renewed for Accounting programs in April 2018 and included six standards organized into three categories: strategic management and innovation; teaching and learning; and academic and professional engagement. According to the institution, the main goal of certification is

> Encourage leading business schools and accounting academic units that voluntarily hold themselves accountable for improving business and accounting practice through scholarly education, engagement, innovation and impactful intellectual contributions. AACSB achieves this purpose by defining a set of criteria and standards, coordinating peer review and consultation, and recognizing high-quality business schools and accounting academic units that align with the standards and participate in the process. (AACSB, 2018, p. 3, underline added)

Engagement is about training the students as citizens and professionals, as well as developing their personal characteristics to stand up to life. Innovation refers to the fact that accredited schools must pursue new experiments and entrepreneurship, setting strategic

innovation as a priority. Impactfulness is about having the schools make a difference using their own high quality inputs (human, physical, and financial) and show how they measure this impact.

To be certified, an HEI must follow the AACSB protocol, starting with the eligibility criteria, aimed to: specify the set of AACSB governing values and principles that the proponent agrees to share; explain how the proponent will be assessed, especially the conditions and the necessary support on the proponent's part for Accounting education and its connections with the market; and address basic characteristics that affect the a program's quality of teaching, research and other activities (AACSB, 2018). Table 43 provides the criteria.

Table 43. Eligibility Criteria for Accreditation in the Business Area

Pool of values	Criterion	Definition	Measurement
Core Values and Guiding Principles	Criterion A: *Ethical Behavior*	The accounting academic unit must encourage and support ethical behavior by students, faculty, administrators, and professional staff.	Evidence of programs, policies, procedures or other activities in force to support the organization's ethics.
	Criterion B: *Collegiate Environment*	The accounting academic unit maintains a collegiate environment in which students, faculty, administrators, professional staff, and practitioners interact and collaborate in support of learning, scholarship, and community engagement.	Evidence of collaborative activities, multi-representational decision environments, student manuals, etc.
	Criterion C: *Commitment to Corporate Social Responsibility*	The accounting academic unit must demonstrate a commitment to address, engage, and respond to current and emerging corporate social responsibility issues through its policies, procedures, curricula, research, and/or outreach activities.	Policies, activities, procedures, research that show concern about issues involving diversity, sustainability, prosperity, and globalization.
General Criteria	Criterion D: *Accounting Accreditation Scope and AACSB Membership*	The institution must be part of an accredited business school or one which is in process to be accredited. The academic entity may be defined as an institution authorized to award bachelor's degrees or higher (in business and accounting) or a business academic unit within such an institution.	Evidence of affiliation or current process to become affiliated.
	Criterion E: *Oversight, Sustainability, and Continuous Improvement*	The accounting academic unit must be structured to ensure proper oversight, accountability, and responsibility for accounting academic operations; it must be supported by continuing resources and it must have policies and processes for continuous improvement.	Documents describing the management, budget, resources, and policies for continuous improvement.
	Criterion F: *Policy on Continued Adherence to Standards and Integrity of Submissions to AACSB*	All AACSB accredited programs must follow accreditation standards and provide ongoing accreditation review.	Documents proving accreditation.

Source: AACSB, 2018.

Accounting-specific accreditation standards make it clear that a high-quality program is one that can prepare students to meet the needs of society through the Accounting courses and throughout the profession; their quality indicators include: job placement and career success of the graduates (BARILLA et al., 2008). Table 44 shows the AACSB standards required to accredit Accounting schools.

Table 44. Specific Standards for Accounting Certification

Level	Variable	Measurement
Strategic Management and Innovation	Standard A1: Accounting institution's mission, impact, and innovation	Foundational documents regulating the undergraduate program
	Standard A2: Accounting Intellectual contribution's impact and alignment with mission	Report from the last five years of scientific publications and institution-wide project reports
	Standard 3: Financial strategies and allocation of resources	Institution's Accounting report presenting its resources and strategy to maintain them
School and its stakeholders	Standard 4: Accounting curricula content, management and assurance of learning	Educational project of the program (curriculum, learning objectives)
	Standard 5: Information technology skills, agility and knowledge for accounting graduates and faculty	Integration of the educational project with information systems, data analysis, and development of information technology.
	Standard 6: Accounting Faculty Sufficiency, Credentials, Qualifications and Deployment	Teaching strategies, relationship between faculty members and market professionals, employment contract, integration between faculty members with different credentials, skills and academic, and professional titles.

Source: AACSB, 2018.

According to the organization itself, the AACSB's role is to promote and develop higher education programs in the field of business (Administration and Accounting). Its mission is to "foster engagement, accelerate innovation, and amplify impact in business education" (AACSB, 2019b). To this end, AACSB is devoted to aligning organizations, schools, faculty members and students so that business education is consistent with market demands. The ultimate expectation is that business education contributes to global prosperity (AACSB, 2019).

The following figures provide a glimpse of the associations' role in the business world: 1) 96% of CEOs on the 2016 Financial Times "From MBA to CEO" graduated from an AACSB-accredited school, 2) 90% of the top MBA programs listed in *The economist* are accredited by the institution, 3) over 3 million students worldwide are enrolled at accredited schools (AACSB, 2019a), 4) currently 831 HEIs in 54 countries are accredited, and 187 institutions have Accounting-specific certification. In Brazil, three schools are AACSB certified in the area of Administration: Fundação Getúlio Vargas São Paulo, Fundação Getúlio Vargas do Rio de Janeiro, and Insper – Instituto de Ensino e Pesquisa (AACSB, 2019a). No Brazilian institution is certified in Accounting.

Assessment can be deemed as a measure of effectiveness of training. Schools seek third-party certification to show their stakeholders that they are providing quality education to their students (Barillha et al., 2008) while professionals seek certification to prove to the market that they are able to perform their duties skillfully (Nagle et al., 2018). In addition to its importance in Accounting training, the AACSB is directly involved in the certification of

new business professionals in partnership with the American Institute of Certified Public Accountants (AICPA).

Created in 1887, the AICPA is responsible for defining the rules and standards governing the work of an accountant. The AICPA is legally responsible for the profession before legislative bodies, public stakeholders, and other professional organizations. It is also directly associated with the promotion of Accounting education. Its activities include: defining standards for audits of private companies, providing informational material for members and non-members, developing and grading the CPA Exam, and monitoring the technical and ethical standards of the Accounting profession (AICPA, 2019). However, most of its activities revolve around the professional accountant certification process through the CPA Uniform exam.

The CPA Exam is applied through a contract between the National Association of State Boards of Accountancy (NASBA), the AICPA and the test administration company Prometric. The NASBA serves as the information center to which Accounting boards in all states send information about eligible exam candidates and from which advisory information and test data are reported to the boards. The AICPA defines the test content, creates questions and determines the scoring method, in addition to conducting statistical analyses of test results (NASBA, 2018).

The AICPA's proposal is to ensure the quality of services provided by its members, which is of public interest. It has developed absolute control over those who may or may not practice within 54 U.S. Jurisdictions (the exam is not applied in the Commonwealth of the Northern Mariana Islands only) (Bline et al., 2016b). The CPA is a certification that enables the accountant to work in the field of auditing and taxation in the public sector; it is, in fact, a condition for a career in this sector. While several benefits of this certification are not clear, improved salaries and career advancement are certainly two of them (NAGLE et al., 2018).

The CPA certification is the most valuable certification in the U.S. accountant career (Trinkle et al., 2016). Nelson et al. (2006) carried out a study on the academic characteristics of students from institutions accredited by the U.S. Federation of Schools of Accountancy (FSA) from 1991 through 2006. Approximately 80% of students said they would take the CPA Exam, while 80% said they were not interested in other certifications, including Certificate Management Accountant (CMA), Certified Internal Auditor (CIA), and Certified Information Systems Auditor (CISA).

To become a certified public accountant, the applicant must bear in mind the three certification Es: Education, Examination, and Experience. Each of the 54 U.S. Jurisdictions defines its rules regarding citizenship, residence and social security number of the U.S. candidates, and may or may not accept applicants from international programs. For incoming applicants to have their certification, they must meet the requirements of one of the U.S. Jurisdictions.

The first criterion (Education) for certification is to have a minimum of 150 hours[19] in business courses: 120 hours in undergraduate courses, and 30 hours in graduate courses (AICPA, 2017). With all the advantages of being a certified public accountant, most students try to obtain the 150 hours within four years to devote the last year to this certification (Nagle et al., 2018). Depending on the U.S. state, there is a minimum time requirement for Accounting and Administration courses, and requirement for Accounting ethics-specific courses (Shelton et al., 2012).

The second criterion (Examination) is passing the Uniform Certified Public Accountant Examination (CPA Exam). The test consists of 4 subtests, namely: auditing and attestation (AUD), business environment and concepts (BEC), financial accounting and reporting (FAR), and regulation (REG). Applicants must achieve a minimum score of 75 out of 100 in each subtest.[20]. The questions are multiple choice, simulation, or written communication (AICPA, 2017). The aim is to assess the applicants' knowledge of Accounting principles, management, concepts, and regulatory practices (Shelton et al., 2012). In considering the Bloom Taxonomy, all four subtests require more high-order skills (apply, analyze, assess, and create) than low-order skills (remember and understand), because the work of an accountant has increasingly required complex skills and thinking (Ugrin & Honn, 2018).

The last criterion (Experience) requires a minimum time of work in the profession, which may vary from one to two years depending on the state. The requirement may also vary for the type of work (audit or tax professional) and type of employer (state or private company) (AICPA, 2017). The exam has been online only since 2004 and can be taken in several parts of the world. It is called uniform because it contains the same content

[19] The state of Florida was the first to require 150 hours of study for certification in 1983. Since then, other states have included 30 hours of graduate studies in their requirements. As of 2019, only Virgin Island requires 120 hours of study, but it has already established a minimum of 150 hours for applications from May 2020 onwards.

[20] For further information about how the score is measured, please refer to "CPA exam scoring" available at https://www.aicpa.org/becomeacpa/cpaexam/psychometricsandscoring.html

regardless of where it is taken, including in Brazil. This ensures fair, consistent assessment, with all test topics widely publicized by the AICPA (Shelton et al., 2012).

The CPA Exam is the most widely accepted exam in academia and the market, although the demand has grown for Certified Management Accountant (CMA) by the Institute of Certified Management Accountant (ICMA). Hargadon and Fuller (2018) argue that these certifications are a major advantage in the labor market, where a double certified professional tends to earn 59% more than non-certified professionals.

Recruitment of accountants is also closely related to CPA certification, and companies offer benefits to employees that accomplish certification within a short period of time, including bonuses and job promotion (Nagle et al., 2018). At KPMG, for example, a newly hired professional earns a US$ 5,000 bonus or a US$ 3,000 bonus if s/he becomes CPA within the first or the second year, respectively (Ovaska-Few, 2016).

For these reasons, several US business schools market their students' CPA pass rate as a way of showing off their quality to the market. However, it begs the question whether the specific characteristics of these programs are truly related to their exam pass rates (Bline, Perreault, & Zhen, 2016b).

6.4 Determinants of Academic Achievement in the US

Table 45 shows the number of CPA applicants has increased over the years; for example, by 13% from 2015 to 2016. In 2017 and 2018, this number reduced mirroring the data number of graduates in Table 42. The CPA pass rate, approximately 60%, has motivated research on the determinants of success on this exam.

Table 45. CPA exam pass rate in the US

Year	New candidates	Candidates who passed	Pass rate
2006	36,078	14,101	39.08
2007	38,513	18,444	47.89
2008	40,839	23,190	56.78
2009	44,204	26,163	59.18
2010	49,597	29,539	59.55
2011	39,083	23,418	59.91
2012	42,157	26,020	61.72
2013	42,962	26,911	62.64
2014	41,266	25,558	61.93
2015	42,653	26,077	61.14
2016	48,004	27,889	58.10
2017	39,630	25,514	64.38
2018	36,827	23,941	65.00

Source: based on AICPA (2017, 2019).

In addition, because of the close relationship of the CPA Exam with academia and the market, studies have been developed in recent years to identify which are the predictors of better academic achievement (Barilla et al., 2008; Bline et al., 2016a; 2016b; Boone et al., 2006; Nagle et al., 2018; Trinkle et al., 2016; Ugrin & Honn 2018), similar to this research analysis on what are good predictors of Enade scores in Brazil. Table 46 summarizes the variables that explain students' achievement on the CPA.

Based on the sociodemographic data from 2005 to 2013, Trinkle et al. (2016) carried out a statistical analysis with time needed for passing the CPA (upon meeting the requirements) as the response variable. The authors noted that the highest pass rates are among the youngest men who attended an AACSB-certified school (in Accounting) in a private university. They also observed that certified professionals (CPA) earn the best salaries and occupy the best jobs.

Table 46. Studies on the determinants of academic achievement in the US

Level	Significant Variables	Studies
Student	Sociodemographic data (gender, ethnicity, income, marital status, parental education)	All studies address at least one of these variables
	Previous knowledge	Ugrin and Honn (2018)
	Admission score	Bline et al. (2016a)
Educational Institution	Selection process	Boone et al. (2006), Nagle et al. (2018), Miller and Nouri (2015)
	On-site or distant education	Shelton et al. (2012)
	Number of course hours	Allen and Woodale (2006), Trinckle et al. (2016)
	Certified schools vs. non-certified schools	Nagle et al. (2018), Barilla et al. (2008), Bline et al. (2016a), Boone et al. (2006), Mittelstaedt and Morris (2017), Miller and Nouri (2015), Trinckle et al. (2016)
	Type of institution accreditation (Administration or Accounting)	Blaine et al. (2016), Nagle et al. (2018), Miller and Nouri (2015), Trinckle et al. (2016)
	Public or private HEI	Mittelstaedt and Morris (2017), Trinckle et al. (2016)
	Faculty productivity	Bline et al. (2016a), Nagle et al. (2018)
	Faculty certifications	Bline et al. (2016a), Nagle et al. (2018)

Source: the author.

Ugrin and Honn (2018) used a database of 1005 applicants to assess the relationship between prior achievement and passing the CPA Exam. They found that the applicants with achievements above the mean (who attended university-equivalent courses) had greater success on the test than those who did not take advanced courses. Applicants with the highest scores on the ACT/SAT[21] were the best performers both in high school and on the CPA Exam. The research is empirical evidence of recommendation # 4 in the Pathways Commission report – adopting measures to attract high-level students to the Accounting

[21] The ACT and the SAT are standardized tests used for admission to U.S. universities.

profession (Pathways Commission, 2012) – as identifying outstanding high school students can be an important factor in recruiting outstanding Accounting professionals.

Similarly, other U.S. studies have related the CPA exam pass rate to the US university admission system, which uses as a rule the scores in standardized tests such as the SAT and the GMAT.[22] Bline et al. (2016b) and Grant et al. (2012) found a significant positive relationship between student achievement on the CPA and their SAT and GMAT scores on university admission. Boone et al. (2006), Miller and Nouri (2015) and Nagle et al. (2018) also identified that the institutions with the most rigorous selection processes were those with the highest CPA exam pass rates.

Recent studies have set out to identify if student achievement differs in online and in-person tests (MILLER; NOURI, 2015; TRINKLE et al., 2016). In investigating disparities between CPA certification requirements across the US states, Shelton et al. (2012) found a significant difference between the achievements of students in distance learning and in-presence learning.

Some studies have related institutional characteristics to achievement on the CPA exam. Some of them have compared student achievement differences on the CPA exam against the type of school accreditation (which may be in Business or Accounting) (Nagle et al., 2018, Bline et al., 2016b; Miller & Nouri, 2015; Trinkle et al., 2016). Others have investigated whether there is difference in the achievement of students from AACSB-certified schools and students from non-certified schools (Nagle et al., 2018; Barilla et al., 2008; Boone et al., 2006).

Barilla et al. (2008) analyzed data from 1985 to 2003 and noted that institutional (Accounting-specific) certification by AACSB contributes to better CPA exam pass rates, while overall business certification by AACSB alone is not capable of explaining the difference in exam pass rates compared to other schools. Miller and Nouri (2015) also found similar results.

Nagle et al. (2018), too, analyzed data at the institutional level, but only for those who took the CPA Exam for the first time. The authors claimed that the results of first-time test takers are important because they guide the educational policies of undergraduate schools of Accounting in the US. They reported that having a graduate degree is directly related to student achievement on the CPA exam. The significant institutional variables affecting the CPA exam pass rate were: business school accreditation, Accounting school accreditation,

[22] The GMAT is a standardized admission test for U.S. graduate programs in business.

student admission process, and the percentage of CPA-certified faculty members. The non-significant variables were the percentage of faculty members with a doctoral degree, institution's ranking in research publication, and type of institution (public or private).

Several of these studies take into account the quality of the Accounting program – as measured through faculty productivity, admission process (typically based on the SAT and ACT scores), and/or course certification (AASCB) – as control variables to compare CPA exam pass rates. The faculty characteristics are also an integral part of the discussions about the variables related to best student achievement.

Nagle et al. (2018) found a direct positive relationship between the percentage CPA-certified faculty members and student achievement on the CPA Exam, while the percentage of faculty members with a doctoral degree was non-significant for such performance. The authors argued that while both professional and academic certifications of faculty are important indicators of school quality, values and priorities may vary between market-oriented and academia-oriented faculty members.

Their study showed that the faculty's professional experience and CPA certification can contribute positively to preparing students for the CPA. Bline et al. (2016b) came to the same conclusion: institutions with the largest number of professionally certified faculty members had the highest exam pass rates on the CPA Exam.

6.5 Accounting in Brazil

In Brazil, the first higher education school of Accounting and Actuarial Sciences was created by Decree-Law No. 7988, as of 22 September 1945 (Pelias et al., 2007) within the Faculdade de Ciências Econômicas e Administrativas (currently Faculdade de Economia, Administração e Contabilidade – FEA) at Universidade de São Paulo (USP). Despite the initial European influence, training in Accounting in Brazil has received strong influence from the American school since the 1960. This influence stems from the North American teaching model disseminated through a textbook of Introductory Accounting written by FEA-USP faculty members and strongly based on *Introductory Accounting* by Finney and Miller (Pelias et al., 2007).

The U.S. influence upon the Brazilian Accounting system started with the modernization of the university administration in Brazil from 1966 to 1972 based on draft agreements between the Brazilian Ministry of Education and the U.S. Agency for

International Development (USAID) (Couto, 1989). Such draft agreements included short-term courses to train faculty and staff of Brazilian universities.

Outgoing USP faculty members who completed graduated studies in U.S. universities opened graduate programs of strong American influence upon their return to Brazil FEA-USP professors who participated in the creation of the institution has had an international experience in the 1960s and 1970s, mainly in the US, at Columbus, Stanford, Harvard and Illinois (Ribeiro, 2009).

Brazil currently has 351,710 accountants and 167,272 Accounting technicians registered with the CFC (Conselho Federal Contabilidade, 2019). According to the Brazilian Register of Employed and Unemployed (CAGED – Cadastro Geral de Empregados e Desempregados), Accountants were one of the six most hired professionals in the country in 2018: nearly 17,000, with a monthly salary of 4,407.55 BRL on average (BRASIL, 2019). Salary may vary depending on company size and level of experience.

The Federal Board of Accountants (CFC – Conselho Federal de Contabilidade) is an autonomous entity incorporated under public law which was created by Decree-Law No. 9,295, as of May 1946. Article 6 in this decree (as amended by Article 76 in Act No. 12,249, as of 2010) sets out its duties: "organize and approve the By-laws defined by the Federal Board; address questions raised in the regional boards; decide on petitions to review penalties applied by the regional boards; publish annual report of its activities; regulate the Accounting principles, the Proficiency Exam, the pool of qualified accountants, and continued learning programs; and set forth the Brazilian Accounting Standards" (Brasil, 2010, Art. 6). In this document, the professional board is not responsible for training Accounting professionals, but only for their continued learning, i.e. CFC is not responsible for initial development, which is held by undergraduate programs, but they offer ongoing professional development.

Decree-Law No. 9,295, in Article 25, defines the technical assignments of an Accounting professional as follows:

> a) organize and execute Accounting services in general;
> b) keep all mandatory and necessary books and prepare the respective balance sheets and statements;
> c) provide judicial or extrajudicial inspections, review of balance sheets and accounts in general, checking of assets, permanent or periodical review of books, judicial or extrajudicial regulations of gross or common malfunctions, assistance to audit committees in corporations, and any other duties of technical nature as established by law. (BRASIL, 1946, Art. 25).

In Brazil, professional certification is relatively recent (from 2011 onwards). Since then, the exam is provided twice: in the first and second semesters of the year. As

shown in Table 47, the exam pass rate has not exceeded 45% of applicants in the last five years (Conselho Federal de Contabilidade, 2019).

Table 47. CFC exam pass rate in Brazil

Year	Exams taken	Exams passed	Pass rate
2011-1	13,383	4,130	30.86
2011-2	18,675	10,886	58.29
2012-1	24,474	11,705	47.25
2012-2	29,226	7,613	26.05
2013-1	33,708	12,000	35.60
2013-2	36,883	15,891	43.14
2014-1	38,116	18,824	49.39
2014-2	32,568	13,591	41.73
2015-1	38,023	20,715	54.48
2015-2	38,022	5,580	14.68
2016-1	41,987	17,576	41.86
2016-2	40,879	8,948	21.89
2017-1	46,949	11,860	25.26
2017-2	44,557	12,041	27.02
2018-1	43,864	13,231	30.16
2018-2	34,259	12,845	37.49
2019-1	36,150	12,626	34.93

Source: CFC, 2019.

Because of these low indicators, several studies have sought to understand the determinants of Accounting students' academic achievement based on the characteristics of students, institutions, and faculty members as explanatory variables. While in Brazil the proxy for academic achievement has been the score on the Enade, in the US it has been in general the CPA exam pass rate – two exams significantly different in nature.

This difference is indicative of how much the labor market and the public and private organizations are directly related to the HEIs in the US. The CPA certification as a proxy of achievement reveals that training for the labor market is the most appropriate measure of Program Quality Level in that society. In contrast, Program Quality Level in Brazil is more related to academic issues, as measured through the National Assessment System for Higher Education (Sinaes).

Also, a note is needed about data availability. In the US, the CPA results have been published by institution, while in Brazil CFC data by institution have only been published since 2017. In addition, the Enade is mandatory, and all students are required to take it every three years. Unlike in America, Brazilian researchers can obtain detailed student-level findings by retrieving students' sociodemographic characteristics from the Enade student questionnaire and combine them with data from the Sinaes reports of quality indicators.

Table 48 presents Brazilian studies on determinants of academic achievement that use the Enade score as a proxy for achievement. In general, the significant variables point to characteristics related to the students, the institution, or the faculty. At the student-level, Brazilian studies in the field of Accounting (Ferreira, 2015; Rodrigues et al., 2016; Rodrigues et al., 2017; Miranda et al., 2015; Miranda et al., 2014; Santos, 2012) have found a significant relationship between demographic/socioeconomic variables and student achievement. Family income and parental education have been the most consistent variables in the studies, while the others have been inconclusive.

Table 48. Studies on the determinants of academic achievement in Brazil

Analysis level	Significant variable	Studies
Student	Sociodemographic characteristics (gender, ethnicity, marital status, income, parental education).	Most research mentions at least one variable
	Previous knowledge	Fernandes and Miranda (2018)
	Absenteeism and study habits	Correa et al. (2012), Miranda et al. (2017), Fernandes et al. (2018)
	Motivation, personal effort, anxiety, and stress	Cornachione Junior et al. (2010), Leal et al. (2013), Rezende et al. (2017)
	Types of learning	Nogueira (2012)
School	On-site or distance learning	Caetano et al. (2015), Rodrigues et al. (2016)
	Characteristics of the program coordinator	Rodrigues et al. (2017)
	Opportunity provided for students to participate in academic activities	Ferreira (2015), Fernandes et al. (2018)
Faculty members	Teacher credentials (pedagogical and professional)	Lemos e Miranda (2015), Miranda (2011), Santos (2012)
	Faculty knowledge and teaching strategies	Cruz et al. (2008), Guerra and Teixeira (2016), Miranda (2011), Santos (2012),

Source: research data.

At the institutional level, studies have pointed to some manageable variables that may influence student achievement, including program coordinator's characteristics (Rodrigues et al., 2017), incentives for students' engagement in research and extension (Ferreira, 2015), school infrastructure, and faculty characteristics and skills (Miranda, 2011). Rodrigues (2017) identified that 9.28% of the total variability in achievement on the Enade (2006 and 2009) was accounted for by the coordination of the undergraduate programs in Accounting, including: coordinator's highest title, incentives for students to attend academic events and do junior research, employment contract, research activities as a learning strategy, ratio of faculty members with a master's and a doctoral degree, type of academic organization.

Ferreira (2015) set out to identify the determinants of Accounting students' achievement on the 2012 Enade and found that 7.11% of the achievement variation could be

ascribed to the differences between the HEIs and faculty members. Significant variables at this level were: administrative category, regional location, number of seniors, percentage of faculty members with a master's degree, pedagogical teaching structure, and infrastructure.

Finally, this dissertation found that the following variables were significant to explain the Enade scores in 2006, 2009, 2012 and 2015:

1) student-related variables – gender, marital status, income, number of books read in the year, number of weekly study hours, and engagement in research and extension;
2) institution-related variables – administrative category, academic organization, infrastructure, and pedagogical teaching structure.

6.6 Dimensions of Quality in Accounting

The accreditation system of business schools, especially for the programs in Accounting, is well accepted by both the academia and the market as a hallmark in education. Certified programs are deemed to have quality as also shown through significant findings in academic research that has assessed the differences in student achievement by comparing certified schools and uncertified schools.

It seems that the US accreditation agencies take a value-for-money approach to quality (Harvey & Green, 1993). This runs counter to Dicker et al. (2018), according to whom the concept of quality changes from one stakeholder to another. For employers, quality is based on external networks with the market; for students, it is based on the engagement with faculty members; for teachers, it is based on the learning environment (Dicker et al., 2018). Concomitantly, quality is also seen as fit for purpose when the CPA passing score is used as a proxy of student achievement (Harvey & Green, 1993).

In general, the AACSB eligibility criteria for accrediting Accounting programs in the US seem to predominantly rest on the notion of quality as perfection or consistency (Harvey & Green, 1993), as it suffices that a given HEI complies with the core values, guiding principles and general criteria required by the Association. The conception of quality as consistency prioritizes processes and conformity with a set of specifications (Bertolin, 2009), precisely what is observed by the AACSB to certify the HEIs. Table 49 categorizes the quality dimensions in this model.

Table 49. Dimension of quality in U.S. undergraduate programs in Accounting

Object of Analysis	Components	Dimension (Harvey and Green, 1993)	Perspective (Bertolin, 2009)
Eligibility criteria for general accreditation at AACSB	Core values and guiding principles	Quality as Perfection or Consistency	Pluralistic Perspective
	General eligibility criteria		Economic perspective
Specific standards for accreditation in Accounting	Standard 1: HEI Mission, Impact and Innovation	Quality as Perfection or Consistency	Pluralistic Perspective
	Standard 2: Intellectual contribution to, impact on and alignment with the institution's mission		Pluralistic Perspective
	Standard 3: Funding and resource allocation strategies	Quality as a Value-for-Money	Economic perspective
	Standard 4: Accounting curriculum content, and learning assurance management	Quality as Perfection or Consistency	Economic perspective
	Standard 5: Students' and faculty members' skills in information technology, agility, and knowledge	Quality as Value-for-Money	Pluralistic Perspective
	Standard 6: Proficiency, Credentials, Qualifications, and Faculty Development		Pluralistic Perspective

Source: research data.

Eligibility criteria A and B (see Table 43) for accrediting Accounting schools make it clear that all stakeholders (faculty, students, staff, and organization leaders) are responsible for a collaborative environment, which emphasizes the idea of quality as perfection (Harvey & Green, 1993), with all stakeholders responsible for it. However, quality as value-for-money (Havey & Green, 1993) also stands out when Accounting-specific accrediting standards are aligned with the needs of the society and the market met by the courses and the Accounting profession, as measured through the quality indicators professional placement and career success of graduates (Barilla et al., 2008).

AACSB is devoted to aligning organizations, higher education institutions, faculty members and students so that business education is consistent with market demands and eventually contributes to global prosperity (AACSB, 2019). The AACSB role is to promote and develop higher education programs in the field of business (Administration and Accounting), and its mission is to promote commitment, accelerate innovation and broaden impact on education (AACSB, 2019). When the AACSB assigns itself a global commitment, one which is focused on marketing dimensions, it also sees quality in education as value-for-money (Harvey & Green, 1993).

In Brazil, the criterion for assessing the quality of undergraduate programs is defined by the National Assessment System for Higher Education, which is disseminated

through its quality indicators – in the present case, the CPC (Preliminary Program Quality Level). Table 50 categorizes the quality dimensions in the Brazilian model.

Table 50. Dimensions of quality observed in indicator Preliminary Program Quality Level (CPC)

Object of Analysis	Components	Dimension (Harvey and Green, 1993)	Perspective (Bertolin, 2009)
Preliminary Program Quality Level	Value added by the training process provided by the undergraduate program	Quality as Transformation	Economic perspective
	Student achievement Characteristics of faculty	Quality as Value-for-Money	
	Student perception of infrastructure Student perception of the pedagogical teaching structure	Quality as Fit for Purpose	
	Student perception of opportunity for further training/learning		Pluralistic Perspective

Source: research data.

Quality as fit for purpose holds sway when the customer defines the purpose and quality is expected to meet the requirements for customer satisfaction (Harvey & Green, 1993). One of the assessment criteria of undergraduate programs in Accounting in Brazil is the student's perception of the program's infrastructure, pedagogical teaching structure, and opportunity for further training/learning. Students answer a questionnaire addressing these three variables based on a Likert scale ranging from totally agree to totally disagree. The students' rates for a given program are averaged, and the program is assigned a rate from 1 to 5 points (Brasil, 2018b). This is a functional use of quality (Bertolin, 2009), as quality is awarded to those programs rated 3 or higher in these requirements.

Similarly, the assessment of student achievement through the Enade measures students' achievement against the syllabus, skills and competences expected to be provided by the program. In this case, the students do not determine what is the quality of their achievement, but they themselves have this quality measured through a standardized test. The assessment of the faculty member's employment contract and ratio of those with a master's and doctoral degrees in a program follows the same logic. According to Sinaes, a program with a CPC of 3 or above has quality in this requirement (Brasil, 2018b).

From the point of view of professional certification, both countries seem to have a view of quality as fit for purpose, because the "product" (the training provided) will be useless if it does not meet the need for which it was created (i.e., to exercise the Accounting profession) (Bertolin, 2009). For both the AICPA and the CFC, applicants must achieve a minimum score to prove quality in the proficiency exam.

Harvey and William (2010) analyzed 15 publications on quality in higher education and concluded that quality is a multifaceted notion with serious political tendencies and should, therefore, be interpreted based on its purpose and its context. The notion of quality is often the result of tensions between a focus on processes and a focus on outcomes, between internal and external stakeholders, between quantitative and qualitative measures, between proposals for diversity or uniformity in the institutions.

In Brazil, where the higher education system is organized, controlled, and regulated by the federal government, it is not surprising that quality as fit for purpose is the dominating notion in the system. In the US, where education is deregulated, organizations are autonomous, and a competitive market is always looking for novelties, the idea of quality as perfection seems to be more suitable. This state of affairs is also reflected in the different views (Bertolin, 2009) about quality in higher education.

In both Brazil and the US, the purpose of training in Accounting is strongly based on potential aspects of economic growth and employability, which Bertolin (2009) called the economic perspective. In particular in the US, professional boards have sought to provide the market with high performance professionals that contribute to growth in both public and private sectors. For this reason, they have developed quality criteria for the higher education programs and a rigorous certification system for new professionals.

Similarly, the certification process of professionals in both countries (i.e., accreditation of schools in the US, and ranking of programs in Brazil) shows how differentiation and relevance are context specific. This is when the education sector is concerned not only with the market, but also with providing a structure that meets local specificities and sociocultural aspects. Bertolin (2009) describes this as the pluralistic view of quality in higher education.

Finally, there is an important mismatch between the notion of quality at the institution's global level and at the departmental level (Krause, 2012). Findings about quality at the organizational level show that quality is related to the dimensions proposed by Harvey and Green (1993); however, findings about quality at the microlevel of the HEIs show that the stakeholders are more concerned with the underlying purpose of ensuring quality than making sure that the system works (Krause, 2012). In other words, several stakeholders are worried about their jobs and not worried about student achievement ultimately. This may be the reason why the dimension of quality as transformative is incipient in this research.

In none of the contexts, the system is concerned with the social inequalities produced by the economic system. There are no eligibility criteria (US), certification criteria

(Brazil and US) or Program Quality Level criteria (Brazil) that attend to aspects of social cohesion.

7 FINAL REMARKS

Brazil has recorded increasing numbers in higher education, including number of enrollments, number of institutions, and public investment in several undergraduate programs (Inep, 2016). Debates have mostly revolved around not only the increased access to this level of education, but also the quality of education and the impact of such investment on economic growth (Bruns & Luque, 2015).

The notion that good undergraduate studies enables young people to meet the challenges of the globalized world and that the quality of higher education in a given country contributes to its international competitiveness has led several countries to prioritize educational investments in higher education (Liu, 2011a). Given the increased public investments, a myriad of measures has been developed building on theoretical models to assess the training process and provide transparency and justification to the expenses.

In Brazil, Act No. 10,861, as of 14 April 2004, established the National Assessment System for Higher Education (Sinaes) to assess all higher education institutions (HEIs), all undergraduate programs and the academic achievement of all students in the country. The ultimate purpose is to improve the quality of higher education in Brazil.

The National Institute for Educational Studies and Research Anísio Teixeira (Inep), responsible for implementing the Sinaes, has published quality indicators of HEIs, programs and students. One of such indicators, the Indicator of Difference between Observed and Expected Achievements (IDD), was chosen as the object of interest in this dissertation. As mentioned in chapters 1 and 2, it is "a quality indicator that measures the value that an undergraduate program adds to the development of its seniors by probing their achievement on the Enade as compared to their developmental characteristics at the beginning of their study track" (Inep, 2017b, p. 1).

The Sinaes has applied four different IDD estimation methodologies since 2004 to obtain the value added of the undergraduate programs. This dissertation set out to investigate the impacts of changing the methodology for estimating the IDD for undergraduate programs in Accounting. This study was carried out in Brazil, one of the few countries to employ an assessment system of student achievement at the end of high school (Enem) and at the end of undergraduate studies (Enade), allowing for probing into how much an undergraduate program contributes to its students' academic achievement through a value-added measure.

The present study used longitudinal data of 30,668 students from 911 undergraduate programs in Accounting who both took the Enade in 2015 and scored above

zero on the Enem in previous years. Accounting is the fourth largest field of graduate studies in the country, with the enrollment of 359,840 students in 1,489 higher education institutions (INEP, 2019a).

When students perform as expected, not only do individuals benefit from investment in education, but society is rewarded through higher tax revenues and social contributions, including reduced public spending on social assistance programs (OECD, 2016). However, several Accounting students have had poor results.

In the 2019/2 CFC Proficiency Exam, only 30% of applicants passed; in the 2018 Enade, 398 (36%) of the 1,101 undergraduate programs in Accounting scored below 3, i.e., they did not reach a satisfactory quality level according to Inep criteria. These results show the need to improve student achievement in Accounting in higher education in Brazil. Policy changes should address this problem aiming at benefits for the entire population.

The crux of this dissertation lied in three primary questions: (1) What is educational quality, specially as it pertains to higher education institutions?; (2) How do we measure quality?; and (3) How can policymakers promote high quality among educational institutions? These are questions that have dominated much discussion among economists in education. This work is grounded in a human capital perspective (rather than the signalist) where the role of educational institutions go beyond sorting individuals.

To that end, the first objective of the dissertation was to analyze the differences between the estimation methodologies for the value-added indicator IDD. The key to understanding the notion of value added is as follows: predicting a senior student's achievement score is assumed to be possible by using data about this student's characteristics before admission to the undergraduate studies, including his/her prior achievement scores and sociodemographic characteristics. Three outcomes are possible through this method: 1) the student may perform as expected (in this case, the result is zero); 2) the student may perform better than expected (a positive result); or 3) the student may perform poorer than expected (a negative result). Within a given institution, all student results are summed, and the final sum is what is called value added, i.e., how much a given program contributed to student achievement.

The Sinaes has used four different methodologies for IDD estimation over the years since its launch in 2004. All of them were based on the analysis of regression (Kim & Lalancette, 2013). The first three models used simple regression statistics, estimated by the ordinary least squares method. The current model is a more complex equation based on a

maximum likelihood-adjusted, linear mixed effects model. It was not only the statistical analysis that was different but also some of the underlying assumptions and modeling.

Statistically, Models 1-3 do not present significant differences for estimating the programs' value added, as their methodologies are only slightly different from one another. In the first model, used from 2006 to 2010, value added was estimated by the residual of a regression equation that had the seniors' Enade scores in a given program as the response variable for the following explanatory variables: freshmen's Enade scores in that year, percentage of parents with higher education, ratio of freshmen and seniors, percentage of faculty members with a master's, percentage of faculty members with a doctoral degree, and employment contract, in addition to program-related variables infrastructure and pedagogical teaching structure. In Model 2, in force in 2011 and 2012, the freshmen's Enade scores were replaced by the students Enem scores. In Model 3, in place in 2013, the program-related variables were changed to include pedagogical teaching structure, infrastructure, and opportunity for further training/learning.

The first limitation of this research follows from this: as the Enade is no longer taken by the freshmen, Model 1 cannot be reproduced exactly as defined in Inep Technical Bulletin (Inep, 2009; 2009a). The freshmen's scores were replaced with those of the seniors who had an available Enem score. Similarly, not all students' Enem scores could be used to adjust Models 2 and 3, because Inep microdata do not link students and the respective program in which they were enrolled. As such, also in Models 2-3, the Enem scores used were those of the students with Enade scores available. Therefore, the first limitation of this research refers to the adjustment performed in the data to reproduce the respective Inep Technical Bulletins.

In Models 1-2, the explanatory variables were significant for academic achievement, but in Model 3, with the restructuring of the measures of the conditions of training and learning, the following variables resulted non-significant: percentages of parents with higher education, percentage of faculty with a master's or a doctoral degree, pedagogical teaching structure, and opportunity for further training/learning. Pearson correlation coefficient between the IDDs of Models 1-3 was very close to 1 (0.98), showing almost perfect fit.

It is also important to note that models 1, 2 and 3 were not really representative of value added modeling as it is contemporarily understood because the controls for previus achievement were really about the academic competitiveness of the incoming Freshman class. That is, models do not really reflect a VAM design if the scores are not from panel data where

the students are the same. A better description is that the model controlled for "quality of the program" as measured by the achievement scores of the incoming Freshman class.

Model 4 presented a very distinct structure and produced statistically different results, as it is a hierarchical model, with data analyzed at two levels: students within programs and students across programs. It also excluded explanatory variables at the institution- and student-levels, so that academic achievement (Enade) was a function of prior knowledge (Enem). Pearson correlation coefficient between Model 4 and 3 was 0.78.

With a view to understanding the outcome for Program Quality Level ranking using different methodologies for IDD estimation, a model-by-model comparative analysis was performed to determine if the programs would experience changes in their CPC Levels. As expected, few programs had their CPC Levels changed across Model 1-3, but more significant changes occurred with the new methodology (Model 4).

Model 4, where the IDD is estimated from the analysis of residuals based on hierarchical linear models, presents more reliable and consistent results about school effect when compared to Model 3, where the value added is estimated by the OLS difference of residuals. This is because the model assumes that there is a difference not only between students but also between programs. In Model 4, there was a reduction in the central quality measures of the programs' categorical CPC (Level 3) and an increase in the extreme CPC Levels of higher quality (Levels 4 and 5) and lower quality (Levels 1 and 2). In other words, Model 4 has greater discrimination power.

Discussing of how each method has an impact on how policymakers characterize the effectiveness of programs is an important contribution this dissertation provides. Variations in ranking and measures of effectiveness have fiscal implications for the programs themselves as well as the national government since funding is tied to assessment. It was interesting to note that the worst and the best programs (Table 23, page 77) are consistent across the models. It is also impressive that Model 4 results in more institutions considered to have achieved acceptable standards. For future research, investigating implications for the budget on these different rankings is suggested. On the one hand, by increasing the number of programs considered to be performing effectively, this should increase budgetary costs. However, if increasing the number of effective programs resulted in decreased auditing and inspection costs, this may have a downward pressure on costs.

An important operational concern in value-added models is identifying what quality differences among students to consider in order not to underestimate or overestimate the value added. Shavelson et al. (2016) argue that students from better socioeconomic

backgrounds can travel, speak other languages, and more easily find internship opportunities due to their parents' networking. Ignoring such characteristics could disadvantage institutions that admit students from low socioeconomic backgrounds when compared to those who admit more privileged students.

Therefore, one of the specific objectives of this research was to assess which are the determinants of academic achievement among Accounting students to identify which variables should be considered to estimate a program's value added. Drawing on an analysis of data from years 2006, 2009, 2012 and 2015, some variables were found significant for achievement on the Enade every year: six variables related to student's personal characteristics (gender, marital status, income, number of books read in a year, weekly study hours, and engagement in research activity, extension and teacher assistance); and five variables related to institution's characteristics (administrative category, academic organization, regional location, infrastructure and pedagogical teaching structure).

In ignoring the significant characteristics of the determinant of achievement, the current Inep model assumes that students are equals and that there is no competition for admission to different programs. In other words, considering that student achievement will be the same across the schools means that any student could be admitted in any school. But this does not hold true in practice. Therefore, this dissertation proposes a new model for estimating value added for undergraduate programs in Account Science in Brazil: Model IDD-VDCF.

It admits that there are differences at the moment of admission, but also in the conditions of training and learning; therefore, it is necessary to level the students based on their personal characteristics in order to isolate this effect from the school contribution. Although Model IDD-VDCF is statistically close to the current Inep model, it seeks to reduce discrepancies between programs by controlling for sociodemographic and institutional variables. The IDD-VDCF distinguishes itself for its lack of significant associations with any of the individual characteristics in contrast to the current and previous IDD measures. The residual related to each program, which allows for ranking the institutions, may carry the weight of unobservable variables, but admittedly capable of influencing academic achievement, such as student-teacher relationship, dedication of program coordinator, learning environment, use of different methodologies in the classroom. These variables are partially captured by the variables measured in the Sinaes: infrastructure, pedagogical teaching structure, and opportunity for further training/learning. Sinaes

Defining the quality of undergraduate programs is not a simple topic that has been explored to exhaustion in Brazil. Due to limited data, this dissertation did oversee some other important variables, which could be the object of further research. One is school dropout. Data from the 2018 Statistical Synopsis of Higher Education show that 10.72% of all students suspended registration, 18.17% had their registration cancelled, and 1% was transferred to another program (INEP, 2019c). Most of these numbers are concentrated in private institutions: 85.68% for suspended registration, 85.17% for cancelled registration, and 76.50% for transferences.

Prestes and Fialho (2018) estimated the cost of dropout at Universidade Federal da Paraíba from 2007 to 2012. They reached an estimated value of BRL 415 million, related only to the institution's financial losses, i.e., neglecting the losses for both families and society. Extrapolating these data to all higher education institutions in Brazil, both private and public, certainly hits an exorbitant value. Thus, discussing quality of undergraduate programs requires data about dropouts, which are harmful to the entire society.

In addition, this research did not collect employability information about the graduates from Accounting programs. Measuring their employability is crucial for assessing the quality of programs. In theory, the more a worker contributes to adding value to a product, the higher his/her salary tends to be. In other words, higher salaries signal higher productivity, which in turn is increased through knowledge and skills that can be acquired in undergraduate studies (Cunha & Miller, 2014).

Besides, it is important to collect data about what graduates do: where they work, their positions, their careers; whether in social organizations, private companies, or public institutions; on what specifically they are working; what is the impact of their work to the community/society. These are aspects that need to be discussed when addressing quality in higher education. Brazil already has a large data system, but in fact a small part of it has been used to improve the education system.The CPC was a quality indicator created by the Sinaes to solve a practical problem in assessing undergraduate programs in Brazil. Act No. 10.861, as of 2004, which regulates the Sinaes, required that all programs should be assessed by a panel of experts at each assessment cycle. However, it is impractical and unfeasible for all programs in Brazil to be visited on site every three years. This would require approximately 100 weekly visits by MEC raters (Verhine, 2015). Therefore, the CPC was created to limit the mandatory on-site visits to those programs showing unsatisfactory quality, that is, those with a CPC lower than 3. Even though the CPC tries to summarize complex information about the quality of programs, it has gained relevance because it is easy for the general public to

understand (Verhine, 2015) and it has even been used for resource allocation in the national system.

In other words, this indicator has taken on greater importance than it should in its original conception and design. The CPC is a measure of complex understanding because it unites three dimensions of Program Quality Level: 1) those related to academic achievement (Enade and IDD rate), 2) those related to faculty characteristics (doctoral degree, master's degree, and employment contract), and 3) those related to the conditions of training and learning (infrastructure, pedagogical teaching structure, and opportunity for further training/learning). The weights of these components for estimating the CPC were defined in Technical Bulletin No. 029, as of 15 October 2012, which also cites Fernandes et al. (2009). In personal communication with one of the authors, she said they are restructuring their model and several changes have been made since its publication.

MEC Administrative Rule No. 840/2019 defines Inep's competences for assessing higher education institutions in Brazil; its Art. 61 says that it is responsible for defining, estimating and disseminating higher education quality indicators that provide supporting evidence for public policies. The CPC Level and the Enade score are two of the major quality indicators that have gained the spotlight through social media and publicity by higher education institutions. In October 2019, Inep released the data for 2018, which reverberated in the Brazilian media, especially because of the poor results. News reports provided comparisons between rates, between learning modalities, between administrative categories and between regional locations to show what are the "quality" programs in the country and where they are located.

However, little is known about the IDD, which has been barely used by organization leaders and passed unnoticed to the general public. This is precisely the indicator that measures how many the institutions actually contributed to their students' academic achievement. The IDD is only one element of the CPC but contributes 35% of the final score of a program. Based on the findings of this research, it is still underestimated what the goal is to compare institutions.

It is also underestimated as an indicator of what institutional and academic variables should obtain more investment if the students are to perform better at the end of their graduate studies. The research shows that despite the IDD estimation methodology, the programs' value added has been significantly related to the characteristics of their faculty (percentages of those with a master's degree and a doctoral degree, and employment contract), as well as their conditions of training and learning (infrastructure, pedagogical

teaching structure, and opportunity for further training/learning). If an academic unit providing undergraduate program in Accounting aims to add value to its students' academic achievement, it should observe these characteristics, which can be improved through internal policies.

Studies on educational output and determinants of academic achievement are relevant because they show how difficult it is to apply some economic models empirically, and their results reveal a range of problems that is much wider than simply understanding the institution management. The results of these studies go beyond school issues, including, for example, wage, status, school funding, impacts of educational quality (Hanushek, 1979).

Governments worldwide, and especially countries with low educational quality, have looked for studies that compare educational opportunities and educational outcomes to develop policies that improve the economic and social perspectives of their population, i.e., to allocate resources productively in such a way that meet society's demands (OECD, 2016).

Despite the importance of standardized achievement tests, such as the Enem and the Enade, important differences that exist throughout Brazil should be reckoned to avoid unreasonable arguments. For instance, one cannot compare on the same basis institutions located in the countryside of Northern Brazil with institutions in large urban centers in the Southeast.

Value added is an important measure because it considers the student's conditions of admission to a given program before estimating what is expected of this student at the end of the studies. In the last year of undergraduate studies, another test measures if a given institution added more or less value to its student's achievement, as expected. As Model IDD-VDCF controls for student-specific and institution-specific variables, it allows for comparing institutions to each other and for their leaders to follow up how much they are actually contributing to academic achievement over time.

In addition, such study also reveals what factors have contributed for some institutions to have more value added than others. After all, what is the purpose of a quality assessment system that does not help improve teaching? A quality assessment system should lend itself to improving what is being assessed.

Nevertheless, the Enem is a test in which the student's greatest motivation lies in earning a place in a university. S/he knows that the higher his/her Enem score is, the greater the chance of being admitted and, better yet, of choosing the HEI. In contrast, the Enade is a mandatory test, but there is no (positive) motivation for the student to take it with the desire to obtain the best score. Scoring 0 or 100 does not affect his/her career. This seems to be one of

the greatest limitations of value-added research in Brazil. The Enade score could at least be on the students' transcript of records if the government wants an adequate measure of their final achievement.

In practice, this dissertation advances in the discussion about the quality indicators released by the Brazilian government. Model IDD-VDCF, as proposed here, allows for comparing programs to each other more fairly than the other models, as it starts from the assumption that students and institutions are different from one another, and this difference can interfere with students' academic achievement regardless of the program management model and the institution's educational policies. In considering these differences for estimating a program's value added, Model IDD-VDCF makes indicator IDD a more accurate measure of the program's contribution to student achievement.

The very nature of competitive admissions among students and the ranking of programs underlie the fact that there is sorting. The role of policymakers in that context, then, is to ensure that they develop policies that allow that sorting to be equitable, where individuals and programs are not merely rewarded for having characteristics privileged by society. Therefore, the major contribution of this dissertation is the development of a measure that disentangles more fully what the contribution of programs are to student learning, and what merely is a reflection of the capacity that a student brought to the program.

This dissertation also contributes to educational leaders of undergraduate programs in Accounting as it provides the variables that are significantly related to a program's value added. The programs that added the most value were those whose students had better reading and study habits, those whose faculty members had better academic qualifications, and those which provided better infrastructure, pedagogical teaching structure, and opportunities for research, teaching, and extension.

Human Capital Theorists not only argue normatively that the purpose of schooling is to enhance the skills of students who attend these organizations, but they also believe that schools actually do enhance student capacity. One role of scholars, then, is to discover the right measure of the knowledge added by educational institutions and the role of policymakers is to make proper use of that data in their policy. This study has shown that investment in faculty should be prioritized in terms of its implications for student achievement.

The CPC, as the main indicator of Sinaes, cannot demonstrate the quality of the programs to the public. In Levels from 1 to 5, this indicator agglutinates valuable information that should be clearly informed to the stakeholders involved in higher education. The IDD

aligned with the Enade score is undoubtedly the main and most suitable mechanisms for the purpose of promoting quality across the undergraduate programs in Brazil. The scores for infrastructure, pedagogical teaching structure and opportunity for further training/learning should not be disclosed as quality indicators, but as indicators of the condition of training and learning. After all, as this research shows, the program's contribution (IDD) is directly related to these variables.

This research, nonetheless, addressed specifically the undergraduate programs in Accounting. It is also necessary to question the assessment system in Brazil, where standardized knowledge tests are the basis of everything. In the Brazilian law, the role of higher education is not only to provide content that are necessary for a given profession or career, but most importantly to educate citizens capable of contributing to the cultural, political, social and economic development of the country. In the US, the quality of Accounting programs is also linked to an economic view. Eligibility criteria for business school accreditation require sustainability and commitment to the community, but above all, professional placement and career success for their graduates. Credentials, qualifications, efficient resource allocation strategies, and innovation are some of the most used words by the AACSB. Perhaps this characteristic is typical of the Accounting profession. The complex global business environment requires professionals with high critical and analytical skills, which justifies the efforts of the Accounting community to provide the market with the best professionals (Ugrin & Honn, 2018).

REFERENCES

Abdullah, A. M. (2011). Factors affecting business students' performance in arab open university: the case of Kuwait. *International Journal of Business and Management*, 6(5), pp.146-155. https://doi.org/10.5539/ijbm.v6n5p146

Alexander, N. A., Jang, S. T., & Kankane, S. (2017). The performance cycle: The association between student achievement and state policies tying together teacher performance, student achievement, and accountability. *American Journal of Education*, 123(3), 413-446. https://doi.org/10.1086/691229

Alexander, N. A. & Jang, S. T. (2019). Expenditures on the professional development of teachers: The case of Minnesota. *Journal of Education Finance*, 44(4), 385-404.

Araújo, A. M. P. de, Santana, A. L. A., & Ribeiro, E. M. S. (2009). Fatores que afetam o processo ensino no curso de ciências contábeis: um estudo baseado na percepção dos professores. *Anais do Congresso ANPCONT*, São Paulo, SP, Brasil, 3.

Association to Advance Collegiate Schools of Business. (2018). *2018 Eligibility Procedures and Accreditation Standards for Accounting Accreditation: Engagement, Innovation, Impact*. Flórida: AACSB. Retrieved from https://www.aacsb.edu/-/media/aacsb/docs/accreditation/accounting/standards-and-tables/2018-accounting-standards.ashx?la=en&hash=8DCDA6CE3B0CEF6AB82D39CBF53995DA96111196.

Association to Advance Collegiate Schools of Business. (2019). *Accreditation and Accredited Schools*. 2019. Retrieved from https://www.aacsb.edu/accreditation/accredited-schools.

Association to Advance Collegiate Schools of Business. (2019a). *History of AACSB*. Retrieved from https://www.aacsb.edu/about/who-we-are/timeline.

Association to Advance Collegiate Schools of Business. (2019b). *What we do*. Retrieved from https://www.aacsb.edu/about/mission.

Association of International Certified Public Accountant. (2017). *2017 Trends: in the supply of accounting graduates and the demand for public accounting recruits*. Durham: Arc Research, 2017. 38 p. Available at <https://www.aicpa.org/interestareas/accountingeducation/newsandpublications/downloadabledocuments/2017-trends-report.pdf>.

Association of International Certified Public Accountant. (2019). *2019 Trends: in the supply of accounting graduates and the demand for public accounting recruits.* Durham: Arc Research, 2019. 44 p. Available at https://www.aicpa.org/content/dam/aicpa/interestareas/accountingeducation/newsandpublications/downloadabledocuments/2019-trends-report.pdf

Association of International Certified Public Accountant. (2019a). *AICPA Mission and History*. Retrieved from https://www.aicpa.org/about/missionandhistory.html.

Austin, A. W. (1980). When does a college deserve to be called high quality? *Current issues in higher education*: improving teaching and institutional quality, 2(1), 1-9.

Bailey, T., & Xu, Di. (2012). Input-Adjusted Graduation rates and College Accountability: what is known from twenty years of research? *Context for Success*, 38p. Available at http://www.hcmstrategists.com/contextforsuccess/papers/LIT_REVIEW.pdf.

Barilla, A. G., Jackson, R. E., & Mooney, J. L. (2008). The CPA Exam as a Postcurriculum Accreditation Assessment. *Journal of Education for Business*, 83(5), 270-274. https://doi.org/10.3200/JOEB.83.5.270-274

Barnett, R. (1992). *Improving higher education: total quality care*. Bristol, USA: SRHE and Open University Press.

Beck, F., & Rausch, R.B. (2014). Fatores que influenciam o processo ensino-aprendizagem na percepção de discentes do curso de Ciências Contábeis. *Contabilidade Vista & Revista*, 25(2), 38-58.

Bertolin, J. C. G. (2009). Qualidade em Educação Superior: da diversidade de concepções a inexorável subjetividade conceitual. *Avaliação*, 14(1), 127-149. https://doi.org/10.1590/S1414-40772009000100007

Bline, D. M., Perreault, S., & Zheng, X. (2016a). An exploratory examination of order effects on CPA exam passage timeliness. *Journal of Accounting Education*, 36, 65-74. https://doi.org/10.1016/j.jaccedu.2016.05.003

Bline, D. M., Perreault, S., & Zheng, X. (2016b). Do accounting faculty characteristics impact CPA performance? An investigation of nearly 700,000 examinations. *Issues in Accounting Education*, 31(3), 291-300. https://doi.org/10.2308/iace-51227

Bogoya, J. D., & Bogoya, J. M. (2013). An academic value-added mathematical model for higher education in Colombia. *Ingeniería E Investigación*, 33(2), 76-81.

Boone, J., Legoria, J., Seifert, D. L., & Stammerjohan, W. W. (2006). The associations among accounting program attributes, 150-hour status, and CPA exam pass rates. *Journal of Accounting Education*, 24, 202-215 https://doi.org/10.1016/j.jaccedu.2006.08.002

Brasil. *Constituição da República Federativa do Brasil de 1988*. Recuperado de http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Brasil. *Decreto nº 7.233, de 19 de julho de 2010*. Dispõe Sobre Procedimentos Orçamentários e Financeiros Relacionados à Autonomia Universitária, e Dá Outras Providências. Retrieved from http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7233.htm

Brasil. *Lei n. 9.394, de 20 de dezembro de 1996*. Estabelece as diretrizes e bases da educação nacional. Retrieved from http://www.planalto.gov.br/ccivil_03/leis/l9394.htm.

Brasil. *Lei nº 10.861, de 14 de abril de 2004*. Institui O Sistema Nacional de Avaliação da Educação Superior – Sinaes e Dá Outras Providências. Retrieved from http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.861.htm.

Brasil. *Lei n. 12.249, de 11 de junho de 2010.* Institui o Regime Especial de Incentivos para o Desenvolvimento [], altera os Decretos-Leis 9.295, []. Retrieved from http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Lei/L12249.htm.

Brasil. *Lei nº 13.530, de 07 de dezembro de 2017*. Altera a lei 10.260, [], a lei 9.394, [], e dá outras providências. Retrieved from http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Lei/L13530.htm.

Brasil. Ministério da Educação. (2016). *Principais Ações e Programas de Responsabilidade do Ministério da Educação no PPA 2012-2015*. Retrieved from http://portal.mec.gov.br/sesu-secretaria-de-educacao-superior/programas-e-acoes.

Brasil. Ministério do Trabalho. (2019). *Programa de Disseminação das Estatísticas do Trabalho: Microdados CAGED (*Relatório Técnico), Brasília, DF, Brasil. Retrieved from http://pdet.mte.gov.br/microdados-rais-e-caged

Brasil. *Portaria nº 209, de 07 de março de 2018. Dispõe Sobre O Fundo de Financiamento Estudantil - Fies, A Partir do Primeiro Semestre de 2018.* Retrieved from http://fies.mec.gov.br/arquivos/portaria_normativa_209_07032018.pdf.

Brasil. *Portaria nº 840, de 24 de agosto de 2018.* Dispõe Sobre Os Procedimentos de Competência do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - Inep Referentes à Avaliação de Instituições de Educação Superior, de Cursos de Graduação e de Desempenho Acadêmico de Estudantes. Retrieved from http://download.inep.gov.br/educacao_superior/avaliacao_institucional/legislacao_normas/2018/portaria_normativa_GM-MEC_n840_de_24082018.pdf

Brasil. *Portaria nº 952, de 2 de maior de 2019*. Dispõe sobre o processo seletivo do Fundo de Financiamento Estudantil - Fies e do Programa de Financiamento Estudantil - P-Fies referente ao segundo semestre de 2019. Retrieved from http://www.in.gov.br/web/dou/-/portaria-n%C2%BA-952-de-2-de-maio-de-2019-86527897.

Brasil. Secretaria do Tesouro Nacional. (2016). *Gasto Social do Governo Central: 2002 a 2015* (Relatório Técnico), Brasília, DF, Brasil. Retrieved from <http://www.tesouro.fazenda.gov.br/documents/10180/318974/Gasto+Social+Governo+Central/c4c3d5b6-8791-46fb-b5e9-57a016db24ec>

Bruns, B., & Luque, J. (2015). Great teachers: how to raise student learning in Latin America and the Caribbean. *Latin American Development Forum*. World Bank Group. https://doi.org/10.1596/978-1-4648-0151-8

Camargo, W. R V., Camargo, R. C. C. P., Andrade, D. F., & Bornia, A. C. (2016). Desempenho dos alunos de Ciências Contábeis na prova Enade/2012: uma aplicação da Teoria da Reposta ao Item. *Revista de Educação e Pesquisa em Contabilidade*, 10(3), 332-355. https://doi.org/10.17524/repec.v10i3.1401

Canziani, Alex et al. (2018). *Financiamento da educação superior no Brasil: impasses e perspectivas* (Estudos Estratégicos Working paper n.11). Brasília: Câmara dos Deputados.

Center for Education Policy Research. (2011). *Value-added measures: How and Why Strategic Data Project Uses them to Study Teacher Effectiveness*. Cambridge: Harvard University. 8 p. Retrieved from https://hwpi.harvard.edu/files/sdp/files/sdp-va-memo_0.pdf.

Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9), 2593-2632. https://doi.org/10.1257/aer.104.9.2593

Clark, V. L. P., & Creswell, J. W. (2015). *Understanding Research: A Consumer's Guide* (2a ed). New Jersey: Pearson.

Clarke, S. P., & Cossette, S. (2000). Secondary analysis: Theoretical, methodological and practical considerations. *Canadian Journal of Nursing Research*, 32(3), 109-129.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26 (6), 673-682. https://doi.org/10.1016/j.econedurev.2007.10.002

Coates, H. (2009). What's the difference? A model for measuring the value added by higher education in Australia. *Higher Education Management and Policy*, 21(1), 1-20. https://doi.org/10.1787/hemp-v21-art5-en.

Conselho Federal de Contabilidade. (2018). *Statistical Reports of the sufficiency of accounting examination.* Available at https://cfc.org.br/registro/exame-de-suficiencia/relatorios-estatisticos-do-exame-de-suficiencia/.

Conselho Federal de Contabilidade. (2019). *Exame de Suficiência de 2019*. Retrieved from <https://cfc.org.br/exame-de-suficiencia-anteriores/2o-exame-de-suficiencia-de-2019/>.

Conselho Federal de Contabilidade. (2019a). *Profissionais Ativos nos Conselhos Regionais de Contabilidade.* Retrieved from <http://www3.cfc.org.br/spw/crcs/ConselhoRegionalAtivo.aspx>.

Cornachione Junior, E. B., Cunha, J. V. A., Luca, M. M. M., & Ott, E. (2010). O bom é meu, o ruim é seu: perspectivas da teoria da atribuição sobre o desempenho acadêmico de alunos da graduação em Ciências Contábeis. *Contabilidade & Finanças*, 21(53), 1-24. https://doi.org/10.1590/S1519-70772010000200004

Couto, A. P. (1989). *A influência norte-americana no sistema contábil brasileiro: elementos para se repensar a história da contabilidade no Brasil*. (Dissertação de Mestrado). Instituto Superior de Estudos Contábeis, Fundação Getúlio Vargas, Rio de Janeiro, RJ, Brasil.

Cunha, J. M., & Miller, T. (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review*, 42, 64-77. https://doi.org/10.1016/j.econedurev.2014.06.001

Darwin, S. (2016). *Student Evaluation in Higher Education: Reconceptualizing the Student Voice*. Switzerland: Springer. https://doi.org/10.1007/978-3-319-41893-3

Deines, D. S., Bittner, J., & Eichman, G. (2012). The Accounting Pilot and the Bridget Project. *Issues in Accounting Association*, 27(1), 113-122. https://doi.org/10.2308/iace-50111

Dicker, R., Garcia, M., Kelly, A., & Mulrooney, H. (2019). What does 'quality' in higher education mean? Perceptions of staff, students and employers. *Studies in Higher Education*, 44(8), 1425-1441. DOI:10.1080/03075079.2018.1445987

Draper, N. R., & Smith, H. A. (1998). *Applied Regression Analysis* (3a ed). New York: John Wiley. https://doi.org/10.1002/9781118625590

Elton, L. (1986). Quality in Higher Education: nature and purpose. *Studies in Higher Education*, 11(1), 83-84. https://doi.org/10.1080/03075078612331378481

Fernandes, A., & Tomazelli, I. (2019, Ago). MEC estuda mudar distribuição de recursos para universidades federais. *O Estado de São Paulo*. São Paulo, p.1-2. Retrieved from <educacao.estadao.com.br/noticias/geral,mec-estuda-mudar-distribuicao-de-recursos-para-universidades-federais,70002973204>.

Fernandes, R., Pazello, E. T., Leitão, T. M. S. P., & Moriconi, G. M. (2009). *Avaliação de Cursos na Educação Superior: a função e a mecânica do Conceito Preliminar de Curso*. Série Documental - Textos para Discussão, Brasília: Inep.

Fernandes, V. D. C.; Miranda, G. J. (2018). Habilidade de leitura e desempenho acadêmico: um levantamento com alunos do curso de Ciências Contábeis. *Anais Encontro Nacional de Estudos e Pesquisas em Administração e Ciências Contábeis*, Porto Alegre, RS, Brasil, 6.

Ferreira, M. A. (2015). *Determinantes do Desempenho Discente no ENADE em cursos de Ciências Contábeis* (Dissertação de Mestrado). Curso de Ciências Contábeis, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil.

Freitas, S. C. (2012). *An Exploratory Study on ENADE Evaluation Report Utilization and its Impact on Undergraduate Accounting Program Performance in Brazil* (Tese de Doutorado). Programa de Pós-graduação em Ciências Contábeis, Universidade de São Paulo, São Paulo, SP, Brasil.

Freitas, S. C., Barbosa, I., Vieira, J. A. G., & Miranda, G. J. (2015). Percepção acerca da qualidade e utilidade do relatório de avaliação do ENADE: um estudo na área de negócios. *Revista Contemporânea de Contabilidade*, 12(27), 117-136. https://doi.org/10.5007/2175-8069.2015v12n27p117

Galiffa, J., & Batalle, P. (2010). Student perceptions of service quality in a multi-campus higher education system in Spain. *Quality Assurance in Education*, 18(2), 156-170. https://doi.org/10.1108/09684881011035367

Gil, A. C. (2008). *Métodos e Técnicas de Pesquisa Social* (6 ed). Atlas: São Paulo.

Grant, T., Ciccotello, C., & Dickie, M. (2002). Barries to professional entry: How effective is 150-hour rule? *Journal of Accounting and Public Policy*, 21, 71-93. https://doi.org/10.1016/S0278-4254(02)00037-6

Green, P. C., Baker, B. D., & Oluwole, J. (2012). The Legal and Policy Implications of Value-Added Teacher Assessment Policies. *Brigham Young University Education and Law Journal*, 2(1), 1-29.

Griboski, C. M. (2012). O Enade como Indutor da Qualidade da Educação Superior. *Estudos em Avaliação Educacional*, 23(53), 178-195. https://doi.org/10.18222/eae235320121920

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7 Ed). New Jersey: Prentice Hall.

Hanushek, E. A. (1979). Conceptual and Empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, 14(4), 351-388. https://doi.org/10.2307/145575

Hanushek, E. A. (2002). Teacher Quality. In: Izumi, L. T., & Evers, W. M. *Teacher Quality*. Stanford: Hoover Press, 2002.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *The American Economic Review*, 100(2), 267-271. https://doi.org/10.1257/aer.100.2.267

Hanushek, E. A., & Woessmann L. (2012). Do better schools lead to more growth?: Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17, 267-321. https://doi.org/10.1007/s10887-012-9081-x

Hargadon, J. M., & Fuller, L. R. (2018). Dual Certification: It's a Win-win! *Strategic Finance*, 40-47.

Harrington, D. R., Kulasekera, K., Bates, R., & Bredahl, M. (2006). *Determinants of Student Performance in an Undergraduate Financial Accounting Class* (Working Paper n. 06, Department of Agricultural Economics and Business - University of Guelph).

Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, 18(1), 9-34. https://doi.org/10.1080/0260293930180102

Horn, A. S., Horner, O. G., & Lee, G. (2019). Measuring the effectiveness of two-year colleges: a comparison of raw and value-added performance indicators. *Studies in Higher Education*, 44(1), 151-169. DOI: 10.1080/03075079.2017.1349741.

Instituto Brasileiro de Geografia e Estatística. (2019). *Indicadores IBGE: Pesquisa Nacional por Amostra de Domicílios Contínua Primeiro Trimestre de 2019*. Brasília, DF, 2019. Retrieved from http://biblioteca.ibge.gov.br/visualizacao/periodicos/2421/pnact_2019_1tri.pdf.

Instituto Brasileiro de Geografia e Estatística. (2019a). *Produto Interno Bruto - PIB*. Retrieved from http://ibge.gov.br/pib.php.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2009). *Cálculo do Conceito ENADE*. Brasília: Inep, 2009. Retrieved from http://download.inep.gov.br/download/enade/2009/Nota_Tecnica_Conceito_Enade.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2009a). *Cálculo do Conceito Preliminar de Curso (CPC)*. Brasília: Inep, 2009. Retrieved from http://download.inep.gov.br/download/enade/2009/Nota_Tecnica_CPC.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2012). *NOTA TÉCNICA N° 029 DE 15 DE OUTUBRO DE 2012.* Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2011/nota_tecnica_indicadores_2011_2.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2013). *INDICADORES DE QUALIDADE DA EDUCAÇÃO SUPERIOR 2012*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2012/nota_metodologica_indicadores_2012.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2014). *Nota Técnica INEP/DAES/CONAES Nº 065*. Brasília: Inep. Retrieved from https://seavi.ufms.br/files/2013/04/Nota_Tecnica_No65_2014_Relatorio_CPA.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2014a). *NOTA TÉCNICA NO. 72: Cálculo do Conceito Preliminar de Curso referente a 2013*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2013/nota_tecnica_n_72_2014_calculo_cpc_2013.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2015). *ENADE 2015: Relatório Síntese de Área - Ciências Contábeis*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2015/ciencias_contabeis.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2015a). *Nota Técnica Daes/Inep no 57/2015.* Brasília: Inep, 2015. Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2014/nota_tecnica_daes_n582015_calculo_do_cpc2014.pdf. Acesso em 20 mai. 2019. (2015a).

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2015b). *Indicadores Financeiros Educacionais*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/indicadores-financeiros-educacionais.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2016). *Sinopse Estatística da Educação Superior 2015*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2017). *NOTA TÉCNICA Nº 3/2017/CGCQES/DAES.* Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2015/nota_tecnica_daes_n32017_calculo_do_cpc2015.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2017a). *NOTA TÉCNICA Nº 33/2017/CGCQES/DAES.* Brasília: Inep. Retrieved from

http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2016/nota_tecnica_n33_2017_cgcqes_daes_calculo_idd.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2017b). *NOTA TÉCNICA Nº 38/2017/CGCQES/DAES*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2016/nota_tecnica_n38_2017_cgcqes_daes_calculo_cpc.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018). *ENADE*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/web/guest/enade.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018a). *NOTA TÉCNICA Nº 16/2018/CGCQES/DAES*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/nota_tecnica_n16_2018_calculo_conceito-enade.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018b). *NOTA TÉCNICA Nº 17/2018/CGCQES/DAES*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/nota_tecnica_n17_2018_calculo_idd.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018c). *NOTA TÉCNICA Nº 18/2018/CGCQES/DAES*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/nota_tecnica_n18_2018_cpc2017.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018d). *NOTA TÉCNICA Nº 19/2018/CGCQES/DAES*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/nota_tecnica_n19_2018_igc2017.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018e). *Sinaes*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/sinaes.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018f). *Sinopse Estatística da Educação Superior 2017*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2019). *Documentos e Legislação*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/web/guest/documentos-e-legislacao12.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2019a). *Censo da Educação Superior: Notas Estatísticas 2018*. Brasília: Inep. Retrieved from http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/censo_da_educacao_superior_2018-notas_estatisticas.pdf.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2019b). *Resultados*. Brasília: Inep. Retrieved from http://portal.inep.gov.br/web/guest/educacao-superior/indicadores-de-qualidade/resultados.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019c). *Sinopse Estatística da Educação Superior 2018* [online]. Brasília: Inep. Retrieved from http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse.

Johnes, J. (2018). University Rankings: What do they really show? *Scientometrics*, 115(1), 585-606. https://doi.org/10.1007/s11192-018-2666-1

Johnston, M. P. (2014). Secondary Data Analysis: A Method of which the Time Has Come. *Qualitative and Quantitative Methods in Libraries*, 3, 619-626.

Kim, H., & Lalancette, D. (2013). *Literature Review on The Value-Added Measurement in Higher Education*. Paris: OECD. 52p.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The Collegiate Learning Assessment: Facts and fantasies. *Evaluation Review*, 31(5), 415-439. https://doi.org/10.1177/0193841X07303318

Klein, S., Freedman, D., Shavelson, R., & Bolus, R. (2008). Assessing School Effectiveness. *Evaluation Review*, 32(6), 511-525. https://doi.org/10.1177/0193841X08325948

Koedel, C., & Betts, J. R. (2009). *Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique* (Working Paper). Retrieved from https://economics.missouri.edu/working-papers/2009/wp0902_koedel.pdf.

Koedel, C., Mihaly, K., & Rockokk, J. E. (2015). Value-added modeling: a review. *Economics of Education Review*, 47, 180-195. https://doi.org/10.1016/j.econedurev.2015.01.006

Krause, Kerri-Lee. (2012). Addressing the wicked problem of quality in higher education: theoretical approaches and implications. *Higher Education Reseacrh & Development*, 31(3), 285-297. https://doi.org/10.1080/07294360.2011.634381

Kundu, G. K. (2016). Quality in Higher Education from Different Perspectives: a Literature Review. *International Journal for Quality Research*, 11(1), 17-34.

Lagrosen, S., Seyyed-Hashemi, R., &Leitner, M. (2004). Examination of the dimensions of quality in higher education. *Quality Assurance in Education*, 12(2), 61-69. https://doi.org/10.1108/09684880410536431

Laros, J. A., & Marciano, J. L. P. (2008). Análise Multinível aplicada ados dados do NELS:88. *Estudos em Avaliação Educacional*, 19(40). https://doi.org/10.18222/eae194020082079

Harvey, L., & Green, D. (1993). Defining Quality. *Assessment and Evaluation in Higher Education*, 18, 9-34. https://doi.org/10.1080/0260293930180102.

Harvey, L., & Williams, J. (2010). Fifteen Years of Quality in Higher Education (Part Two). *Quality in Higher Education*, 16(2), 81-113. https://doi.org/10.1080/13538322.2010.485722

Liu, L. (2011). Measuring value-added in higher education: conditions and caveats - results from using the Measure of Academic Proficiency and Progress (MAPP^{TM}). *Assessment and Evaluation in Higher Education*, 36(1), 81-94. https://doi.org/10.1080/02602930903197917

Liu, L. (2011a). Value-added assessment in higher education: a comparison of two methods. *Higher Education*, 61, 445-461. https://doi.org/10.1007/s10734-010-9340-8

Marçal, R.R., Matos, V. S., Carvalho, T.F.M., & Carvalho, M.S. (2019). Avaliações de Desempenho no Ensino Contábil Brasileiro: uma análise comparativa entre IES diante do exame de suficiência do CFC. *Revista de Administração Contabilidade e Economia*, 18(2), 363-384. https://doi.org/10.18593/race.19638

Marginson, S. (2014). University rankings and social science. *European Journal of Education*, 49(1), 45–59. https://doi.org/10.1111/ejed.12061

Martins, G.A., & Theóphilo, C.R. (2017). *Metodologia da Investigação Científica para Ciências Sociais Aplicadas* (3 ed). São Paulo: Altas.

McClain, C. J., Krueger, D. W., & Taylor, T. (1986). Northeast Missouri State University's value-added assessment program: A model for educational accountability. *International Journal of Institutional Management in Higher Education,* 10(3), 252-261.

Melguizo, T., & Wainer, J. (2016). Toward a set of measures of student learning outcomes in higher education: Evidence from Brazil. *Higher Education*, 72, 381-401. https://doi.org/10.1007/s10734-015-9963-x

Melguizo, T., Zamarro, G., Velasco, T., & Sanchez, F. J. (2017). The Methodological Challenges of Measuring Student Learning, Degree Attainment, and Early Labor Market Outcomes in Higher Education. *Journal of Research on Educational Effectiveness*, 27(2), 424-448. https://doi.org/10.1080/19345747.2016.1238985

Merton, R. K. (1968). The Matthew Effect in Science: the reward and communication systems of science are considered. *Science*, 159(3810), 56-63. https://doi.org/10.1126/science.159.3810.56

Milla, J., Martín, E. S., & Bellegem, S. V. (2016). Higher Education Value Added Using Multiple Outcomes. *Journal of Educational Measurement*, 53(3), 368-400. https://doi.org/10.1111/jedm.12114

Miller, G. J., & Nouri H. (2015). An examination of the relationship between obtaining AACSB accounting accreditation and certified public accountant (CPA) exam pass rates. *International Journal of Economics and Accounting*, 6(2), 179-194. https://doi.org/10.1504/IJEA.2015.069905

Miranda, G. J. (2011). *Relações entre as qualificações do professor e o desempenho discente nos cursos de graduação em contabilidade no Brasil* (Tese de Doutorado). Universidade de São Paulo, São Paulo, SP, Brasil.

Miranda, G. J., Araújo, T. S., & Marcelino, I. A. (2017). O absenteísmo acadêmico e suas consequências mais óbvias. *Gestão Universitária na América Latina*, 10(1), 172-189. https://doi.org/10.5007/1983-4535.2017v10n1p172 https://doi.org/10.5007/1983-4535.2017v10n1p172

Miranda, G. J., Casa Nova, S. P. C., & Cornachione, E. B. (2013). The Accounting Education Gap in Brazil. *Chinese Business Review*, 12, 361-372. https://doi.org/10.17265/1537-1514/2013.04.003

Miranda, G. J., Lemos, K. C. S., Oliveira, A. S., & Ferreira, M. A. (2015). Determinantes do Desempenho Acadêmico na Área de Negócios. *Meta: Avaliação*, 7(20), 175-209. https://doi.org/10.22347/2175-2753v7i20.264

Miranda, G. J., Mamede, S. P. N., Marques, A. V. C., & Rogers, P. (2014). Determinantes do Desempenho Acadêmico em Ciências Contábeis: Um Análise de Variáveis Comportamentais. *Anais do Congresso USP*, São Paulo, SP, Brasil, 14.

Mitchell, D. E., &Mitchell, R. E. (2003). The Political Economy of Education Policy: The Case of Class Size Reduction. *Peabody Journal of Education*, 78(4), 120-152. https://doi.org/10.1207/S15327930PJE7804_07

Monk, D. H. (1989). The Education Production Function: Its Evolving Role in Policy Analysis. *Educational Evaluation and Policy Analysis*, 11(1), 31-45 https://doi.org/10.3102/01623737011001031

Mustafa, S. T., & Chiang, D. (2006). Dimensions of quality in higher education: How academic performance affects university students' teacher evaluation. *The Journal of American Academy of Business,* 8(1), 294-303.

Nagle, B. M., Menk, K. B., & Rau, S. E. (2018). Which accouting program characteristics contribute to CPA exam success? A study of institutional factors and graduate education. *Journal of Accounting Education*, 45, 20-31. https://doi.org/10.1016/j.jaccedu.2018.09.003

National Research Council. *Getting Value out of Value-Added*: Report of a Workshop. Washington, DC: The National Academies Press.

Nelson, I. T., Quirin, J. J., Vendrzyk, V. P, & Kovar, S. (2006). Trends in Accounting Student Characteristics: Results from a 15-year Longitudinal Study at FSA Schools. *The FSA Archive*, 1-25.

Nonis, S. A., & Hudson, G. I. (2010). Performance of College Students: Impact of Study Time and Students Habits. *Journal of Education for Business*, 85(4), 229-238. https://doi.org/10.1080/08832320903449550

Organisation for Economic Cooperation and Development. (2019). *Education at a Glance 2019: OECD Indicators*. Retrieved from http://oecd.org/education/education-at-a-glance.

Organisation for Economic Cooperation and Development. (2018). *Measuring improvements in learning outcomes: Best practices to assess the value-added for schools*. Paris: OECD.

Ovaska-Few, S. (2016). What Large Firms Expect From New Accounting Grads: Recruiters weigh in on what they want to see from new hires. *Extra Credit*, Retrieved from <https://www.aicpa.org/interestareas/accountingeducation/newsandpublications/what-large-cpa-firms-expect-from-new-accounting-grads.html>.

Pathways Comission. (2012). *The Pathways Commission: Charting a National Strategy for the Next Generation of Accountants.* Florida: AAA. Retrieved from <http://commons.aaahq.org/posts/a3470e7ffa>

Pedrosa, R. H. L., Amaral, E., & Knobel, M. (2013). Assessing higher education learning outcomes in Brazil. *Higher Education Management and Policy*, 24(2), 55-71. https://doi.org/10.1787/hemp-24-5k3w5pdwk6br

Peleias, I. R., Silva, G. P., Segreti, J. B., & Chiroto, A. R. (2007). Evolução do ensino da contabilidade no Brasil: uma análise histórica. *Revista de Contabilidade e Finanças*, 30, 19-32. https://doi.org/10.1590/S1519-70772007000300003

Pike, G. R. (2016). Assessment Measures: Considerations when Using Value-Added Models in Higher Education Assessment. *Assessment Update*, 28(5), 8-10. https://doi.org/10.1002/au.30073

Prestes, E. M. T., & Fialho, M. G. G. (2018). Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba. *Ensaio: Avaliação e Políticas Públicas em Educação*, 26(100), 869-889. https://doi.org/10.1590/s0104-40362018002601104

Primi, R., Silva, M. C. R., & Bartolomeu, D. (2018). A validade do Enade para avaliação de cursos superiores: uma abordagem multinível. *Revista Examen*, 2(2), 128-151.

R CORE TEAM. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from: https://www.R-project.org/.

Ribeiro, S. L. S. et al. (2009). *Contando História: o Departamento de Contabilidade e Atuária – FEA/USP entre números e palavras*. São Paulo: D'Escrever Editora.

Rodgers, T. (2005). 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*, 13(2), 95-106. https://doi.org/10.1108/09684880510594355

Rodrigues, B. C. O., Resende, M. S., Miranda, G. J., & Pereira, J. M. (2016). Determinantes do desempenho acadêmicos dos alunos dos cursos de Ciências Contábeis no ensino à distância. *Enfoque: Reflexão Contábil*, 35(2), 139-153. https://doi.org/10.4025/enfoque.v35i2.30105

Rodrigues, B. C. O., Miranda, G. J., Santos, N. A., & Pereira, J. M. (2017). A Instituição, a Coordenação e o Desempenho Acadêmico dos Alunos de Ciências Contábeis. *Contabilidade, Gestão e Governança*, 20(2), 231-251. https://doi.org/10.21714/1984-3925_2017v20n2a4

Sanders, W. L., & Horn, S. P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256. https://doi.org/10.1023/A:1008067210518

Santos, N. A. (2012). *Determinantes do Desempenho Acadêmico dos Alunos dos Cursos de Ciências Contábeis* (Tese de Doutorado). Universidade de São Paulo, São Paulo, SP, Brasil.

Santos, N. A., Cunha, J. V. A. D., & Cornachione Junior., E. B. (2006). Análise do Desempenho dos Cursos de Ciências Contábeis do Estado de Minas Gerais no ENADE/2006. In: *Anais do Congresso ANPCONT*, São Paulo, SP, Brasil, 1.

Shavelson, R. J., Domingue, B. W., Mariño, J. P., Mantilla, A. M., Forero, A. M., & Wiley, E. E. (2016). On the Practices and challenges of measuring higher education value-added: the case of Colombia. *Assessment & Evaluation in Higher Education*, 41(5), 695-720. http://dx.doi.org/10.1080/0260293.8.2016.1168772

Shelton, M. L., Thompson, S., & Serrett, R. (2012, nov.). A Uniform CPA Exam, but No-So-Uniform Requirements: Education and License Rules Vary among States. *The CPA Journal*, 62-65.

Simons, K.A., Lowe, D.R, & Stout, D.E. (2003). Comprehensive literature review: factors influencing choice of accounting as a major. *Academy of Business Education Conference*, Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.543.1218&rep=rep1&type=pdf.

Souza, E. S. (2008). *ENADE 2006: Determinantes do Desempenho dos Cursos de Ciências Contábeis* (Dissertação de Mestrado). Programa Multiinstitucional e Inter-Regional de Pós-Graduação em Ciências Contábeis, Universidade de Brasília, Brasília, DF, Brasil.

Srikanthan, G., & Dalrymple, J. F. (2002). Developing A Holistic Model For Quality in Higher Education. *Quality in Higher Education*, 8(3), 215-224. https://doi.org/10.1080/1353832022000031656

Srikanthan, G., & Dalrymple, J. (2003). Developing alternative perspectives for quality in higher education. *International Journal of Educational Management*, 17(3), 126-136. https://doi.org/10.1108/09513540310467804

Steedle, J. T. (2012). Selecting value-added models for post-secondary institutional assessment. *Assessment & Evaluation in Higher Education*, 37(6), 637-652. https://doi.org/10.1080/02602938.2011.560720

Steedle, J., Kugelmass, H., & Nemeth, A. (2010). What do they measure? Comparing three learning outcomes assessments. *Change*, 42(4), 33-37. https://doi.org/10.1080/00091383.2010.490491

Taylor Research and Consulting Group. (2000). *AICPA Student and Academic Research Study***.** Final Quantitative Report, New York: AICPA.

Tood, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113, 3-33. https://doi.org/10.1111/1468-0297.00097

Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of higher education learning outcomes: Feasibility study report*. Paris: OECD.

Trinkle, B. S., Scheiner, J., Baldwin, A. A. & Krull Jr, G. W. (2016). Gender and other Determinants of CPA Exam Success: A Survival Analysis. *The Accounting's Educator Journal*, 27, 101-117.

Ugrin, J. C., & Honn, D. D. (2018). Advanced placement and CPA exam performance: Implications for recruitment of quality students into college accounting programs. *Advances in Accounting*, 42, 96-109. https://doi.org/10.1016/j.adiac.2018.04.002

United States of America. Tennessee Department of Education. (2019) *SAS® EVAAS®: Statistical Models and Business Rules of TVAAS Analyses.* State Report. Retrieved from https://tvaas.sas.com/support/TVAAS-Statistical-Models-and-Business-Rules.pdf.

Usher, A., & Medow, J. (2009). A global survey of university rankings and league tables. In: Kehm, B. M., & Stensaker, B. (2009). *University rankings, diversity, and the new landscape of higher education* (p. 3–18). Rotterdam: Sense Publishers. https://doi.org/10.1163/9789087908164_002

Verhine, R. E. (2015). Avaliação e regulação da educação superior: uma análise a partir dos primeiros 10 anos do Sinaes. *Avaliação*, 20(3), 603-619. https://doi.org/10.1590/S1414-40772015000300003

Vogt, M., Degenhart, L., & Biavatti, V. T. (2016). Relação entre Formação Docente, Metodologias de Ensino e Resultados do Exame Nacional de Desempenho dos Estudantes de Ciências Contábeis. *Revista Catarinense de Ciência Contábil*, 15(45), 63-77. http://dx.doi.org/10.16930/2237-7662/rccc.v15n45p63-77.

Voluntary System of Accountability. *About VSA Analytics*. Retrieved from https://www.vsaanalytics.org/about.

Vroeijenstijn, A. I. (2003). Towards a quality model for higher education. *Journal of Philippine Higher Education Quality Assurance*, 11(1), 78-94.

Xu, Z., Hannaway, J., &Taylor, C. (2011). Making a difference? The effects of Teach for America in high school. *Journal of Policy Analysis and Management*, 30(3), 447-469. https://doi.org/10.1002/pam.20585

Yeh, Stuart S. (2010) Understanding and addressing the achievement gap through individualized instruction and formative assessment, *Assessment in Education: Principles, Policy & Practice*, 17(2), 169-182, https://doi.org/10.1080/09695941003694466

Yeh, S. S. (2020). Educational Accountability, Value-Added Modeling, and the Origin of the Achievement Gap. Education and Urban Society. https://doi.org/10.1177/0013124519896823

Yunker, J. A. (2005) The dubious utility of the value-added concept in higher education: the case of accounting. *Economics of Education Review*, 24, 355-367. https://doi.org/10.1016/j.econedurev.2004.06.003

Zhang, L. (2009). A value-added estimate of higher education quality of US states. *Education Economics*, 17(4), 469-489. https://doi.org/10.1080/09645290701838079

Zlatkin-Troitschanskaia, O., Pant, H. A., & Coates, H. (2016). Assessing student learning outcomes in higher education: Challenges and international perspectives. *Assessment & Evaluation in Higher Education*, 41(5), 655–661. https://doi.org/10.1080/02602938.2016.1169501

Zoghbi, A. C. P., Moriconi, G. M., & Oliva, B. T. (2010). Aumentando a eficácia e a eficiência da avaliação do ensino superior: a relação entre o Enem e o Enade. *Estudos em Avaliação Educacional*, 45. https://doi.org/10.18222/eae214520102024

APPENDIX

Level of Analysis	Name of Variable (Code)	Type of Variable	Description of Variable
Response Variable	Overall Enade score (c)	Numerical	0 to 100
Explanatory variable (student related)	Mean Enem score ($\overline{Enem}$)	Numerical	0 to 1000
	Score in Enem section 'Natural Sciences' (CN)	Numerical	0 to 1000
	Score in Enem section 'Humanities' (CH)	Numerical	0 to 1000
	Score in Enem section 'Mathematics' (MT)	Numerical	0 to 1000
	Score in Enem section 'Languages' (LT)	Numerical	0 to 1000
	Ratio of parents with high education (w)	Numerical	0 to 1
Explanatory variable (program related)	Standardized score for program's faculty members with a doctoral degree (Doc)	Numerical	1 to 5
	Standardized score for program's faculty members with a master's degree (Me)	Numerical	1 to 5
	Standardized score for faculty members' employment contract in the program (RT)	Numerical	1 to 5
	Standardized score for program infrastructure (IFF)	Numerical	1 to 5
	Standardized score for program's pedagogical teaching structure (ODP)	Numerical	1 to 5
	Standardized score for program's opportunity for further training/learning (OAF)	Numerical	1 to 5