UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Kelly Alice Barbosa Macêdo

Valor acrescentado, eficácia diferencial e equidade social: evidências longitudinais da educação básica no Brasil

Juiz de Fora
2026

Kelly Alice Barbosa Macêdo

Valor acrescentado, eficácia diferencial e equidade social: evidências longitudinais da educação básica no Brasil

Tese apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Economia da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Doutora em Economia.
Área de concentração: Economia.

Orientador: Prof. Dr. Marcel de Toledo Vieira

Co-orientadora: Prof.ª Drª Maria Eugénia Ferrão

Juiz de Fora

2026

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Macêdo, Kelly Alice Barbosa.
Valor acrescentado, eficácia diferencial e equidade social : evidências longitudinais da educação básica no Brasil / Kelly Alice Barbosa Macêdo. -- 2026.
130 p. : il.

Orientador: Marcel de Toledo Vieira
Coorientador: Maria Eugénia Ferrão
Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2026.

1. Valor acrescentado. 2. Modelo multinível. 3. Eficácia escolar. 4. Equidade social. 5. Avaliação educacional. I. Vieira, Marcel de Toledo , orient. II. Ferrão, Maria Eugénia , coorient. III. Título.

Kelly Alice Barbosa Macêdo

Valor acrescentado, eficácia diferencial e equidade social: evidências longitudinais da educação básica no Brasil

Tese apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutora em Economia. Área de concentração: Economia.

Aprovada em 04 de março de 2026.

BANCA EXAMINADORA

Dr. Marcel de Toledo Vieira - Orientador
Universidade Federal de Juiz de Fora

Dr.ª Maria Eugénia Ferrão - Coorientadora
Universidade da Beira Interior

Dr. Admir Antonio Betarelli Junior
Universidade Federal de Juiz de Fora

Dr. Ricardo da Silva Freguglia
Universidade Federal de Juiz de Fora

Dr. Alcino Fernando Ferreira Pinto Couto
Universidade da Beira Interior

Dr.ª Denise Britz do Nascimento Silva

Instituto Brasileiro de Geografia e Estatística

Juiz de Fora, 11/02/2026.

Documento assinado eletronicamente por **Marcel de Toledo Vieira**, **Professor(a)**, em 05/03/2026, às 11:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Admir Antonio Betarelli Junior**, **Professor(a)**, em 05/03/2026, às 13:20, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Maria Eugénia Neto Ferrão da Silva**, **Usuário Externo**, em 05/03/2026, às 13:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Alcino Fernando Ferreira Pinto Couto**, **Usuário Externo**, em 05/03/2026, às 15:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Denise Britz do Nascimento Silva**, **Usuário Externo**, em 05/03/2026, às 16:44, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ricardo da Silva Freguglia**, **Professor(a)**, em 09/03/2026, às 17:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2871068** e o código CRC **ED48B2D2**.

AGRADECIMENTOS

Uma tese não se constrói sozinha.

Este trabalho representa a conclusão de um processo longo e desafiador. Neste espaço, expresso minha profunda gratidão a Deus e a todas as pessoas que estiveram ao meu lado, apoiando-me e inspirando-me ao longo dessa jornada. Sou especialmente grata àqueles que escolhi ter comigo e que, igualmente, escolheram estar presentes, contribuindo para tornar possível esta conquista. Sou grata a todos os colegas que compartilharam comigo este percurso e aos que conheci durante a trajetória.

À minha família, agradeço pelo amor, apoio e pelo esforço em me proporcionar as melhores condições para que eu pudesse chegar até aqui. Ao meu companheiro, meu sincero agradecimento pelo incentivo constante e pelo suporte durante as longas madrugadas de estudo.

Agradeço ao meu orientador, professor Marcel de Toledo Vieira, por seu apoio, paciência, compreensão e confiança. Sou especialmente grata por sua humanidade e gentileza, na condução da orientação. Mesmo com uma agenda extremamente ocupada, mostrou-se sempre disponível e dedicado, incentivando de forma constante meu desenvolvimento acadêmico. Sua orientação foi fundamental para o amadurecimento teórico e metodológico desta pesquisa.

Sou profundamente grata também à professora Maria Eugénia Ferrão, minha coorientadora, cuja orientação trouxe estrutura, clareza e entusiasmo ao meu trabalho. Acompanhou-me com paciência, atenção aos detalhes e compromisso com meu aprendizado e desenvolvimento. Sou imensamente grata por sua confiança e pela generosidade intelectual com que compartilhou conhecimentos ao longo de todo o percurso. Sinto-me privilegiada por ter contado com a orientação desses dois professores, cuja colaboração formou uma equipe engajada e estimulante, essencial para meu desenvolvimento acadêmico e meu crescimento como pesquisadora.

Agradeço igualmente aos membros da banca de qualificação e defesa — professores Denise Britz, Alcino Couto, Admir Betarelli e Ricardo Freguglia — pela disponibilidade, pelo rigor acadêmico e pelas valiosas contribuições oferecidas ao aprimoramento deste trabalho. Agradeço ao PPGE/UFJF pelo acolhimento e a todos os professores do programa que contribuíram, direta ou indiretamente, para a realização e a conclusão deste ciclo. Registro também meu agradecimento ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e a professora Maria Eugénia apelos microdados utilizados nesta pesquisa. Por fim, agradeço pelo apoio financeiro concedido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) — Brasil, Código de Financiamento 001.

RESUMO

Esta tese analisa o uso dos modelos de Valor Acrescentado (VA) na avaliação educacional, articulando seus fundamentos teóricos e lacunas metodológicas a uma aplicação empírica sobre eficácia diferencial e equidade social no sistema educacional brasileiro. O trabalho está organizado em três ensaios. O primeiro, revisão de escopo (1971–2024), e o segundo, revisão sistemática ibero-americana baseada no protocolo PRISMA, mostram que a literatura em países em desenvolvimento ainda é incipiente e marcada por limitações metodológicas, sobretudo pelo predomínio de modelos simplificados de dois níveis que tendem a subestimar fatores intraescolares e contextuais. Para superar essas limitações, o terceiro ensaio investiga a eficácia diferencial e a equidade social no Brasil, contrastando os estados do Maranhão e de Minas Gerais. Utilizando dados longitudinais pareados do SAEB (2011–2015), a análise combina um modelo multinível linear de cinco níveis (aluno, turma, escola, município e estado) para estimar a proficiência e um modelo logístico multinível para analisar o sucesso escolar e a progressão regular. Os resultados mostram que mais de 90% da variação no desempenho ocorre dentro das escolas, concentrando-se nos níveis do aluno e da turma. A eficácia escolar diferencial é limitada, indicando que as escolas têm baixa capacidade de compensar os efeitos da origem socioeconômica e das defasagens de aprendizagem prévia. O modelo binário revela ainda dinâmicas de desigualdade não captadas apenas pelos testes padronizados: embora as meninas apresentem menor proficiência em Matemática, elas têm quase o dobro de chances de sucesso escolar, e as penalidades raciais na progressão são mais fortes em contextos educacionais mais estruturados, como Minas Gerais. Conclui-se que a reprodução das desigualdades educacionais no Brasil ocorre predominantemente por meio de processos intraescolares e mecanismos de seleção dentro das próprias instituições. Assim, os modelos de VA devem ser compreendidos não apenas como instrumentos de responsabilização, mas como ferramentas analíticas capazes de orientar políticas públicas voltadas à promoção da equidade educacional.

Palavras-chave: valor acrescentado; modelo multinível; eficácia escolar; equidade social; avaliação educacional.

ABSTRACT

This dissertation examines the use of Value-Added (VA) models in educational evaluation, linking their theoretical foundations and methodological gaps to an empirical application on differential school effectiveness and social equity in the Brazilian education system. The study is structured around three progressive essays. The first, a scoping review (1971–2024) and the second, an Ibero-American systematic review based on the PRISMA protocol, show that the literature in developing countries remains limited and characterized by methodological constraints, particularly the predominant use of simplified two-level models that tend to underestimate within-school and contextual factors. To address these limitations, the third essay investigates differential school effectiveness and social equity in Brazil through a comparison of the states of Maranhão and Minas Gerais. Using matched longitudinal data from the Sistema de Avaliação da Educação Básica (SAEB) for the period 2011–2015, the analysis combines a five-level linear multilevel model (student, classroom, school, municipality, and state) to estimate academic proficiency with a multilevel logistic model to examine school success and regular student progression. The results show that more than 90% of the variation in student performance occurs within schools, concentrating primarily at the student and classroom levels. Differential school effectiveness is limited, indicating that schools have a constrained capacity to compensate for the effects of socioeconomic background and prior learning deficits. The binary model further reveals dynamics of inequality that are not captured by standardized test scores alone: although girls display lower proficiency in Mathematics, they have nearly twice the odds of school success, and racial penalties in student progression are more pronounced in more structured educational contexts, such as Minas Gerais. The findings suggest that the reproduction of educational inequalities in Brazil operates predominantly through within-school processes and selection mechanisms within institutions themselves. Accordingly, VA models should be understood not only as accountability metrics but also as analytical tools that highlight the urgent need for public policies focused on internal school processes in order to promote genuine educational equity.

Keywords: value-added; multilevel model; school effectiveness; social equity; educational evaluation.

CONTENTS

VALOR ACRESCENTADO, EFICÁCIA DIFERENCIAL E EQUIDADE SOCIAL: EVIDÊNCIAS LONGITUDINAIS DA EDUCAÇÃO BÁSICA NO BRASIL

1. GENERAL INTRODUCTION

Persistent educational inequalities observed across different countries and contexts demonstrate that the challenges facing education systems extend beyond access to school and include the quality of learning and equitable distribution of educational opportunities. Recent evidence suggests stagnation and regressions in learning outcomes, coupled with a strong correlation between student performance and socioeconomic status (OECD, 2018; UNESCO, 2024). This places the focus of the debate on not only how much students learn, but also on the extent to which schools contribute to that learning while considering students' initial conditions and contextual factors. This scenario highlights the necessity of analytical tools that can distinguish the specific role of schools and education systems in promoting learning, especially in contexts marked by structural inequalities.

Although educational assessment has evolved from an input-focused analysis toward greater attention to learning outcomes, the isolated use of raw proficiency indicators, such as standardized test averages, presents significant limitations for measuring educational effectiveness (Fernandes & Gremaud, 2009), both at the school and teacher levels. These indicators tend to largely reflect students' socioeconomic profiles, making it difficult to distinguish between the influence of the family context and the value added by the school (Coleman, 1968; Sirin, 2005). This methodological limitation can lead to inaccurate diagnoses, penalizing schools that serve more vulnerable populations while overestimating those in more advantaged contexts. In this regard, there is growing recognition that educational assessments should incorporate mechanisms that explicitly account for students' prior performance (Ferrão, 2018; Ferrão & Couto, 2013; Jerrim et al., 2019; Muñoz-Chereau, 2019; Muñoz-Chereau & Thomas, 2016; Taut et al., 2016) and the context in which learning occurs (Albernaz et al., 2002; Alves et al., 2007; Bonamino et al., 2010; Casillas, 2006; César & Soares, 2001; Coleman, 1968; Creemers, 2005; Medeiros & Oliveira, 2014; Mortimore et al., 1988; Reynolds et al., 2014; Scheerens & Bosker, 1998), thereby allowing the specific contribution of schools and education systems to academic progress to be isolated, independently of students' background conditions.

It is within this debate that the concept of Value-Added (VA) emerges, originally developed in the field of economics of education and later incorporated into educational assessment. The VA framework shifts the focus from comparing final outcomes to controlling learning progress over time, allowing for the estimation of the relative contribution of schools, teachers, and education systems to student learning while controlling individual and contextual characteristics (Bryk & Weisberg, 1976; Hanushek, 1971). Methodologically, this framework has been predominantly operationalized through multilevel models, which recognize the hierarchical nature of educational data and allow for the decomposition of student performance variance (Levy et al., 2019). Over the past decades, VA models have assumed a central position in both academic research and assessment systems, as well as in public policy debates in several countries, being used for monitoring, accountability, resource allocation, and support for school improvement (Amrein-Beardsley et al., 2013; Braun, 2013; Ferrão & Couto, 2014; Hanushek, 2019; Koedel et al., 2015). This widespread adoption, combined with the diversity of uses and interpretations of VA, underscores the importance of examining its conceptual foundations, analytical assumptions, and interpretive limitations.

Despite their widespread use, VA models remain the subject of significant theoretical and methodological debate. The literature highlights challenges related to the validity, stability, and interpretation of estimates (Baker et al., 2010; Ferrão, 2014; Paufler & Amrein-Beardsley, 2014; Rothstein, 2009, 2010; Soares et al., 2017), as well as risks associated with their use in high-stakes decision-making (Newton et al., 2010). Moreover, although the field has advanced considerably in developed countries, our understanding of how schools mitigate structural inequalities in developing contexts remains limited. Traditional cross-sectional approaches have struggled to isolate the true contribution of schools from students' socioeconomic backgrounds in such contexts. In settings characterized by high levels of social inequality and pronounced school segregation, the direct adoption of VA models developed for other institutional arrangements may present important limitations, as key assumptions—such as the stability of school effects and the persistence of learning over time—do not always hold empirically in contexts of greater socioeconomic vulnerability (Andrabi et al., 2011; Muñoz-Chereau, 2019). Therefore, it is essential that analyses incorporate local institutional and socioeconomic specificities to avoid reinforcing preexisting inequalities rather than contributing to the promotion of social equity.

These limitations are even more pronounced in the Ibero-American context, which encompasses education systems at varying stages of institutional development. Although research on VA models has expanded, studies show considerable methodological heterogeneity, geographic concentration, and reliance on indicators limited to standardized test performance (Ferrão, 2022a; Murillo & Martínez-Garrido, 2019). The limited exploration of more complex hierarchical structures and broader educational indicators constrains understanding of the role of schools in promoting both educational effectiveness and equity across diverse socioeconomic contexts. These theoretical, methodological, and contextual limitations highlight the need for an integrated and critical analysis of VA models, particularly in settings marked by persistent inequalities.

In this context, discussions on educational effectiveness, differential effectiveness, and social equity are particularly important. Evidence suggests that schools do not affect all students equally and that their capacity to promote academic progress varies according to students' socioeconomic backgrounds and contextual conditions (Reynolds et al., 2014; Sammons, 2007). Understanding these differences during the early years of basic education – a critical stage for the school's role in reducing educational and social inequalities (Ferrão et al., 2018) – helps identify school practices associated with closing the achievement gap and informs the debate on education policies that address the structural conditions of education systems.

The overall objective of this dissertation is to critically examine the use of VA models in educational assessment. This examination links the models' conceptual foundations and patterns of empirical application across different institutional contexts with their implications for analyzing differential effectiveness and social equity in education systems. The focus is on basic education, a stage critical to schools' role in reducing educational and social inequalities (Ferrão et al., 2018) in contexts marked by profound social disparities. This thesis examines VA models, from their conceptual foundations to their empirical applications in Ibero-American contexts, as analytical tools for diagnosing differential school effectiveness and social equity in education systems characterized by structural inequalities.

The dissertation is organized into three essays that reflect a progressive analytical trajectory. The first essay examines the conceptual and historical foundations of VA models, highlighting existing gaps and ongoing theoretical debates, and has been accepted for publication in Journal Ensaio: Avaliação e Políticas Públicas em Educação. The second essay, accepted for publication in Journal Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación (REICE), investigates the empirical application of these models in the Ibero-American context, addressing associated methodological limitations and heterogeneities.[1] The third essay, which has been submitted for publication in Revista de Educación, focuses on Brazilian basic education using data from the 2011-2015 period, analyzing differential school

[1] The first and second essays were accepted for publication in January 2026, and the third essay was submitted in January 2026.

effectiveness and social equity, with particular emphasis on comparisons across federal units and the decomposition of performance variance at multiple levels.

The empirical analysis was initially planned for the periods 2015–2019 and 2019–2023. However, operational constraints in accessing restricted microdata led to a more limited scope for this thesis, covering 2011-2015. This period still allows for a robust examination of student learning outcomes and the structural mechanisms underlying educational inequality in Brazil.

This thesis makes significant contributions by integrating conceptual, methodological, and empirical dimensions. From a theoretical and conceptual standpoint, it clarifies the foundations of VA models within international literature and examines their application in Ibero-America, identifying gaps in regional research. Methodologically, it demonstrates the importance of multilevel models for variance decomposition, the analysis of differential school effectiveness, and social equity, particularly in contexts marked by inequality. Empirically, it provides evidence of differential school effectiveness and regional disparities in Brazil, revealing mechanisms of inequality often overlooked by simplified approaches. From a policy perspective, it informs decisions regarding resource allocation, support strategies for vulnerable schools, and teacher development programs aimed at promoting equity and improving learning outcomes.

This integration deepens our understanding of the role of schools in contexts marked by structural inequalities, highlighting the potential of VA models as robust analytical tools for informing education policies aimed at improving school effectiveness, rather than merely serving as measurement instruments. The following chapters expand on this investigation, tracing the analysis from the theoretical clarification of VA models to their practical application in various contexts, as well as the assessment of differential school effectiveness and social equity in Brazil.

2. ESSAY 1: A JOURNEY THROUGH VALUE ADDED IN EDUCATION: INSIGHTS AND FUTURE DIRECTIONS

ABSTRACT

Value-Added (VA) models are widely used to evaluate educational effectiveness, yet their conceptual and methodological development remains insufficiently systematized. This study conducts a scoping review, complemented by an exploratory bibliometric analysis of studies published between 1971 and 2024, focusing on the seminal contributions of Hanushek (1971) in education economics and Bryk and Weisberg (1976) in educational statistics. The findings reveal persistent gaps in the literature, notably the limited evidence from developing countries and the underexplored roles of school leadership and causal inference within VA frameworks. The review highlights a growing diversification of methodological approaches alongside an increasing alignment with accountability policies. From a policy perspective, the results underscore the need for context-sensitive VA applications, stronger school leadership, improved longitudinal data systems, and closer collaboration between researchers and policymakers to support more effective and equitable educational interventions.

Keywords: educational evaluation; educational effectiveness; school effectiveness; social equity; scoping review.

2.1 INTRODUCTION

The term “value-added” is commonly used in economics to describe the value created throughout a production process, reflecting the contribution of each stage to the generation of new value. Eric Hanushek, a distinguished economist specializing in education and a pivotal figure in discussions on educational reforms and public policy, was a pioneer in applying this concept to the educational context. In his 1971 study, titled “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data”, Hanushek examined how teacher and classroom characteristics influence students' academic performance over time. This work has become a seminal contribution to the field of education economics by introducing the concept of Value-Added (VA) to educational evaluation.

> “Families obviously have considerable impact on education through physical conditions, attitude formation and direct involvement in the educational process. Since these factors tend to be highly correlated with socioeconomic status of the family, this aspect of education is proxied by father’s occupation and family structure. The influence of peers is much the same as that from the family, and, thus, this aspect is proxied by aggregate measures of the socioeconomic status of individuals in a given class or school. While innate abilities are included in the conceptual model, there is no direct measure of this aspect. However, there is reason to believe that biases in the school parameters due to this missing variable are minimal. First, the model with initial achievement measures the “value added” of various inputs and biases will occur only if the missing portion of innate abilities is correlated with the rate of learning (as opposes to the level). Second, at least for whites, it is reasonable to assume that this factor is captured fairly well in the family background variables. This is the case if innate abilities tend to be hereditary and if social mobility is highly correlated with ability. Severe problems, at least in the school portion of the model, do not arise unless there is a mechanism which leads to the correlation of the ‘nonhereditary’ part of innate abilities and specific school resources.” (Hanushek, 1971, p. 281-282)

Hanushek (1971) introduced a multiple regression framework to estimate VA in education, designed to measure the contribution of schools and teachers to student learning while accounting for initial differences in achievement. The model incorporates variables such as family circumstances, socioeconomic status, and peer influences, which are typically proxied by father’s occupation, family structure, and aggregate class or school socioeconomic measures. Although innate abilities cannot be directly observed, the framework assumes that any resulting bias in school estimates is likely minimal, except in cases where nonhereditary components of ability are systematically correlated with specific school resources.

In this context, VA models are directly linked to empirical approaches that condition current achievement on prior achievement, with the aim of capturing the contribution of educational inputs to students’ rate of learning rather than to their absolute level of performance. VA models do not correspond to a specific statistical technique; instead, they are characterized by their analytical purpose and by the interpretation of estimated effects. VA, therefore, refers to the incremental gains in achievement that can be attributed to educational inputs, beyond what would be expected given students' starting points.

Within this VA framework, the work of Bryk and Weisberg (1976) represents a seminal contribution to the application and formalization of the VA concept in education. They were among the first researchers to formalize the VA term in the field. In their study, Value-Added Analysis: A Dynamic Approach to the Estimation of Treatment Effects, they proposed a methodology for evaluating educational interventions when randomized experiments are not feasible, modeling the expected growth attributable to the intervention to estimate its VA. This methodology allowed for a direct assessment of program effects and played a key role in advancing both the development and practical application of the VA concept in educational evaluation.

> "Our alternative view of educational processes focuses on individual growth. In the ideal situation, we would like to have available complete descriptions of growth under "natural" conditions. Thus for each subject, prior to the implementation of the program, the expected outcome could be predicted. The observed value could then be compared with the predicted value. The difference, or value-added, represents the additional benefit attributable to the program, and is thus a natural measure of treatment effect." (Bryk & Weisberg, 1976, p. 131-132)

Taken together, Hanushek (1971) and Bryk and Weisberg (1976) helped establish the logic of VA approaches by conceptualizing VA as learning growth relative to prior performance. Their contributions shifted educational evaluation away from static outcome levels toward estimating the incremental effects of schools, teachers, and educational interventions on student learning. This perspective formed the analytical basis for later research on educational effectiveness, understood here as the capacity of educational systems to generate learning gains beyond students' initial characteristics and analytically decomposed into school effectiveness and teacher effectiveness.

Methodologically, the field has progressively converged toward the use of multilevel models and longitudinal designs as widely adopted approaches for capturing both the hierarchical structure of educational systems and individual learning trajectories (Goldstein, 2003; Koedel et al., 2015; Soares et al., 2017). The literature on VA models (Amrein-Beardsley et al., 2016; Everson, 2017; Ferrão & Couto, 2013; Levy et al., 2019; Saunders, 1999, 2000) has addressed the relationship between students' SES and achievement showing inspiring results that demonstrate the effectiveness of some schools (Albernaz et al., 2002; Alves & Franco, 2008; Brooke et al., 2014; Charalambous et al., 2018; Kyriakides et al., 2019; Sammons, 2007; Sammons et al., 1995; Soares et al., 2017; Thomas et al., 2007), and in particular their differential effectiveness (Palardy, 2008; Strand, 2010, 2012; Twersky, 2022) in closing the gaps. However, concerns on VA methodological issues, purposes and uses are still on the agenda of scientific research (AERA-American Educational Research Association, 2015; Morganstein & Wasserstein, 2014).

In this regard, this first essay aims to examine the evolution of the concept of VA in education by analyzing the articles that first introduced and applied the term within the fields of educational economics and educational statistics. Covering publications from 1971 to 2024 and drawing on research conducted in the Web of Science and Scopus databases, a scoping review was adopted as the most appropriate strategy to identify patterns, emerging trends, shifts in research priorities, as well as persistent gaps in literature. By tracing the theoretical and methodological development of VA over time, this essay establishes the conceptual and historical foundation of the thesis, clarifying why an understanding of the origins, and evolution of VA models is essential before examining their application in contexts marked by high levels of inequality.

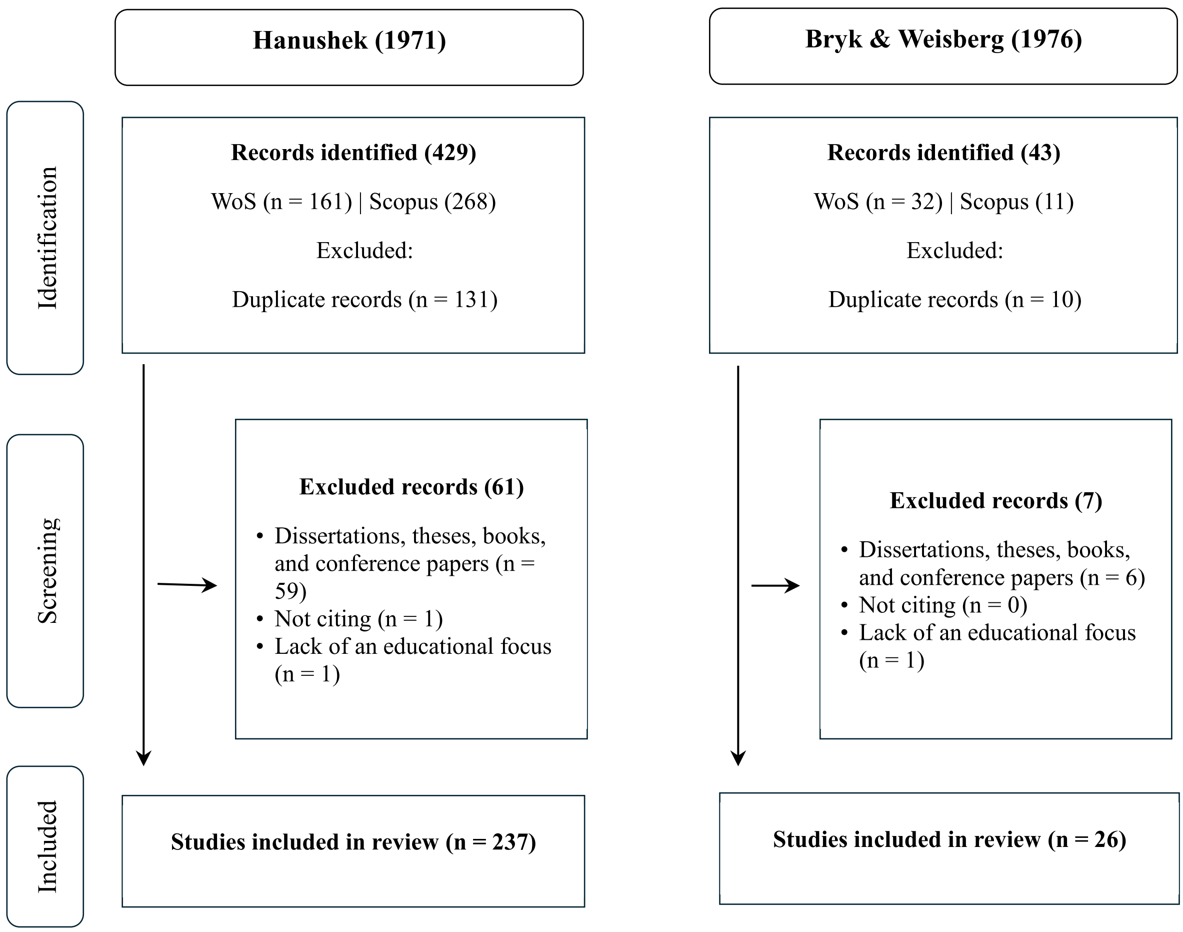
This essay is organized into three sections. The next section describes the methodology and the process of study selection. The third section presents the results and discussion, while the final section provides the study's conclusions.

2.2 METHODOLOGY

In line with the objective of mapping the development of VA research in education, this study addresses the following research question: What are the main conceptual, methodological, and thematic developments in VA research in education since its introduction by Hanushek (1971) and Bryk and Weisberg (1976)? To address this question, we have chosen to employ the scoping review method, which imparts an exploratory nature to the study and contributes to the literature by identifying trends and gaps based on existing research, without critically evaluating individual studies. This method covers a broader range of studies and methodologies compared to systematic reviews (Arksey & O'Malley, 2005; Braunack-Mayer et al., 2020; Pham et al., 2014). The scoping review is conducted following a research protocol that ensures rigorous and transparent practices. This protocol provides a clear framework for the review, detailing the research question, article search strategies, inclusion and exclusion criteria, and methods for data screening, extraction, and analysis.

In the initial stage of the study, research was conducted for studies citing Hanushek (1971) and subsequently for those citing Bryk and Weisberg (1976) within the Scopus and Web of Science databases. These databases were used to ensure comprehensive interdisciplinary coverage and to capture a broad spectrum of studies. As part of this scoping review, an exploratory bibliometric analysis was conducted using the Bibliometrix package in R (Aria & Cuccurullo, 2017) and its graphical interface Biblioshiny. Data was collected on July 5, 2024, and the search results were downloaded in BibTeX format. Three bibliometric techniques were employed as exploration stages within the broader scoping review: (1) Annual Scientific Production, to examine temporal trends in research output; (2) Collaboration Network Analysis, to explore co-authorship and institutional collaboration patterns; and (3) Thematic Mapping of Terms, to identify central, emerging, and peripheral research themes across the literature. These analyses provided an overview of trends, connections, and thematic patterns and guided the selection of studies for the subsequent detailed scoping review.

Figure 2.1 – Identification of studies via databases



Source: Own elaboration (2026).

To ensure that only studies relevant to the research objectives were considered, the inclusion criteria required: (1) articles in Portuguese, English, and Spanish; (2) specific citation of Hanushek (1971) between 1971 and 2024, or Bryk and Weisberg (1976) between 1976 and 2024, to capture the influence of these studies; and (3) educational focus. To avoid including records that could compromise the focus of the review, the exclusion criteria were applied: (1)

dissertations, theses, books, and conference papers; (2) studies not citing any of the specified works; and (3) studies lacking an educational focus.

A search of the Scopus and Web of Science (WoS) databases identified 429 articles citing Hanushek (1971) and 43 articles citing Bryk & Weisberg (1976). Figure 2.1 provides a detailed account of the article selection process. All studies included in the review were initially deemed relevant due to their citation of the specified authors, thereby eliminating the need for title and abstract screening. Thus, after removing duplicates, the sample for evaluating the development of the VA concept in studies citing Hanushek (1971) consisted of 298 original articles, of which 61 were eliminated based on exclusion criteria 1 (n = 59) and 2 (n = 1), respectively. The sample studies citing Bryk & Weisberg (1976), after removing duplicates, consisted of 33 original studies, of which 7 were eliminated due to exclusion criteria 1 (n = 6) and 3 (n =1).

The categories of analysis were year of publication, objective of the article, context in which Hanushek (1971) or Bryk & Weisberg (1976) was cited, methodology, main findings, and the impact factor of the journal in which it was published. The analysis of this information aims to fulfill our goal of providing a comprehensive overview of the development of the VA concept in studies citing Hanushek (1971) and Bryk & Weisberg (1976) in education literature.

2.3 RESULTS AND DISCUSSION

This section provides a comprehensive review of studies citing Hanushek (1971) and Bryk and Weisberg (1976), examining the evolution and key characteristics of the VA concept while identifying patterns, trends, and research gaps[2]. The review spans diverse methodological approaches, including qualitative analyses, quantitative models, and meta-analyses, and covers multiple subfields within education, reflecting the broad and lasting influence of these seminal works. In addition, we analyze collaboration networks among citing authors to identify key actors and the dynamics of scholarly cooperation. These patterns not only map publication trends but also reflect broader theoretical shifts in the VA framework and its increasing relevance to educational policy and accountability debates.

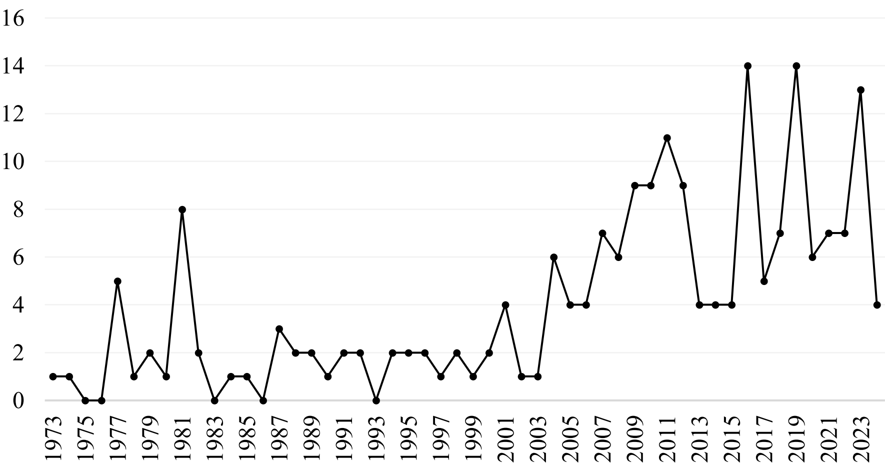
2.2.1 Analysis of studies citing Hanushek (1971)

[2] The studies included and analyzed in this scoping review are marked with an asterisk (*) in the reference list.

As part of the scoping review, a bibliometric analysis of collaboration networks reveals that the most influential authors were predominantly from the United States, Germany, the United Kingdom, the Netherlands, and Italy, countries with greater financial resources, research infrastructure, and access to relevant data. This geographic concentration underscores the need to expand studies in developing countries to achieve a more global understanding of educational effectiveness, particularly given evidence that achievement gaps have narrowed in specific contexts and periods (see e.g., Armor et al., 2018; Ferrão, 2022b; Hanushek et al., 2022). For instance, the Ibero-American region, which includes many developing countries, remains underrepresented in the literature – see, e.g., Larrea et al. (2024), Martínez-Abad et al. (2020), or Santelices et al. (2017a) – despite its diverse educational contexts and policy challenges.

The number of studies citing Hanushek's work has increased significantly in recent decades, especially in the last 20 years, as shown in Figure 2.2. These studies reflect a concentration of academic production mainly in the fields of economics (38%) and education (32%).

Figure 2.2 – Annual scientific output of articles citing Hanushek (1971)



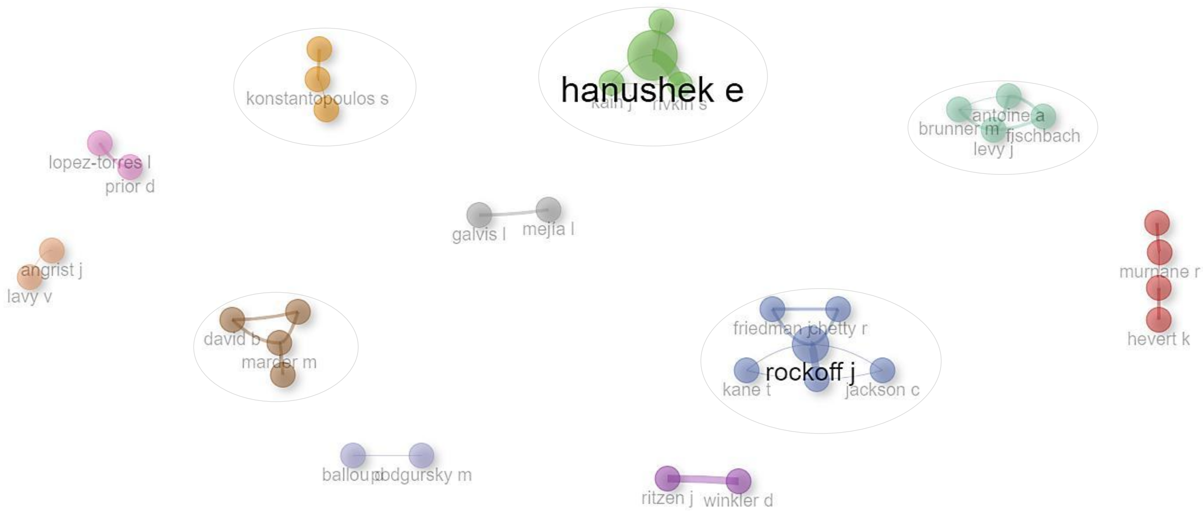
Source: Own elaboration (2026).

2.2.1.1 Collaboration Network Analysis of Authors

The influential clusters identified around Hanushek’s work not only document empirical advances in VA estimation but also illustrate how methodological developments have progressively affected policy-relevant interpretations of teacher and school effectiveness. The collaboration network, shown in Figure 2.3, served as the basis for selecting the articles listed in Table A2 in Appendix A. The focus is on the most relevant groups identified over time, considering the number of citations. The size of the nodes and the number of connections indicate the relative prominence of authors, highlighting five main groups associated with Konstantopoulos, Hanushek, Rockoff, Marder, and Levy, which reflect co-authorship relations.

Studies citing Hanushek's work appear in prestigious journals such as the American Economic Review and Econometrica, as well as in educational publications such as Educational Assessment, Evaluation and Accountability (see Table A1 in Appendix A). The diversity of journals underlines the cross-disciplinary relevance of the study. Analysis of the clusters reveals central themes: Hanushek focuses on educational policy and economic performance; Rockoff examines school practice and management; Marder evaluates educational programs with a focus on STEM[3] teachers; Konstantopoulos studies educational inputs and teacher effectiveness; and Levy explores VA models in education with a focus on accuracy and methodology.

Figure 2.3 – Collaboration network among authors citing Hanushek (1971), 1971-2024



Source: Own elaboration (2026).

2.2.1.2 Analysis of influential cluster studies

Hanushek (1971) set the stage for understanding the relationship between school inputs and student performance by introducing the concept of VA to quantify the contributions of teachers and schools. His findings on resource effects prompted further investigation, most notably by Hedges et al. (1994), whose meta-analysis highlighted the significant role of school investment in student achievement. Subsequent studies, including Nye et al. (2004) and Rivkin et al. (2005), applied multilevel and fixed-effects models to demonstrate that teacher experience

[3] Science, Technology, Engineering, and Mathematics.

and school quality substantially influence student outcomes. More recent research by Jackson et al. (2014), Chetty et al. (2014a, 2014b) and Levy et al. (2019) reflected methodological advances in VA estimation, incorporating strategies to account for unobserved student characteristics, refine model specifications, and improve transparency. Together, these studies illustrate the maturation of VA research and its increasing analytical sophistication. Table A2 in Appendix A summarizes the main characteristics of studies within this cluster.

Building on these methodological advances, some of the studies identified in this scoping review have examined the substantive implications of VA models for educational outcomes and public policy. Research by Chetty et al. (2014a, 2014b), Levy et al. (2020, 2023), and Emslander et al. (2022) demonstrated that VA estimates are associated with positive long-term outcomes, including higher college enrolment rates and future earnings. These patterns are consistent with broader methodological discussions in the literature, which emphasize the relevance of incorporating contextual covariates and advanced statistical techniques to enhance the reliability of VA measures (Amrein-Beardsley et al., 2013; Sirin, 2005). Within the reviewed corpus, studies on teacher effectiveness consistently highlight the persistence of teacher effects and show that teacher turnover undermines both student achievement and the stability of VA scores, a concern also widely discussed in the methodological literature (Braun, 2013; Koedel et al., 2015). At a broader level, Hanushek & Woessmann (2011) extended the VA framework to the macroeconomic sphere, suggesting that improvements in educational effectiveness contribute to long-term economic growth in OECD countries.

Against this backdrop of evidence on the substantive and long-term implications of VA estimates, a subset of studies identified in the scoping review examines the effects of VA based accountability policies on teaching quality and teacher distribution. Analyses of large-scale reforms, such as the *No Child Left Behind Act* (Hanushek & Rivkin, 2010), indicate that while accountability policies may yield gains in test scores, particularly among marginalized students, they are also associated with consequences such as increased teacher turnover. These tensions, which emerge from the reviewed studies, echo broader concerns in the methodological literature regarding the use of VA estimates for high-stakes decision-making, given their potential instability and sensitivity to model specification (Amrein-Beardsley, 2023; Newton et al., 2010).

Similarly, evidence on teacher turnover indicates that isolated policy measures, such as merit pay, are insufficient to address structural constraints in teacher recruitment and allocation, highlighting the role of institutional design and managerial capacity in improving educational effectiveness (Bertoni et al., 2024; Marder, 2012; Muñoz & Prem, 2024). Taken together, these findings suggest that even valid VA models may generate unintended outcomes when used to inform high-stakes accountability decisions in weak or fragmented institutional contexts, emphasizing that educational effectiveness relies on both measurement quality and institutional capacity (Bertoni et al., 2024; Muñoz & Prem, 2024). In contexts characterized by high institutional heterogeneity, such as Brazil, these findings are particularly relevant to ongoing debates on educational quality, as they highlight the risks of adopting accountability models without adequate consideration of local inequalities in school infrastructure.

In studies citing Hanushek (1971), VA is primarily used to evaluate educational policies, centering on the contributions of schools and teachers to student achievement. Influential author clusters have driven both methodological innovation and insights relevant to educational policy, although research remains predominantly concentrated in high-income countries. In practice, VA indicators are increasingly used to guide teacher evaluation systems, resource allocation, and institutional accountability. However, these findings highlight that VA should be applied with caution in high-stakes decisions. In the Brazilian context, where large-scale assessments are increasingly used to inform educational policy and management, these findings highlight the risks of applying VA indicators in high-stakes decisions without adequate consideration of institutional capacity and persistent inequalities across schools and regions.

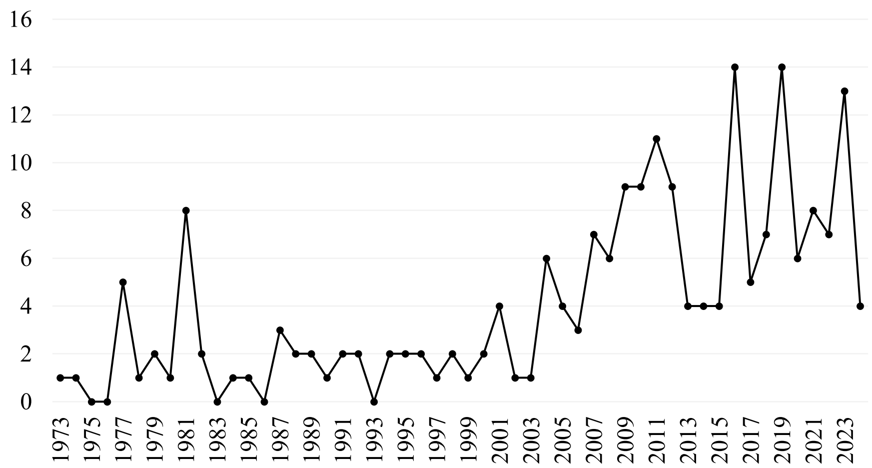
In addition, improving educational outcomes also depends on careful attention to teacher training and retention, particularly given persistent challenges such as high teacher turnover and the unintended consequences that can arise from high-stakes accountability policies. Overall, even methodologically robust VA models can produce unintended effects, underscoring that educational effectiveness depends not only on measurement quality but also on institutional design, teacher selection, and managerial capacity.

2.2.2 Analysis of studies citing Bryk & Weisberg (1976)

Building on early conceptualizations of VA in education, Bryk & Weisberg (1976) introduced methodological innovations that allowed for more precise estimation of educational effects, particularly when randomized experiments were not feasible. A total of 26 studies published between 1976 and 2024 were identified. Network analysis shows that the most influential authors are primarily based in the United States, reflecting the geographic concentration of VA research in education. Although Bryk & Weisberg's (1976) work has had a smaller citation footprint compared to Hanushek (1971), the continued presence of references over time indicates that their methodological contributions remain relevant in specific research contexts.

Figure 2.4 presents the annual scientific output of articles citing Bryk and Weisberg (1976) over the analyzed period. These publications reflect a multidisciplinary approach with a focus on education, psychology and applied statistics, integrating quantitative methods and educational interventions. Publication in high-impact journals, such as Psychological Methods and Psychological Bulletin, suggests that this study has made significant contributions to specific areas of educational psychology and research methodology (see Table A1 in appendix A).

Figure 2.4 – Annual scientific output of articles citing Bryk & Weisberg (1976)

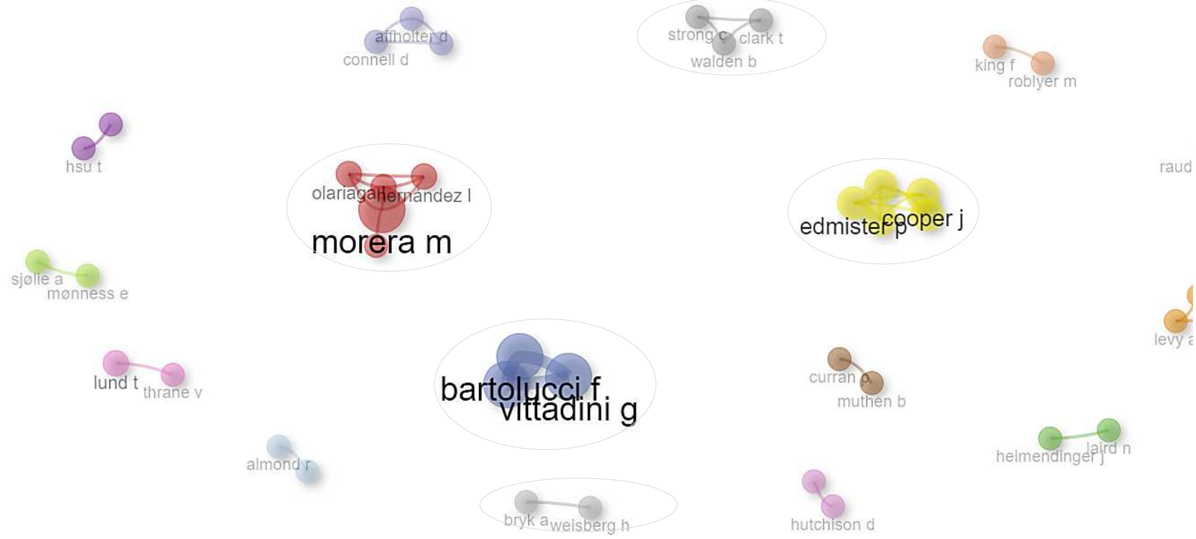


Source: Own elaboration (2026).

2.2.2.1 Analysis of the authors' collaboration network and influential clusters

The analysis of the authors' collaboration network citing Bryk & Weisberg (1976), presented in Figure 2.5, highlights a limited number of influential clusters over time. Node size and network connections indicate author influence, revealing three main clusters led by Bartolucci, Cooper, and Morera, as well as two additional clusters associated with Muthen and Bryk based on citation counts. Table A3 in Appendix A summarizes the main characteristics of the studies within each identified cluster. Except for the clusters led by Bartolucci and Morera, most clusters comprise a single article, reflecting the fragmented diffusion of citations within this body of literature. Accordingly, the in-depth analysis focused on the seven studies that formed substantial collaborative clusters, to capture consolidated methodological lineages rather than isolated or dispersed citations.

Figure 2.5 – Collaboration network of authors citing Bryk & Weisberg (1976), 1976-2024



Source: Own elaboration (2026).

Despite their limited size, the identified clusters are substantively informative regarding the evolution of statistical modeling in educational evaluation. The work by Bryk & Weisberg (1977) critically assessed traditional cross-sectional approaches and proposed linear growth models based on pre- and post-test data to more accurately capture learning trajectories over time, a perspective later developed and operationalized in the broader literature on VA and growth curve literature (Palardy, 2008; Primi et al., 2010). This approach also contributed to the development of subsequent longitudinal and multilevel modeling frameworks, which extend VA analyses to more complex educational data structures (Goldstein, 2003; Hox et al., 2017; Singer, 1998).

Building on this framework, researchers such as Muthen and Curran (1997), Morera et al. (2009), and Bartolucci et al. (2011, 2023) advanced the analysis of educational data through the development of latent variable models, longitudinal designs, and causal extensions of difference-in-differences approaches that explicitly account for school- and classroom-level effects. These contributions expanded the analytical capacity of evaluation models, allowing for more precise estimation of educational intervention effects and of the dynamic structure of educational processes. Taken together, these contributions provide a stronger methodological foundation for assessing educational effectiveness across heterogeneous and complex contexts.

In summary, studies citing Bryk and Weisberg (1976) commonly employ VA approaches to evaluate educational interventions. Their methodological innovations laid the groundwork for longitudinal, multilevel, and latent variable approaches in educational assessment, allowing the effects of programs to be estimated more precisely and capturing the dynamic structure of learning processes. These studies show that VA can provide nuanced insights into the impact of interventions at the level of individual students and classrooms, supporting the design of targeted programs and accountability policies. The advances inform debates on VA modeling and educational effectiveness, particularly in contexts where randomized experiments are not feasible and evaluations rely heavily on administrative and large-scale assessment data, as in countries such as Brazil. However, the uneven adoption of these methods highlights that methodological progress must be supported by adequate research infrastructure, funding, and data access to produce robust evidence capable of informing public policy.

2.2.3 Thematic analysis of studies

While Sections 2.3.1 and 2.3.2 examined the evolution of VA research through distinct citation lineages, the following thematic synthesis integrates these trajectories to identify broader patterns and implications for educational evaluation. Rather than merely indicating keyword frequency, thematic maps[4] are interpreted here as signals of how the VA field has expanded.

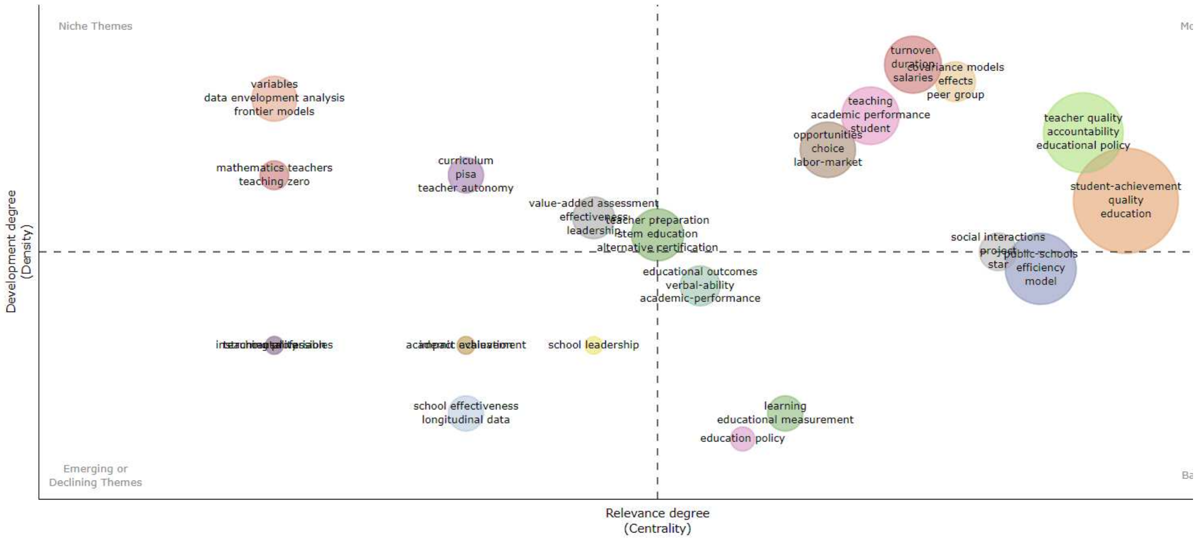
This analysis highlights themes and issues likely to be central to educational research in the coming years, providing guidance for future studies and academic approaches. Figure 2.6 illustrates the thematic clusters on a graph with two dimensions: centrality (X-axis), which refers to the importance of a given theme, and density (Y-axis), which refers to the development of the theme. Clusters in Quadrant 1 (Motor Themes) are central and well-developed. Quadrant 2 (Niche Themes) contains specialized themes with high density but lower influence. Quadrant 3 (Emerging or Declining Themes) features areas of low density and centrality, suggesting either reduced development or emerging topics. Finally, Quadrant 4 (Foundational Themes) shows relevant themes that are well researched and integrated as foundational knowledge in the field.

Figure 2.6 highlights the central themes of the 207 studies cited by Hanushek (1971). The dominant themes reflect an interest in how school factors and teacher effectiveness affect student achievement and underscore the importance of evaluating resource efficiency in public education. Analysis of quadrants 1 and 3 not only identifies dominant and emerging themes but also illustrates the evolving focus of the field. The clusters in quadrant 1 (driver themes), such as student achievement, teacher quality, covariance models, opportunity, and attrition

[4] Thematic maps were created using all studies from each database without applying any filters, as the aim is to identify the general directions of each group of studies.

(Hanushek, 2011; Hanushek et al., 2016a; Levy et al., 2023; Rivkin et al., 2005), are well-developed, have a significant impact on education policy and practice, and influence research on educational evaluation. In contrast, emerging themes in quadrant 3 — such as “school leadership”, “school effectiveness”, and “longitudinal data” (Kuzmanic et al., 2024; Levy et al., 2020; Valle & Lillejord, 2023) — reflect a shift in the literature toward understanding systemic school processes and the temporal dynamics of learning. This movement is crucial for analyses of equity and differential effectiveness, as longitudinal data enables more accurate measurement of VA and its variation across student populations.

Figure 2.6 – Thematic map of studies citing Hanushek (1971), 1971-2024



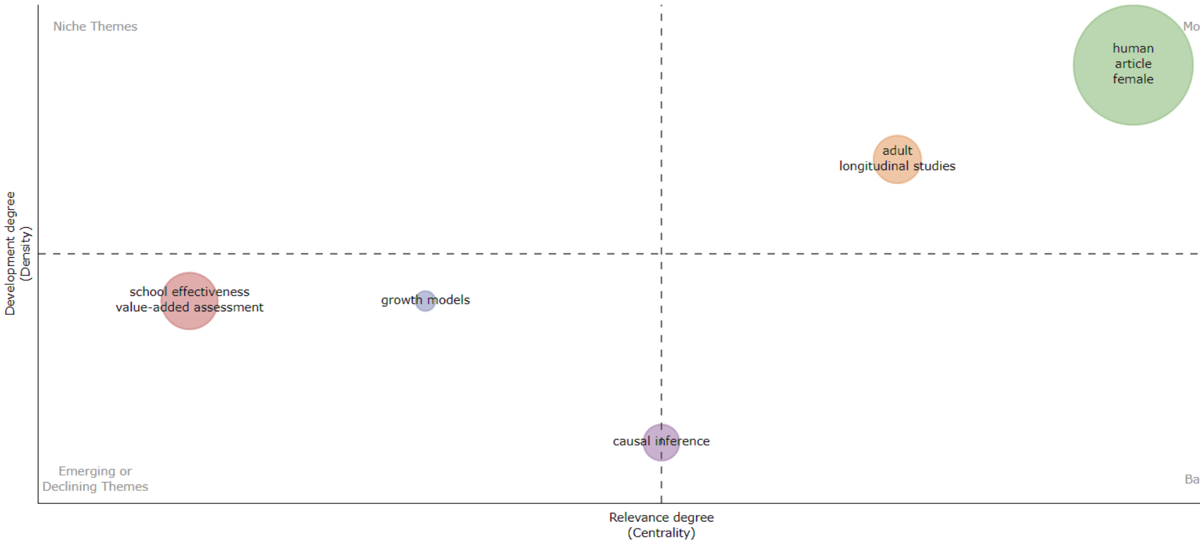
Source: Own elaboration (2026).

This shift toward educational systemic processes and temporal dynamics, identified in Quadrant 3, aligns with contemporary evidence highlighting the need for more robust models to capture educational complexity. The literature supports the view that the use of longitudinal data is essential for ensuring the stability and consistency of VA estimates, addressing the limitations of cross-sectional analyses (Ferrão & Couto, 2013; Morera et al., 2009). Regarding school leadership, recent studies, such as Muñoz & Prem (2024), highlight the increasing importance of this topic by demonstrating how managerial effectiveness directly impacts student outcomes. Taking it together, these findings suggest that future research in educational evaluation will increasingly focus on school-level factors, managerial effectiveness, and the longitudinal analysis of learning trajectories, offering new avenues for understanding and improving VA outcomes in educational evaluation.

The thematic analysis of studies citing Bryk & Weisberg (1976), shown in Figure 2.7, reveals both established and emerging areas of interest in educational research. Quadrant 1 encompasses well-developed themes, emphasizing human factors and longitudinal analyses

(Bartolucci et al., 2011, 2023). In addition, the topic of causal inference (Raudenbush & Schwartz, 2020) has emerged between Quadrants 3 and 4, occupying an intermediate stage of development, yet gaining prominence as indicated by its increasing frequency as a keyword in recent years. Other themes in Quadrant 3, including "school effectiveness" and "value-added assessment", represent potential avenues for future research. While these topics have been extensively discussed in educational literature (e.g., Sammons, 2007; Santelices et al., 2017; Soares et al., 2017; Strand, 2010, 2016), they continue to evolve, incorporating new approaches and perspectives that warrant further exploration and refinement.

Figure 2.7 – Thematic map of studies citing Bryk & Weisberg (1976), 1976-2024



Source: Own elaboration (2026).

The thematic analysis of studies citing Hanushek (1971) and Bryk & Weisberg (1976) reveals a field predominantly focused on traditional factors such as teacher effectiveness, while also showing increasing attention to emerging areas, including school leadership, longitudinal data, and causal inference. This shift toward a more comprehensive and methodologically sophisticated understanding of school effectiveness is particularly relevant in contexts like Brazil, where longitudinal approaches are essential for producing stable and equitable VA measures (Brooke et al., 2014; Ferrão; Couto, 2013; Soares et al., 2017). At the same time, the literature highlights persistent limitations related to the stability, interpretability, and high-stakes application of VA models, emphasizing the need to complement quantitative analyses with qualitative evidence and engagement with school actors, such as teachers, students and families. Together, these trends underscore the potential of VA as a contextualized tool for educational evaluation, while informing educational policies (Amrein-Beardsley et al., 2013; Braun, 2013).

2.4 CONCLUSION

In response to the research question, the findings indicate that VA research in education has matured substantially. The field has moved beyond the simple measurement of test scores toward the analysis of more complex learning processes, showing that educational effectiveness depends not only on methodological sophistication but also on the institutional context of schools.

This scoping review examined the development of the VA concept in educational research by analyzing studies citing Hanushek (1971) and Bryk and Weisberg (1976) published between 1971 and 2024. The analysis combined bibliometric techniques, collaboration network mapping, and thematic synthesis, highlighting the evolution of VA models from simple estimates of teacher and school effects to more complex, longitudinal, and increasingly causally informed approaches. The review underscores that methodological sophistication alone is insufficient: VA assessments must be complemented by qualitative evidence and the active engagement of teachers, students, and families, reinforcing that these models are intended to support, rather than replace, contextualized professional judgment.

Important limitations remain in the literature, particularly the limited number of studies conducted in developing countries and the need to expand research into thematic areas such as school leadership, academic performance, longitudinal data, school effectiveness, and causal inference. A more systematic examination of contextual variables shaping educational effectiveness is also required to better adapt interventions to diverse institutional settings. In addition, the literature highlights that education policies prioritizing the training and retention of qualified teachers are central to improving student outcomes. As a scoping review, this study presents inherent limitations, including an emphasis on breadth rather than depth, the absence of formal methodological quality appraisal, reliance on seminal citations, and the execution of screening and data extraction by a single author, which calls for caution in generalizing the findings.

Based on the patterns identified, the findings contribute to educational policy and practice by emphasizing the role of school leadership in shaping school culture and improving student outcomes, as well as the importance of longitudinal data for monitoring academic progress and supporting evidence-based policy adjustments. The results also highlight the potential of VA assessments to guide practice, enhance educational effectiveness, and promote equity, including more informed resource allocation and the implementation of targeted programs for underserved schools. In contexts such as Brazil, where persistent social inequalities and institutional constraints affect educational outcomes, these insights are particularly relevant. At the same time, generating robust evidence capable of informing public policy requires adequate research infrastructure, sustained funding, and reliable access to data, alongside closer collaboration between researchers and policymakers.

In light of these findings, future research should expand empirical VA studies in developing countries, investigate the relationship between school leadership and VA, deepen the use of longitudinal designs, explore causal inference strategies, and incorporate a broader set of contextual variables into VA models. The conduction of Systematic Literature Reviews (SLR) is also essential to consolidate evidence and strengthen the methodological and policy relevance of VA research. Overall, this study seeks to stimulate further research and debate aimed at fostering more equitable and effective education systems.

As the opening essay of this thesis, this study establishes the conceptual and methodological foundations for a critical examination of VA models in educational assessment, clarifying how schools' contributions to student learning have been theorized, measured, and debated in the literature. By mapping the evolution of VA research and identifying gaps, particularly regarding institutional context, school leadership, longitudinal data, and causal inference, this essay motivates the broader research agenda of the thesis and lays the groundwork for the subsequent analyses.

3. ESSAY 2: VALUE-ADDED MODELS IN IBERO-AMERICAN EDUCATION: A SYSTEMATIC REVIEW

ABSTRACT

The implementation of Value-Added (VA) models is becoming a central instrument in educational evaluation and public policy guidance. Studying these models in diverse contexts provides valuable insights into their application and impact. This systematic review examines the use of VA models in educational systems across Ibero-American countries. Following the PRISMA protocol, empirical studies estimating VA are identified and analyzed across four dimensions: objectives, methodology, variables used, and main findings. Most studies focus on assessing school and teacher effectiveness, and rely on two-level multilevel models with standardized test scores as outcomes, and consistently show that family, school, and socioeconomic factors influence student performance, while also revealing a scarcity of emotional or alternative educational indicators and limited regional collaboration, as well as the need to incorporate broader performance measures such as grade repetition and dropout. Overall, the review indicates that VA models are useful for evaluating educational effectiveness in Ibero-America, but their potential remains limited by methodological heterogeneity and a narrow geographic scope, highlighting the need for advances that ensure more robust and comparable evidence across the region.

Keywords: value-added model; multilevel models; school effectiveness; student performance; systematic literature review.

3.1 INTRODUCTION

As evidenced in the scoping review presented in the previous essay, the theoretical foundations of Value-Added (VA) models were solidified largely in developed countries. VA models are a tool in educational evaluation employed for analyzing the educational effectiveness (Chetty et al., 2014; Ferrão & Couto, 2013; Paufler & Amrein-Beardsley, 2014). These models allow for the estimation of the VA to students' human capital by accounting for the factors and resources that influence the learning process over a given period and how learning accumulates over time. This enables the analysis of the effects of schools and teachers on students' academic performance (Chetty et al., 2014; Ferrão & Couto, 2013; Koedel et al., 2015; Paufler & Amrein-Beardsley, 2014), considering factors ranging from family resources to school-related aspects, such as infrastructure (Blaskó et al., 2022; Burger, 2019; Guilherme et al., 2024; Jensen et al., 2018; Ma et al., 2018; Nurse & Melhuish, 2021; Schneeweis, 2011). Furthermore, these models offer valuable insights for the design of more effective education policies aimed at enhancing quality and reducing educational inequality. They also support the evaluation of educational effectiveness, understood as integrating teacher and school effectiveness, and guide the allocation of resources based on students' specific needs (Amrein-Beardsley et al., 2013; Braun, 2013; Hanushek, 2019).

Political factors significantly influence the development of educational assessment models based on standardized testing (Libâneo, 2016; Miranda & Santos, 2012). In response to demands for greater accountability and transparency, educational reforms have been introduced that incorporate VA metrics to investigate the impact of schools and teachers on student performance. Initiatives such as the Tennessee Value Added Assessment System (TVAAS), the No Child Left Behind Act (NCLB) in the United States, and the Contextual Value Added (CVA) in the United Kingdom reflect different approaches to incorporating VA models into educational evaluation and accountability systems. These initiatives draw on student performance data and contextual factors to inform policy decisions, support school improvement efforts, identify effective pedagogical practices, and optimize resource allocation (Amrein-Beardsley et al., 2013; Braun, 2013; Hanushek, 2019; Levy et al., 2019). By controlling contextual variables, these models offer a more accurate assessment of the impact of schools and teachers on student performance (Rubin et al., 2004; Sirin, 2005; White, 1982).

Over the course of several decades, novel methodological approaches, and academic debates have refined these models, thereby consolidating their role in educational assessment (Koedel et al., 2015; Thomas et al., 2007). Although various approaches exist, linear regression models and multilevel models are the most discussed in methodological literature. A few systematic reviews have addressed the application, usefulness, validity, and limitations of VA models in educational policies and teacher evaluations (Amrein-Beardsley et al., 2023), as well as technical modeling aspects, such as the methodological rigor required. This methodological rigor is crucial to ensure the validity of estimates and the accuracy of applying these approaches (Levy et al., 2019). In specific contexts, such as those of the Iberian Peninsula, researchers have also advanced the field by exploring different specifications of VA models, often with the aim of strengthening educational systems, promoting school autonomy, and enhancing the role of educational assessment within the public policy cycle (Ferrão, 2025).

The implementation of these approaches has been met with a degree of controversy. Braun (2013) underscore issues that can compromise the precision of the models, including school choice bias, measurement errors in assessments, and missing data. In this regard, Kersting et al. (2013) emphasizes the importance of two crucial aspects to ensure the reliability of results: consistency, which ensures that different models produce similar results, and stability, which refers to the consistent ranking of schools over time (Ferrão, 2014; Soares et al., 2017). These aspects enable VA scores to be employed in decision-making processes concerning resource allocation and professional development, fostering advancements in the educational system. However, studies by Baker et al. (2010), Paufler & Amrein-Beardsley (2014) and Rothstein (2009, 2010) suggest that these estimates can be biased and unstable, rendering them unsuitable for high-risk educational management decisions (Newton et al., 2010). Given the potential implications for educational decisions, VA estimates should be interpreted with caution and empirically validated for accuracy and stability.

Building on the conceptual foundations of the previous essay, which showed that VA models are well-established in developed countries, such as the United States and the United Kingdom, this study focuses on the Ibero-American region, where evidence remains fragmented. This region combines developed and developing countries, allowing the analysis of VA models across diverse educational and socioeconomic contexts. While countries like Chile, Brazil, Portugal, and Spain have developed significant research on the topic in the region – see, e.g., Ferrão (2022b), Larrea et al. (2024), Martínez-Abad et al. (2020), Murillo &

Martínez-Garrido (2019) or Santelices et al. (2017a) –, there is still a need to expand investigations to assess the effectiveness and adaptation of these models to distinct educational and socioeconomic contexts. The primary objective of this Systematic Literature Review (SLR) is to synthesize and critically analyze empirical evidence on the application of VA models in educational research across Ibero-American contexts, identifying methodological tendencies, commonly used variables, and research gaps in the existing literature. Accordingly, the review addresses the following research questions:

RQ1. What research objectives have guided studies employing VA models in education?

RQ2. Which methodological designs and analytical approaches predominate in these studies?

RQ3. Which variables and indicators are most frequently used in the estimation of VA models?

RQ4. What are the main findings, gaps, and limitations reported in these studies?

To answer the proposed research questions, this SLR followed the PRISMA protocol to ensure transparency and replicability throughout all stages. Searches were conducted in ERIC, Scopus, and Web of Science using multilingual strings (Portuguese, English, and Spanish) to identify studies on VA models within Ibero-American education systems. These databases were selected because they index the most relevant journals in education and because they provide extensive coverage of peer-reviewed research on VA models. After applying the inclusion and exclusion criteria, the data extraction from the final sample was organized into three main dimensions: general study characteristics, methodological aspects, key findings and reported limitations. This framework enables a systematic examination of VA research by outlining its objectives and methodological approaches (RQ1-RQ2), identifying the most frequently used variables and indicators (RQ3), and synthesizing the main findings and limitations (RQ4). By identifying methodological patterns and gaps in the literature, this review also lays the groundwork for the subsequent empirical study of VA models in Brazilian schools (Essay 3) in contexts of high inequality.

In addition to this introduction, the following section describes the methodology adopted in this study. The third section of this study presents the results. The fourth section integrates discussion and conclusions, addressing the findings and their implications, and summarizing key insights and recommendations.

3.2 METHODOLOGY

The SLR is a rigorous and structured method designed to identify, assess, and synthesize available evidence to answer a research question and expand knowledge on a given topic (Cumpston et al., 2019). This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines, which establish systematic stages for the identification, screening, eligibility assessment, and inclusion of studies, ensuring transparency and reproducibility throughout the process (Page et al., 2021).

With the research questions and objectives clearly defined, the systematic review protocol encompassed the selection of databases, the formulation of the search string, and the adoption of inclusion and exclusion criteria for the final corpus of studies. The data collection phase was executed in November 2025, and the resultant data was downloaded in BibTeX format and subsequently imported into Parsifal[5], an online tool designed to provide methodological support for conducting systematic reviews. This tool aims to ensure rigor, organization, and transparency throughout all stages of the review process.

Search strategy

This SLR utilizes the Education Resources Information Center (ERIC), Scopus, and Web of Science databases, which complement each other, ensuring a representative selection of studies on VA model applications in education and economics within the Ibero-American context[6]. Scopus and Web of Science databases were chosen due to their international prominence and multidisciplinary coverage, complemented by the ERIC database, which was specifically included for its focus on educational research.

Following the selection of the databases, search strings were applied to each database in the fields of titles, abstracts, and keywords, including terms such as “education”, “school”, “teacher”, “student”, “value added model” and “Ibero-America” (Table B1, in the Appendix B). To ensure broad coverage of relevant publications, the search was conducted in Portuguese, English, and Spanish. The search strings differ in their consideration of geographic location:

---

[5] Accessible at: https://parsif.al/

[6] According to the Organization of Ibero-American States (OEI), the member countries are as follows: Andorra, Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Spain, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Portugal, the Dominican Republic, Uruguay, and Venezuela.

one restricts the search to the location mentioned in the title, abstract, and keyword fields, while the other uses the location information recorded in the database, covering a larger number of relevant publications.

Eligibility criteria

To assess the use of VA models in Ibero-American education, studies published in Portuguese, Spanish, and English were included. Portuguese and Spanish capture the primary scientific production of the region, while English was included as the dominant language of international scientific communication. Conference proceedings, gray literature, and non-peer-reviewed studies were deliberately excluded (see Table 3.1), as were studies conducted outside the Ibero-American region or focused on higher education. This was done to ensure that the analysis centered on learning outcomes at the school level. A preliminary search indicated that no prior systematic reviews specifically examine the application of VA models in Ibero-American education. Moreover, systematic reviews and meta-analyses were disregarded, given that the objective of this review is to examine original empirical evidence. This decision also aimed to avoid double counting results and ensure comparability across studies. In the interest of maintaining the rigor and integrity of the research process, studies that did not apply VA models or that did not address educational contexts and learning outcomes were excluded. This approach ensured the relevance, methodological consistency, and overall quality of the research incorporated into this review.

Table 3.1 - Inclusion and exclusion criteria used to define the systematic review sample

Inclusion criteria	**Exclusion criteria**
Studies employing VA models	Do not apply VA models, systematic reviews, and meta-analyses
Research focused on educational contexts or learning outcomes	Studies not focused on the educational contexts or learning outcomes
Research conducted in Ibero-American countries	Studies conducted outside the Ibero-American region
Studies analyzing VA in primary and secondary education	Studies analyzing VA in higher education
Peer-reviewed studies available in full text	Studies not peer-reviewed or unavailable in full text

Source: Own elaboration (2026).

Data extraction

The data collection from the reviewed studies encompasses three primary domains: general information about the studies (journal, year, authors, analyzed location, etc.); methodological aspects (statistical model, variables, etc.); and key results and limitations. Analyzing this information allows us to address the research questions, map the literature on educational assessment and VA model development in the Ibero-American region, and identify potential research gaps. During the data extraction process, each article was evaluated for methodological adequacy, transparency in the specification of variables, and the validity and consistency of findings. To ensure methodological rigor and uniformity in this assessment, quality appraisal was guided by established quality criteria, focusing on whether VA models were appropriately specified, whether dependent, independent, and control variables were theoretically justified and clearly described, and whether the interpretation of results aligned with the analytical strategy and acknowledged limitations.

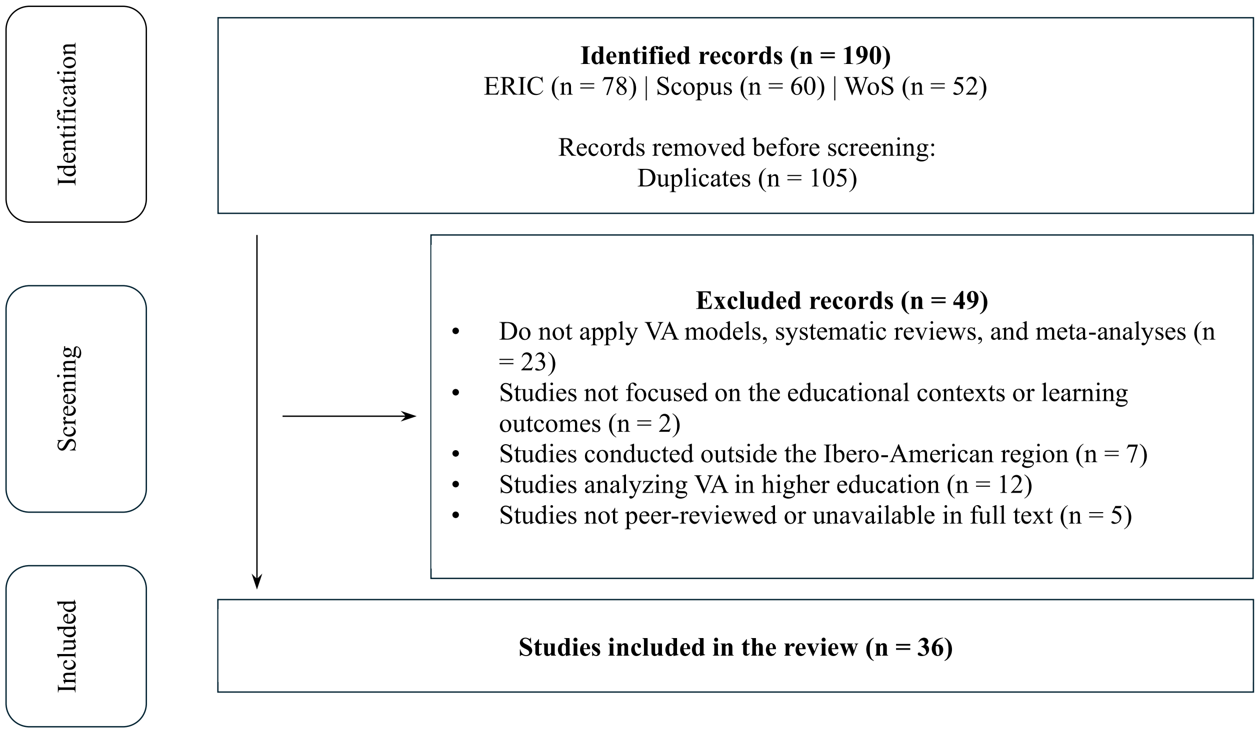
3.3 RESULTS

To analyze the evolution of VA models in Ibero-American countries, an initial global search was conducted for studies applying these models in educational. The search yielded 1,589 articles, reflecting the high academic interest in the topic. However, when the search was restricted to Ibero-American countries, the number dropped to 190 studies, which was further reduced to 36 after screening, highlighting the need for more evidence for this specific context. Most of the selected studies were published in English, with fewer in Spanish and Portuguese. The detailed selection process is illustrated in Figure 3.1.

The studies included in the review were conducted in the following countries: Chile (N = 15), Brazil (N = 8), Portugal (N = 4), Spain (N = 4), Peru (N = 3), and Mexico (N = 2)[7]. One of the studies analyzed both Brazil and Portugal, resulting in a total of 36 studies in the analyzed set. The reviewed studies did not include any research from the following countries: Andorra, Argentina, Bolivia, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Dominican Republic, Uruguay, and Venezuela.

[7] The studies included and analyzed in this systematic review are marked with an asterisk (**) in the reference list.

Figure 3.1 - Flow diagram of the systematic review according to the PRISMA model



Source: adapted from Page et al. (2021, p.6).

3.3.1 Objectives of the reviewed studies

This topic addresses RQ1, focusing on the research objectives that have guided VA studies in Ibero-American education systems (a summary of the reviewed studies is provided in Table B2, in the Appendix B). Within this scope, research on teachers' VA has examined the association between higher education levels or greater years of experience and greater contributions to student learning progress (Filho, 2019; Santelices et al., 2015, 2017). Other studies use teachers' VA scores as a benchmark to compare with various evaluation tools, either to assess how well hiring instruments predict teacher effectiveness (Bertoni et al., 2024), or to examine the relationship between VA scores and the outcomes of teacher performance evaluations (Taut et al., 2016).

In parallel, models have begun exploring the relationship between school characteristics and student outcomes (Eigbiremolen et al., 2020; Ferrão, 2014; Ferrão & Couto, 2014; Ferrão & Goldstein, 2009; Ortega et al., 2018; Page et al., 2017; Troncoso et al., 2016; Troncoso, 2019; Vivanco, 2013), controlling by students' socioeconomic variables (Castro-Morera et al., 2015; Ferrão, 2009; Singh, 2020; Torres, 2018). In this context, an important issue that has gained traction in the academic literature is the role of schools in not only enhancing academic outcomes but also in addressing gender disparities (see, e.g., Muñoz-Chereau, 2019) and mitigating broader social inequalities (see, e.g., Ferrão, 2022a). To this end, increasingly sophisticated VA models have been proposed and applied across different educational and social contexts. These include frontier models (Thieme et al., 2016), multilevel models (Brooke et al., 2014; Soares et al., 2017), and dynamic specifications that incorporate temporal dependency (Page et al., 2024). Recent studies have further expanded the use of VA models to analyze managerial effectiveness in public education systems (Muñoz & Prem, 2024), gender-specific teacher effectiveness (Barrios-Fernández & Riudavets-Barcons, 2024), and even family-related influences on child development (Berthelon et al., 2020), through fixed-effects specifications. Altogether, these advances have facilitated a more comprehensive understanding of the persistence and determinants of school effectiveness over time.

Several studies have examined the methodological applications of VA models. For instance, Ferrão & Couto (2013) investigates the choice of statistical specification, particularly within multilevel models, and shows the impact both the magnitude and the stability of school effects. Similarly, López Martin et al. (2014) find that linear and nonlinear approaches may lead to divergent classifications of schools or teachers, underscoring the sensitivity of VA estimates to model structure. Other studies explore how the configuration of the data affects VA estimation. Morera et al. (2009) report that longitudinal designs produce more robust estimates by explicitly modeling students' prior achievement trajectories, while Hernández & Olariaga (2009) emphasize the importance of comparable longitudinal assessments to avoid bias in growth estimates. Muñoz-Chereau & Thomas (2016) further expands this discussion by showing that estimating effects across multiple levels (municipal, school, and classroom) can substantially alter conclusions about where educational value is generated. These findings highlight how strongly VA results depend on modeling decisions, reinforcing the need for careful specification and transparent reporting when interpreting school or teacher effects.

A subset of the reviewed studies examines how external, familial, and health-related conditions affect students' academic trajectories. In northeastern Brazil, Gomes-Neto et al. (1997) investigate the effects of visual acuity, nutrition, and general well-being on the achievement of rural children. Other studies focus on early years and home practices: Bartholo et al. (2019) and Berthelon et al. (2020) analyze how daycare experiences and harsh parenting influence cognitive and socio-emotional development, while Jerrim et al. (2019) estimate the impact of homework time using a twin fixed-effects design. Bartholo et al. (2023) further document learning losses associated with COVID-19 school closures. Together, these studies highlight the wide range of non-school factors that influence learning outcomes and underscore the need to interpret academic performance within its broader social and familial context.

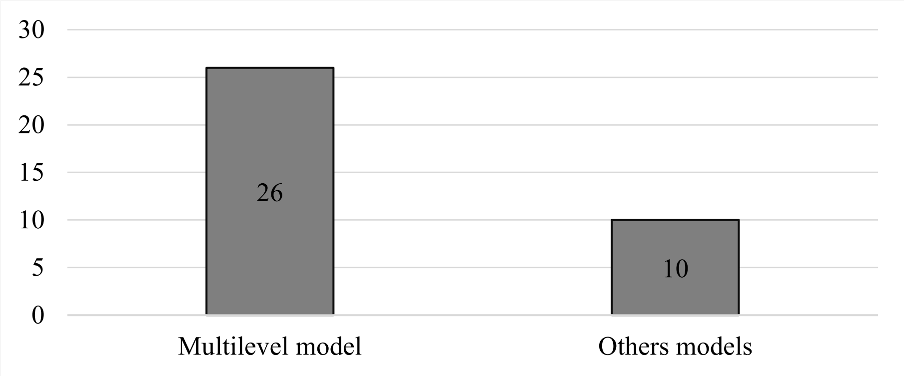
3.3.2 Statistical models

In recent decades, VA modeling has been refined to yield more rigorous and contextually grounded analyses in education. During this period, multilevel models and other methodologies have garnered particular attention, as classified in Figure 3.2. The 'other models' category includes ten studies that chose alternative models to multilevel methods – such as quantile regression techniques, fixed effects models, and more dynamic approaches – based on their specific analytical needs and the absence of a requirement to directly model the hierarchical structure of the data. This discussion contributes to addressing RQ2.

Bartholo et al. (2019) and Gomes-Neto et al. (1997) used multivariate regression models to explore linear associations and examine multiple factors simultaneously. Hernández & Olariaga (2009) advanced the use of regression models combined with Item Response Theory (IRT)[8], allowing for the capture of nuanced student performance and the inclusion of more reliable measures in VA analysis. Models such as dynamic OLS (Eigbiremolen et al., 2020), fixed-effects (Barrios-Fernández & Riudavets-Barcons, 2024; Berthelon et al., 2020; Jerrim et al., 2019; Muñoz & Prem, 2024), and quantile regressions (Page et al., 2017) have been employed, respectively, to control for individual variations over time, isolate specific effects, and deal with latent variables and unobserved heterogeneity. Thieme et al. (2016) uses a frontier model that allows for the evaluation of school efficiency by comparing its performance with a reference frontier, thereby reinforcing the importance of contextual variables in explaining performance differences.

[8] The IRT is a statistical method that evaluates students' latent abilities on tests, as opposed to the conventional approach of deriving a score based on the total points earned on the test.

Figure 3.2 – Statistical model types

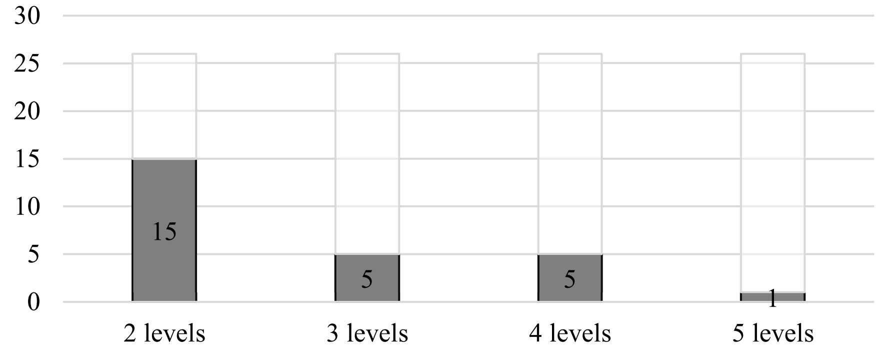


Source: Own elaboration (2026).

Multilevel models are statistical models that have gained significant attention in the academic community for their ability to manage dependencies across various levels of analysis of precise form. Compared to multiple regression or fixed effects models multilevel models have a distinct advantage in capturing contextual effects and intricate interactions. In this review of literature 76% of the studies analyzed estimate VA using multilevel models. Figure 3.3 presents the studies of 'multilevel model' category categorized by the levels included in their analyses.

Figure 3.3 shows 15 two-level studies, of which 13 adopt a student-school structure and 2 a student-class/teacher structure. They analyze the role of socioeconomic variables (Ferrão, 2009; Ferrão & Couto, 2014; Ferrão & Goldstein, 2009; Morera et al., 2009), the persistence of school effectiveness (Page et al., 2024), the impact of teacher quality on student learning (Taut et al., 2016), and the effects of teacher training in early elementary grades (Filho, 2019). Some studies also investigate the repercussions of school closures during the pandemic (Bartholo et al., 2023). Methodological issues are likewise addressed, including the consistency and stability of VA estimates (Ferrão, 2014; Ferrão & Couto, 2013) and the choice of statistical specifications, such as linear versus nonlinear models (López-Martin et al., 2014), accelerated growth curve models (Ortega et al., 2018), status models (Soares et al., 2017), and spatial techniques (Vivanco, 2013), which allow for the control of geographic factors and capture local variation in educational outcomes. This predominance of two-level models indicates a primary focus on student-school relationships, while the use of diverse methodological approaches reflects efforts to enhance the precision of school effectiveness estimates.

Figure 3.3 – Frequency of multilevel models by levels



Source: Own elaboration (2026).

The extant studies that employ three-level model (Bertoni et al., 2024; Brooke et al., 2014; Muñoz-Chereau, 2019; Santelices et al., 2017; Torres, 2018), four-level models (Ferrão, 2022a; Muñoz-Chereau & Thomas, 2016; Santelices et al., 2015; Singh, 2020; Troncoso et al., 2016), or five-level models (Troncoso, 2019) demonstrate an effort to capture more complex hierarchical structures within educational data. Additional levels into the model enhance its capacity to address the intricacies and interdependencies inherent in the data, thereby facilitating a more meticulous and precise examination of the numerous factors that influence outcomes. However, these models are not without limitations, including the need for larger datasets and greater computational demands, which can complicate their implementation.

In the reviewed literature, most studies rely on sample data (78%), while a smaller proportion employs population-level data (22%). Although population studies generally provide more precise estimates, the prevalence of sample-based research highlights the practical challenges of obtaining comprehensive data. Many studies adopt longitudinal designs (80%), which, combined with the widespread use of multilevel models, allow for a more nuanced understanding of factors influencing educational outcomes. Within this context, the robustness of estimates is frequently assessed through several strategies, including the stepwise inclusion of contextual and socioeconomic variables to test sensitivity (Santelices et al., 2015; Muñoz-Chereau & Thomas, 2016), analysis of temporal stability across years or educational cycles (Ferrão & Couto, 2014; Ferrão, 2014), and controls for potential endogeneity of teachers and students (Filho, 2019; Eigbiremolen et al., 2020). These methodological checks contribute to more consistent and generalizable estimates, ensuring that the observed effects reflect genuine educational phenomena rather than distortions resulting from model misspecification or sample-related characteristics.

Additionally, the studies highlight common limitations in research on educational VA and student performance. Key challenges include omitted contextual variables, non-random allocation of students and teachers, reliance on self-reported or administrative data, and limited generalizability due to sample restrictions, such as the exclusion of private schools or specific regions (Santelices et al., 2015, 2017; Muñoz-Chereau, 2019; Torres, 2018; Bertoni et al., 2024). Non-multilevel studies additionally face challenges related to linearity assumptions, difficulty capturing nested data structures, and limited control of unobserved heterogeneity (Troncoso, 2019; López-Martin et al., 2014). More generally, several studies emphasize the challenge of fully accounting for external factors, including family support, socioeconomic conditions, or interventions over time, which may influence student performance (Page et al., 2017; Bartholo et al., 2019; Brooke et al., 2014). These limitations underscore the need for careful model specification, the use of richer contextual data, and caution in generalizing results beyond the studied populations, as they may affect model estimates and the robustness of inferences drawn from the findings.

3.3.3 Variables in VA models

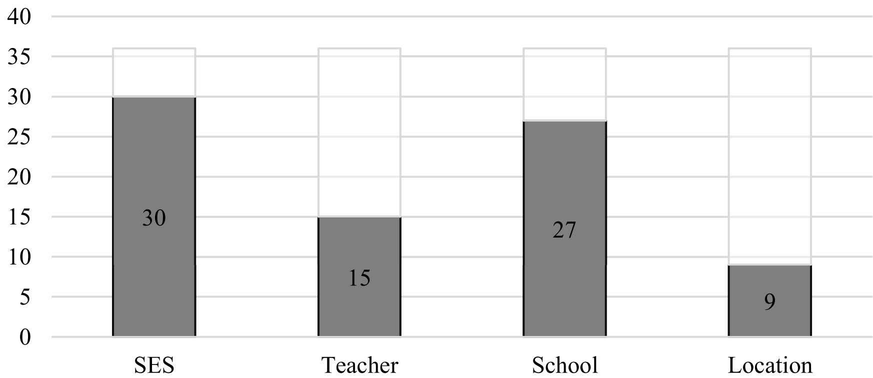
Addressing RQ3, which concerns the variables and indicators most frequently used in the estimation of VA models, this section examines how the selection of explanatory variables for educational performance has become increasingly sophisticated as VA modeling evolved. Across the reviewed studies, student proficiency is the conventionally utilized dependent variable to assess learning outcomes. However, the study by Gomes-Neto et al. (1997) stands out for expanding this approach by incorporating not only proficiency (via multiple regression) but also the probability of school dropout and regular grade progression as dependent variables, which were estimated using probit models. A significant aspect of the reviewed studies is the analysis of student performance based on standardized tests in mathematics, language, or cognitive skills. These tests are widely recognized and utilized, providing a standardized metric for performance evaluation. Notably, evaluations such as the National System of Measurement of Education Quality (SIMCE) in Chile and the National System for Basic Education Assessment (SAEB) in Brazil are widely adopted for this purpose.

Initially, the explanatory variables focused on individual student characteristics, including age, gender, prior performance, and socioeconomic factors. However, beginning in

2009, with the wider adoption of multilevel models, the scope broadened to include additional factors such as motivation, school infrastructure, and teacher training. After 2015, there was an increase in the incorporation of psychosocial and emotional variables, while the most recent period (2020-2024) has highlighted variables related to the impact of the pandemic, and nutritional aspects. This evolution exemplifies the ongoing endeavor to capture the multifaceted influences on educational performance.

Figure 3.4 presents the most used types of explanatory variables in the VA models analyzed in this review. Generally, studies incorporate not only individual student characteristics but also variables from Socioeconomic Status (SES), school, and local contexts within the model structure. Figure 3.3 shows the predominance of two-level models, which accounts for the higher frequency of variables related to students (30 of the 36 studies reviewed) and schools (27 of the 36 studies reviewed), with less emphasis on teacher and locality variables in Figure 3.4. The use of locality variables is less frequent due to the necessity of introducing additional levels into the model, which increases the complexity of the analysis and necessitates larger data sets to ensure precise estimation.

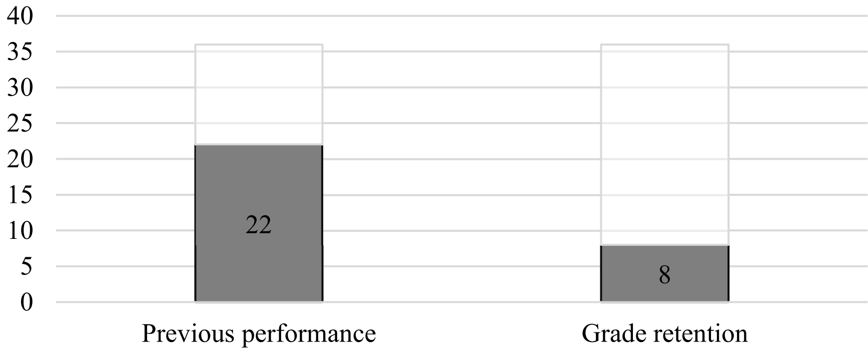
Figure 3.4 – Types of explanatory variables



Source: Own elaboration (2026).

Students' academic history, including prior achievement and grade repetition, is a central dimension in the analysis of educational performance. Prior academic performance is identified as one of the most robust predictors of academic success. Students who demonstrate elevated prior performance are more likely to sustain high performance over time, as evidenced by 22 of the 36 studies reviewed, as illustrated in Figure 3.5. Analyses of factors such as grade retention have likewise identified negative effects on educational performance (Brooke et al., 2014; Castro-Morera et al., 2015; Ferrão, 2009; Ferrão & Couto, 2014; Gomes-Neto et al., 1997; Troncoso, 2019; Troncoso et al., 2016). Furthermore, studies have shown that structural variables, such as the quality of school infrastructure and teacher characteristics, are important for academic development (Bertoni et al., 2024; Castro-Morera et al., 2015; Ferrão, 2022a; Morera et al., 2009; Ortega et al., 2018; Page et al., 2024; Santelices et al., 2015, 2017; Torres, 2018; Vivanco, 2013). Other variables have also shown significant effects, such as participation in extracurricular programs and study time (Eigbiremolen et al., 2020; Hernández & Olariaga, 2009; López-Martin et al., 2014), as well as distance to school and access to income transfer programs (Bartholo et al., 2019; Eigbiremolen et al., 2020).

Figure 3.5 – School trajectory variables



Source: Own elaboration (2026).

3.3.4 Main results of VA models

This section focuses on the corpus of studies that met the criteria of methodological adequacy, data transparency, and clarity in findings, which constitute the corpus examined in this topic. It contributes to the discussion raised in RQ4. Within these studies, the explanatory variables, including socioeconomic and demographic conditions, as well as student attitudes, are shown to influence educational performance in diverse ways depending on the context (Bartholo et al., 2019; Brooke et al., 2014; Ferrão, 2009, 2014, 2022a; Ferrão & Goldstein, 2009; Gomes-Neto et al., 1997; López-Martin et al., 2014; Morera et al., 2009; Muñoz-Chereau, 2019; Muñoz-Chereau & Thomas, 2016; Soares et al., 2017; Taut et al., 2016; Troncoso, 2019). The reviewed literature suggests that the effect of these variables tends to be positive when students have family support and are motivated (see, e.g., Hernández & Olariaga, 2009). Conversely, in contexts characterized by social vulnerability, marked by the dearth of support and resources, the effect can be detrimental (see, e.g., Vivanco, 2013). Furthermore, psychosocial and emotional variables, such as parental mental health and levels of neuroticism, have been demonstrated to exert a substantial influence on educational performance, manifesting adverse effects in situations of vulnerability (Berthelon et al., 2020; Muñoz-Chereau & Thomas, 2016).

The issue of gender disparity and its impact on educational performance is also extensively discussed in the literature (Bartholo et al., 2019; Brooke et al., 2014; Ferrão, 2014; Morera et al., 2009; Muñoz-Chereau, 2019; Muñoz-Chereau & Thomas, 2016; Taut et al., 2016; Troncoso et al., 2016). For instance, the works of Brooke et al. (2014) and Ferrão (2014), focusing on Brazil, explore factors explaining variations in learning outcomes across schools, revealing negative but minor effects on boys' performance in Portuguese, while in Mathematics, boys tend to perform slightly better than girls. Conversely, Muñoz-Chereau (2019) and Troncoso et al. (2016) investigated the performance gap between boys and girls in Chilean schools and found insignificant differences in progress between genders. These findings suggest that gender disparities in educational performance typically present differences of low magnitude.

Participation in extracurricular programs and the amount of time spent studying have been linked to improved academic performance, emphasizing the importance of factors external to the classroom (Eigbiremolen et al., 2020; Hernández & Olariaga, 2009; Muñoz-Chereau & Thomas, 2016; Troncoso, 2019). Research has demonstrated that variables such as the distance to school and access to income transfer programs are particularly salient to academic performance in developing countries (Eigbiremolen et al., 2020). Other studies have placed the socioeconomic context at the center of educational performance analysis, moving beyond treating it merely as a control variable and incorporating VA models into the discussion of social inequalities (Ferrão, 2022a). Recent research has sought to incorporate the effects of the pandemic[9] or nutritional factors, thereby introducing new dimensions to the analysis of educational performance, and demonstrating that improved health conditions are linked to better academic outcomes (Berthelon et al., 2020; Eigbiremolen et al., 2020; Gomes-Neto et al., 1997).

The corpus of articles examining student performance variation suggests that disparities in achievement are not solely attributable to individual student characteristics but also to school-related factors. In the case of intra-school variation, the teacher qualification, the quality of

[9] The repercussions of the novel Coronavirus (Covid-19) pandemic have had a substantial impact on educational performance, largely attributable to the necessity of adapting to remote learning modalities, ensuring access to technological resources, and re-evaluating teaching methodologies. The interrelated variables associated with these factors have been shown to exert a considerable negative influence on student learning outcomes (Bartholo et al., 2023).

teaching, the pedagogical approaches, and the learning environment are factors that significantly influence performance (Bertoni et al., 2024; Filho, 2019; Jerrim et al., 2019; Torres, 2018). Conversely, inter-school disparities are more closely associated with structural inequalities across schools, including resource availability and infrastructure (Filho, 2019; Muñoz & Prem, 2024; Thieme et al., 2016; Torres, 2018). These disparities help explain why schools in vulnerable regions or with limited support tend to perform worse than those with superior resources and higher investment. However, the reviewed studies highlight that educational policies, such as support programs for vulnerable schools and teacher training strategies, can reduce variation between schools and improve performance within schools.

3.4 DISCUSSION AND CONCLUSIONS

This systematic review aims to synthesize the extant literature on the application of VA models in education across Ibero-American contexts, examining study objectives, methods, variables used, and main findings. The evidence suggests that, although VA models are well-established tools for assessing school and teacher effectiveness in developed contexts (Koedel et al., 2015; Amrein-Beardsley et al., 2023; Levy et al., 2019; Chetty et al., 2014), their use in Ibero-America remains incipient (Muñoz-Chereau et al., 2020), heterogeneous, and concentrated in a small number of countries, particularly Chile (41% of the reviewed studies) and Brazil (22%). Only a limited number of studies have been identified in Portugal and Spain. However, analyses such as those by Ferrão (2025) indicate that research output in these countries is more extensive, suggesting that their underrepresentation in this review reflects limitations in the scope of the search.

With respect to the studies' objectives (RQ1), the findings indicate that most research focuses on evaluating teacher and school VA, as well as on analyzing contextual and social factors such as gender inequalities, family conditions, and pandemic-related impacts. This incorporation of local factors reflects an adaptation of established practices from other countries to the specific realities of the region, consistent with literature highlighting the need to contextualize VA models across different educational systems (Chetty et al., 2014; Paufler & Amrein-Beardsley, 2014). Consequently, despite growing academic interest, substantial gaps remain in both geographical coverage and the institutionalization of VA-based evaluation practices, limiting the potential of these analyses to inform equitable education policies.

Concerning the application of statistical models (RQ2), the widespread use of multilevel models underscores the well-established consensus that educational data possesses a hierarchical structure (Goldstein, 2003; Hox et al., 2017; Levy et al., 2019). However, in the Ibero-American context, most analyses remain confined to two levels (student-school), with limited attention to more complex structures involving classrooms, municipalities, or regions. This simplification may underestimate broader contextual effects and structural influences on learning, factors whose importance in VA studies is well documented (Blaskó et al., 2022; Guilherme et al., 2024; Ma et al., 2018; Nurse & Melhuish, 2021). These findings align with Amrein-Beardsley et al. (2023), who warn that simplified methodological choices may compromise the accuracy, stability, and validity of VA models, reducing their reliability for teacher evaluation in high-stakes contexts. Moreover, the use of diverse analytical approaches, including multivariate regression, fixed-effects models, dynamic models, frontier models, and quantile regression, reveals a lack of methodological standardization. This diversity hinders cross-study comparability and limits the development of robust evidence to inform education policy.

The analysis of the variables employed (RQ3) reveals a substantial evolution, shifting from conventional indicators, such as individual characteristics and SES, to the incorporation of psychosocial, motivational, and contextual factors, as well as more recent variables related to the pandemic and nutrition. This trend, well documented in the international literature (Branco et al., 2015; Durlak et al., 2011; Greenberg et al., 2003; Hoffman, 2009; Rubin et al., 2004; Sirin, 2005), indicates that VA models are being progressively refined to capture the multifaceted nature of the learning process. Such refinement recognizes that academic performance is influenced not only by cognitive skills but also by socioemotional competencies and family, social, and institutional contexts.

In addition, most reviewed studies employ standardized test scores in mathematics and language as the dependent variable. An exception is Gomes-Neto et al. (1997), who also consider the probability of school dropout and regular grade progression, thereby broadening the conceptualization of academic performance. This more comprehensive approach aligns with international literature arguing that indicators such as school dropout (Cao et al., 2024; Chan & Dai, 2023), grade retention (García-Pérez et al., 2014; Martin, 2011), and student success (Ferrão, 2023; Hancock et al., 2017) also constitute key factors of academic outcomes.

Although some of the studies primarily focus on higher education, their arguments reinforce the importance of expanding performance assessment beyond standardized tests.

In line with the broader literature on educational effectiveness (Barbosa & Fernandes, 2001; Braun, 2013; Charalambous et al., 2019; Chetty et al., 2014; Ferrão et al., 2001; Hanushek et al., 2005; Hanushek, 2019; Levy et al., 2019; Marder et al., 2020; Strand, 2010, 2012, 2016), the principal findings of this review (RQ4) indicate that contextual and social factors exert a substantial influence on school performance. Family support, student motivation, adequate health conditions, high-quality instruction, and sufficient infrastructure contribute positively to learning, whereas social vulnerability, limited resources, and low-quality instruction exacerbate inequalities both within and across schools.

Notwithstanding these advances, the reviewed studies reveal substantial gaps. For instance, there is a scarcity of research incorporating psychosocial and emotional variables, such as parental mental health, children's socioemotional development, and levels of neuroticism, even though these factors are increasingly linked to academic outcomes in recent studies (Clark et al., 2021; Loeb et al., 2019; Morando & Platt, 2022). This gap may reflect limited data availability or challenges in collecting such information. There is also a notable shortage of studies that examine additional contextual factors, including distance to school, time spent studying outside of school, participation in extracurricular activities, and access to income-transfer programs. Another underexamined dimension in the Ibero-American context concerns the relationship between school effectiveness and social inequality. Comparative analyses across institutional contexts, such as different states, municipalities, or school networks, are rare, and collaboration among researchers in the region is limited and sporadic, restricting the production of robust and comparable evidence. Finally, our understanding of the implications of applying VA models in high-stakes accountability contexts remains highly limited, likely due to the susceptibility of VA estimates to bias and instability (Baker et al., 2010; Paufler & Amrein-Beardsley, 2014; Rothstein, 2009, 2010), which renders them less suitable for high-stakes decision-making (Newton et al., 2010).

The primary limitation of this systematic review is the disproportionate concentration of scientific production on VA in only a few countries, particularly Chile and Brazil, which substantially reduces regional representativeness. This geographic concentration limits the review's ability to capture the institutional, socioeconomic, and territorial diversity that characterizes Ibero-America and, consequently, restricts a fuller understanding of how educational inequalities manifest across different contexts. As a result, the evidence base risks being overly influenced by the experiences of a small subset of countries, leading to interpretations that are less applicable or ill-suited. Most studies employ simplified multilevel models, primarily limited to two levels (student-school), which may underestimate broader contextual effects associated with classrooms, teachers, municipalities, or school networks. Finally, a key limitation is that this review did not focus on detailed empirical data from the individual studies, such as covariate distributions, effect-size comparability, or specific modeling specifications. By not centering on these concrete data points, the review provides a broad overview but limits the depth of quantitative insights and the capacity to assess methodological robustness across studies. Taken together, these limitations highlight the need for methodological refinement and for expanding research efforts to underrepresented Ibero-American countries.

In sum, this systematic review highlights both the advances and the persistent gaps in the existing literature, empirically corroborating, in the regional context, the diagnosis presented in the first essay regarding the still limited application of VA models in developing countries. At the same time, it provides clear guidance for future research. Considering the limitations identified, this study argues that research on VA in the Ibero-American region should adopt more sophisticated multilevel models with random coefficients and multiple levels (student, teacher, school, municipality, and state) to more accurately capture regional heterogeneity and broader contextual influences that two-level models cannot adequately address. Addressing this methodological limitation, the empirical study presented in the next chapter (Essay 3) implements a five-level model, enabling a more precise decomposition of variance and a robust analysis of differential effectiveness and social equity in Brazilian schools.

When levels such as municipalities or states are omitted, studies may overestimate the variability attributed to teachers or schools and underestimate structural factors that influence learning. Regarding variables, future research should incorporate psychosocial and emotional factors, additional contextual indicators (e.g., distance to school, time spent studying outside of school, and participation in extracurricular activities), and learning outcomes beyond standardized test scores (e.g., school dropout and grade retention). Following this recommendation, the subsequent essay complements the traditional proficiency analysis with a binary model of 'educational success', explicitly investigating the factors associated with students' regular progression and flow, encompassing dimensions often overlooked in the regional literature. The absence of these indicators suggests that the current literature offers only a partial view of student performance, overlooking factors that may exacerbate educational inequalities. Lastly, it is also essential to overcome the concentration of research in a limited number of countries by expanding geographic coverage and fostering sustained regional collaboration.

Taken together, these efforts are essential for building a more substantial and contextually grounded body of evidence capable of informing equitable public policies. Notably, without more diverse data and the use of models that accurately capture the multifaceted structure of educational systems, policies derived from VA studies risk reinforcing existing inequalities rather than mitigating them. This review not only provides a critical synthesis of the state of the art but also seeks to outline a methodological agenda for future VA research in Ibero-America, one that calls for richer datasets and greater analytical complexity to advance the development of more effective and inclusive educational policies. This thesis responds directly to this call in the third and final essay. By shifting from the diagnostic scope of this review to an empirical application using rigorous multilevel modeling and comparative data from contrasting Brazilian states (Minas Gerais and Maranhão), we seek to demonstrate how VA models can effectively diagnose equity gaps and inform public policy in unequal educational systems.

4. ESSAY 3: DIFFERENTIAL EFFECTIVENESS AND SOCIAL EQUITY: A VALUE-ADDED ANALYSIS IN BRAZIL AND IN THE STATES OF MARANHÃO AND MINAS GERAIS

ABSTRACT

Against the backdrop of global educational stagnation and the limited evidence on differential school effectiveness in Ibero-America, this study examines the capacity of Brazilian schools to promote learning and educational equity. The analysis combines a national perspective with a comparative focus on the states of Maranhão and Minas Gerais. Using longitudinal SAEB data that track students from 5th to 9th grade between 2011 and 2015, the study applies a five-level linear model to estimate academic achievement and a binary logistic model to analyze school success, controlling for prior achievement, socioeconomic status, and other relevant covariates. The results indicate that more than 90% of the variance in achievement is located within schools, with classroom-level effects suggesting systematic patterns of student grouping. Differential school effectiveness is low, indicating that schools differ little in their capacity to moderate the effects of prior achievement and students' socioeconomic background on academic outcomes. At the regional level, school infrastructure shows stronger associations with achievement in Maranhão, while racial penalties are more pronounced in Minas Gerais. The logistic model also reveals a gender paradox, whereby girls exhibit higher probabilities of school success despite lower mathematics performance. Overall, the findings highlight the predominance of within-school variation in educational outcomes and the limited dispersion of school effects, underscoring the importance of examining intra-school processes when assessing learning and equity.

Keywords: value-added models; differential school effectiveness; social equity; multilevel model.

4.1 INTRODUCTION

The global educational landscape, as documented in the 2024 Global Education Monitoring Report (UNESCO, 2024), is characterized by stagnation and regression with uneven effects: 33% of children and youth remain out of school in low-income countries, compared to only 3% in high-income countries. At the same time, learning levels have declined by approximately three points since 2018, reflecting a deterioration in educational quality. These setbacks have been compounded by a decline in public investment in education, which fell from 4.4% to 4% of global Gross Domestic Product (GDP) between 2015 and 2022, with four out of ten countries failing to meet minimum financing benchmarks. This context underscores that educational inequalities are not limited to access to schooling but extend to the quality of instruction and the availability of resources, disproportionately affecting children and youth in lower-income settings.

Educational disparities are sustained by mechanisms such as social stratification as it manifests within education systems, fostering school segregation (Bartholo et al., 2020; Krüger, 2020; Nurse & Melhuish, 2021; Pickett, 2014; Ribeiro, 2018). This phenomenon reduces opportunities for interaction among students from different backgrounds, which may constrain learning, particularly for the most disadvantaged (Agostinelli et al., 2020). In addition, other factors contribute to social inequalities in education and reinforce segregation, including students' socioeconomic origins (Ackerman et al., 2004; Breen et al., 2009; Gustafsson et al., 2011; Jackson, 2013; Marks et al., 2006; Melhuish et al., 2015; Shavit & Blossfeld, 1993; Sylva et al., 2004), family resources (Burger, 2019; Jensen et al., 2018; Schneeweis, 2011), and school-level factors such as infrastructure (Blaskó et al., 2022; Guilherme et al., 2024; Ma et al., 2018; Nurse & Melhuish, 2021), which are fundamental to educational success (Owens, 2018; Reardon & Owens, 2014).

In this context, the discussion of educational effectiveness, which integrates school effectiveness and teacher effectiveness, gains relevance by examining what makes a school "good" and how schools can be improved (Reynolds et al., 2014), thereby contributing to the reduction of educational inequalities within education systems (Burger, 2016; Lavy, 2016). At the school level, effectiveness refers to schools' capacity to promote student learning by considering pedagogical practices, management, and attention to students' socioeconomic and demographic conditions (Casillas, 2006; Closs et al., 2024; Coleman, 1968; Creemers, 2005;

Mortimore et al., 1988; Reynolds et al., 2014; Scheerens & Bosker, 1998). However, school effects do not manifest uniformly across students, as they are influenced by socioeconomic and demographic factors, highlighting the need to analyze differential effectiveness (Ferrão, 2022a, 2022b; Mortimore et al., 1988; Mortimore & Whitty, 2000; Sammons et al., 1997; Strand, 2010). It is generally assumed that schooling tends to reduce performance differences associated with socioeconomic background, particularly by benefiting students from less advantaged contexts (Burger, 2016; Lavy, 2016). It is at this juncture that the concept of educational equity becomes central, understood as the education system's commitment to ensuring fair access to and opportunities for learning, removing barriers that constrain students' development, and compensating for inequalities of origin (McLaughlin, 2010).

Research on the phenomenon of differential effectiveness remains limited, yet its investigation is highly relevant, as variations in schools' capacity to promote the progress of students from different socioeconomic backgrounds make it possible to identify practices and strategies associated with reducing inequalities, thereby informing equity-oriented educational policies (Kyriakides, 2004). Given that regional and socioeconomic disparities are directly reflected in achievement levels, comparisons across different educational contexts are essential to elucidate how these factors affect school effectiveness and student learning (Albernaz et al., 2002; Alves et al., 2007; Bonamino et al., 2010; Casillas, 2006; César & Soares, 2001; Coleman, 1968; Creemers, 2005; Medeiros & Oliveira, 2014; Mortimore et al., 1988; Reynolds et al., 2014; Scheerens & Bosker, 1998). Within this framework, the present study examines school effectiveness in Brazil, with particular attention to regional disparities. Ferrão et al. (2018, p. 283) reinforce this perspective by noting that "students who attend schools in the North and Northeast regions exhibit lower proficiency compared to their peers in other regions," underscoring the importance of analyzing regional differences in school performance.

In response to this context, this study was designed to examine the phenomenon within the Brazilian context, aiming to overcome the methodological limitations identified in the systematic review presented in the previous chapter. By implementing a five-level model (student, classroom, school, municipality, and state) and incorporating progression indicators, such as school success, the analysis allows for a more robust decomposition of educational achievement variance than that provided by the traditional two-level models predominant in the regional literature. Thus, synthesizing the trajectory of this thesis, which builds on the conceptual foundations in the first essay and the regional diagnostic in the second, this third and final study fulfills the objective of estimating school effectiveness, with a focus on differential effectiveness and social equity, through an analysis at both national and regional levels. To this end, the states of Maranhão and Minas Gerais were selected, which ranked among the lowest and highest educational outcomes, respectively, in 2011 and 2015 (Figure C1, Appendix C). The selection of these states allows for the exploration of clear contrasts in school performance while maintaining regional representativeness: Minas Gerais represents the Southeast, with more favorable educational and socioeconomic conditions, whereas Maranhão, in the Northeast, faces persistent challenges associated with poverty, inequality, and limited educational infrastructure (PNUD, IPEA, FJP, 2019).

These contextual differences are reflected in the average performance in Reading and Math (Figure C1, Appendix C). Maranhão (21) exhibits lower performance levels, whereas Minas Gerais (31) maintains higher averages throughout the analyzed period. To examine these contrasts, the study employs a multilevel Value-Added (VA) model, which estimates the differentiated contribution of schools to students' academic progress while controlling for prior achievement, socioeconomic status (SES), and other relevant covariates. The model highlights how structural inequalities between schools and regional differences are associated with school effectiveness, while also informing patterns of equity in learning.

Furthermore, the structural differences between these two states highlight the heterogeneity in the distribution of educational resources, such as the Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb)[10]. Although both Minas Gerais (31) and Maranhão (21) receive Fundeb resources, as shown in Figure C2 in Appendix C, the per capita allocation differs due to enrollments in contexts with higher weighting factors, such as full-time education, rural areas, and Indigenous or quilombola communities. In Maranhão, these resources are higher, likely due to a greater proportion of students in vulnerable situations. Minas Gerais, on the other hand, exhibits a lower concentration of these characteristics, receiving relatively smaller per capita transfers. The redistributive nature of Fundeb, combined with federal complements, aims to reduce socioeconomic and racial inequalities, explaining the differences in allocations even among states with distinct fiscal capacities.

---

[10] Fundeb is a special accounting fund (composed of 27 individual funds) intended to finance public basic education. It consists of portions of state, Distrito Federal, and municipal taxes, supplemented by federal contributions, functioning as a redistributive mechanism to ensure a minimum standard of quality and per-student funding (Constitutional Amendment No. 108/2020 and Law No. 14,113/2020).

In light of this context, the empirical analysis presented in the following sections deepens the understanding of school effectiveness and its implications for educational equity in Brazil and across distinct regional contexts. The discussion is grounded in the literature review (Section 4.2) and the institutional framework (Section 4.3), which guide the empirical strategy employed (Section 4.4), the analysis of results (Section 4.5), their interpretation (Section 4.6), and the concluding remarks (Section 4.7).

4.2 EDUCATIONAL EFFECTIVENESS AND SOCIAL EQUITY: FOUNDATIONS AND EVIDENCE

The pursuit of quality education has become a central pillar of the global policy agenda due to its fundamental role in shaping new generations and reducing poverty (Qi & Wu, 2019; Tilak, 2002). Educational reforms and policies have proven to be effective tools for promoting greater equality of opportunity, particularly for students from more vulnerable socioeconomic backgrounds (Alves & Ferrão, 2020; Brauw et al., 2015; Morais et al., 2021; Ryu et al., 2020). In this regard, research on the role of schools in student development has been conducted since the 1960s, reflecting sustained academic interest in school effectiveness (Amrein-Beardsley et al., 2016; Bryk & Weisberg, 1976; Coleman, 1968; Dumay et al., 2014; Ferrão, 2012, 2022a, 2022b; Ferrão & Couto, 2014; Hanushek, 1971; Hanushek et al., 2022; Karino & Laros, 2017; Kyriakides et al., 2019; Levy et al., 2019; Mortimore, 2014; Mortimore et al., 1988; Reynolds et al., 2002; Sammons, 2007).

While early studies indicated limited effects of schools on student achievement (Bernstein, 1970; Coleman et al., 1966; Jencks et al., 1972), later research emphasized the importance of pedagogical practices and school management (Edmonds, 1979; Mortimore et al., 1988; Rutter, 1980). In this context, educational effectiveness is defined as the capacity of schools and teachers to promote student learning by integrating pedagogical practices, school management, and students' socioeconomic and demographic contexts (Casillas, 2006; Closs et al., 2024; Coleman, 1968; Creemers, 2005; Mortimore et al., 1988; Reynolds et al., 2014; Scheerens & Bosker, 1998; Valle & Lillejord, 2023).

Beyond average effects, research on educational effectiveness has increasingly distinguished between mean effectiveness, differential effectiveness, and social equity. Within this broader perspective, studies in the field, has emphasized the importance of looking "inside" the school (Teddlie et al., 2002; Teddlie & Reynolds, 2001), recognizing that differences in student achievement are intrinsically linked both to the school's socioeconomic composition and to internal school characteristics, including leadership and the quality of the institutional environment. Analyses of the relationship between students' SES and their achievement highlight the effectiveness (Charalambous et al., 2018; Kyriakides et al., 2019; Sammons, 2007; Sammons et al., 1995; Thomas et al., 2007) and differential effectiveness (Palardy, 2008; Strand, 2010, 2011, 2016) of some schools in reducing educational inequalities. Differential effectiveness refers to schools' capacity to reduce achievement disparities by adapting pedagogical and management practices to the needs of subgroups defined by socioeconomic, ethnic, or gender factors, thereby promoting more equitable education (Ferrão, 2022a, 2022b; Mortimore et al., 1988; Mortimore & Whitty, 2000; Sammons et al., 1997; Strand, 2010).

In analytical terms, differential effectiveness operates along more than one dimension, capturing heterogeneity in schools' capacity to promote learning gains both across students with different prior achievement levels and across socioeconomic groups, the latter being interpreted in this study as social equity. From a theoretical standpoint, this distinction reinforces the view that effectiveness and equity are analytically distinct yet interdependent dimensions of educational systems. As argued by Ferrer-Esteban (2016), processes of social sorting that generate unequal distributions of students – particularly the concentration of socioeconomically disadvantaged students in specific schools or classrooms – tend to undermine overall system effectiveness, as the negative effects of segregation outweigh the more limited gains associated with the concentration of advantaged students.

A common approach to evaluating school effectiveness is the VA[11] model. This model allows for estimating the specific contribution of schools to student achievement while controlling for demographic variables, such as gender, race/ethnicity, and family characteristics (Kyriakides et al., 2019; Muñoz-Chereau, 2019; Muñoz-Chereau & Thomas, 2016; Nuttall et al., 1989; Soares et al., 2017; Strand, 2010; Taut et al., 2016; Troncoso, 2019); school-related factors, including infrastructure, instructional resources, teacher composition, school management, and class size (Blaskó et al., 2022; Filho, 2019; Guilherme et al., 2024; Ma et al., 2018; Marder et al., 2020; Martínez-Abad et al., 2020; Nurse & Melhuish, 2021; Santelices et al., 2015; Soares et al., 2017; Torres, 2018; Troncoso, 2019); and socioeconomic factors, given

[11] As discussed in the first essay of this thesis, this model was introduced into education by economist Eric A. Hanushek in 1971, in the study titled "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data", extending its application beyond the economic context.

that students' social background is one of the main determinants of educational disparities (Ackerman et al., 2004; Albernaz et al., 2002; Alves et al., 2007; Bonamino et al., 2010; Breen et al., 2009; César & Soares, 2001; Gustafsson et al., 2011; Marks et al., 2006; Medeiros & Oliveira, 2014; Melhuish et al., 2015; Shavit & Blossfeld, 1993; Soares et al., 2017). Another key predictor for calculating VA is students' prior achievement (Ferrão, 2018; Ferrão & Couto, 2013; Jerrim et al., 2019; Muñoz-Chereau, 2019; Muñoz-Chereau & Thomas, 2016; Taut et al., 2016).

From the perspective of economics of education, cognitive development is fundamentally a cumulative process in which current achievement reflects the entire history of family and school investments, as well as individual endowments (Todd & Wolpin, 2003, 2007). Because complete historical information on these factors is rarely available, value-added models address this omitted-variable problem by conditioning current outcomes on students' prior achievement. In this framework, prior achievement serves as a proxy for the accumulated stock of past investments and unobserved abilities, enabling the model to more accurately isolate the contemporaneous contribution of schools (Todd & Wolpin, 2007).

The most common dependent variables in VA models are standardized test scores, although some studies incorporate alternative indicators, such as school dropout (Cao et al., 2024; Chan & Dai, 2023; Gomes-Neto et al., 1997), grade retention (García-Pérez et al., 2014; Gomes-Neto et al., 1997; Martin, 2011), and student success (Ferrão, 2022b; Hancock et al., 2017), which also constitute key determinants of academic achievement. While recognized as relevant, these alternative variables have been less frequently employed in the literature as measures of school outcomes.

Such models can be estimated using various methodologies, including multilevel models (Goldstein, 2003), which are widely employed (Levy et al., 2019), as well as techniques such as latent transition models (Bartolucci et al., 2023) or growth curve models (Ortega et al., 2018; Palardy, 2008; Primi et al., 2010). Beyond assessing the effect of schools on student achievement, VA models also provide an essential perspective for understanding the relationship between educational effectiveness and social equity. The intersection of these concepts reflects the complexity of the education system and the challenges involved in reducing inequalities associated with factors such as SES, ethnicity, and gender.

The Brazilian literature on school effectiveness and differential effectiveness is still limited (Bartholo et al., 2019, 2023; Brooke et al., 2014; Ferrão, 2022a; Ferrão & Couto, 2013, 2014; Filho, 2019; Gomes-Neto et al., 1997; Soares et al., 2017), particularly with regard to comparative approaches across states or regions that integrate socioeconomic and school management factors. This essay seeks to address this gap by empirically assessing how schools operating within distinct institutional contexts (Maranhão, Minas Gerais, and Brazil) contribute to mitigating social inequalities. Considering this discussion, the following hypotheses are formulated for empirical evaluation:

Hypothesis 1: The effects of prior achievement, socioeconomic background, gender, and race on learning persist consistently across contrasting institutional contexts, indicating that such inequalities are independent of the overall level of development of the education system.

Hypothesis 2: Although average inequalities persist between contexts, patterns of differential effectiveness and social equity manifest heterogeneously across contrasting educational systems.

Hypothesis 3: The patterns of differential effectiveness and social equity observed in standardized tests are also reflected in alternative indicators of academic performance, such as students' regular progression throughout their schooling trajectory.

Given the discussion, it becomes essential to understand how the principles of educational effectiveness and equity are translated into Brazilian public policies. The institutional framework for basic education reflects, to varying degrees, this concern with quality and the reduction of inequalities by establishing goals, guidelines, and monitoring mechanisms aimed at improving learning outcomes and mitigating disparities. Within this context, the next section presents the institutional framework guiding the Brazilian education system, highlighting the strategies adopted to address structural challenges and promote greater equity across schools and school networks.

4.3 CONTEXT OF BASIC EDUCATION IN BRAZIL

The provision of basic education in Brazil is embedded within a complex federal structure, in which the guarantee of the right to education results from a collaborative arrangement among the Union, the states, the Distrito Federal, and the municipalities. According to the National Education Guidelines and Framework Law (LDB)[12], basic education comprises Early Childhood Education, Elementary Education, and Secondary Education. This

[12] Law n° 9.394/1996.

study focuses on Elementary Education, the initial stage of basic schooling, understood as a critical period for schools to contribute to reducing educational and social inequalities (Ferrão et al., 2018). The 1988 Federal Constitution and the LDB establish Elementary Education as a public right, assigning municipalities primary responsibility for providing Early Childhood and Elementary Education, while states share responsibility, particularly for the later grades. This division of responsibilities creates a heterogeneous landscape of administrative, technical, and financial capacities across the federative entities, with direct implications for teaching quality and educational equity.

In this context, financing plays a central role in attempts to mitigate structural inequalities. Municipalities, states, and the Distrito Federal are required to allocate at least 25% of their tax revenue to the Maintenance and Development of Education (MDE). A portion of these resources, automatically withheld at the source (20% of a basket of taxes and constitutional transfers), forms the Fund for the Maintenance and Development of Basic Education and the Valorization of Education Professionals (Fundeb)[13], the main redistributive mechanism within the Brazilian education system. This fund pools contributions from the federative entities and redistributes them based on student enrollment, weighted by factors that differentiate per-student costs (such as full-time, rural, Indigenous, or quilombola education), with federal supplementation to ensure a minimum per-student allocation (Constitutional Amendment No. 108, 2020). In this way, Fundeb serves a complementary and redistributive function, aiming to reduce disparities arising from the differing fiscal capacities of states and municipalities.

Despite this institutional design, significant differences persist in the effective capacity for educational investment. Although Fundeb functions to ensure minimum standards, the fund alone does not eliminate inequalities associated with the differing fiscal capacities of the federative entities. States and municipalities with lower own-source revenue face more severe resource constraints, which are reflected in disparities in school infrastructure. These limitations are particularly evident in the North and Northeast regions, where a substantial proportion of schools still lack basic services such as electricity and sanitation (IBGE, 2018). In Maranhão (Northeast), for example, these structural constraints are especially pronounced, whereas

[13] Originally established by Constitutional Amendment No. 53 of 2006 to replace Fundef (Fund for the Maintenance and Development of Elementary Education and the Valorization of Teaching), Fundeb operated on a temporary basis between 2007 and 2020. It is currently a permanent financing instrument, established by Constitutional Amendment No. 108/2020 and regulated by Law No. 14,113/2020.

federative entities with greater fiscal capacity, such as Minas Gerais (Southeast), enjoy greater financial autonomy to supplement educational expenditures beyond Fundeb allocations.

The functioning and outcomes of the education system are primarily monitored through the National System of Basic Education Assessment (SAEB), coordinated by the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Conducted biennially, SAEB evaluates student performance through standardized proficiency tests in Reading and Math and collects contextual information via questionnaires administered to students, teachers, and school principals (Decree No. 9,432, Art. 5, 2018). Complementarily, the Basic Education School Census, conducted annually by Inep, provides detailed data on students, classes, schools, and education professionals, covering both public and private institutions nationwide (Decree No. 6,425, 2008). By encompassing both the initial and final stages of Elementary Education, such as the 5th and 9th grades, these datasets allow for the analysis of educational inequalities and the investigation of schools' roles in promoting learning.

The guidelines and objectives that steer Brazilian educational policy are formalized in the National Education Plan (PNE 2014–2024), which sets targets aimed at the universalization of elementary education, literacy at the appropriate age, and the promotion of educational quality with equity. Although the PNE and the State Education Plans (PEE) were approved after part of the period analyzed in this study (2011–2015), these documents systematize structural challenges that were already present at that time and therefore serve as a normative reference for understanding the priorities and institutional constraints under which education networks operated.

At the state level, states incorporate national guidelines through their own strategies. The PEE of Maranhão emphasizes addressing social, racial, and territorial inequalities, reflecting a context marked by greater socioeconomic vulnerability and more severe structural constraints. In contrast, the PEE of Minas Gerais, while also aligned with national goals of equity and quality, highlights teacher valorization and the consolidation of educational quality as central pillars of action. These differences illustrate that, even under a common institutional framework, schools operate within distinct environments of resources and expectations.

Understanding this institutional arrangement, which integrates the federal structure, financing mechanisms, assessment systems, and normative guidelines, is essential for the empirical analysis conducted in this study. The regional and institutional inequalities described influence the functioning of schools and provide the necessary context for interpreting the results on school effectiveness and educational equity presented in the following sections. The next section outlines the empirical strategy used to assess the differentiated contribution of schools to student academic progress and to the mitigation of educational inequalities across distinct regional contexts.

4.4 EMPIRICAL STRATEGY

4.4.1 Econometric model

To identify the factors influencing student achievement, a multilevel linear modeling approach with random coefficients was employed, due to the suitability of this approach for the hierarchical structure of the data (Ferrão, 2022a; Goldstein, 2003; Hox et al., 2017; Singer, 1998). This approach allows for the examination of associations between variables across different levels of clustering. The models include variables intended to explain students' gains in proficiency, with prior achievement and students' SES highlighted as key explanatory variables, while controlling students' sociodemographic characteristics, school context, and leadership.

The specification of a five-level model is based on the systematic review conducted in the second essay of this thesis, which indicated that simplifications to two- or three-level models tend to overestimate school effects and underestimate heterogeneity between classrooms and municipal contexts. Most of the variance in student achievement is concentrated within schools (Chetty et al., 2014; Ferrão, 2022a; Hill et al., 2011; Koedel et al., 2015). Given this pattern, the present essay examines the decomposition of within-school variance by explicitly incorporating the classroom level, allowing for the distinction between effects associated with class organization and those that are strictly individual.

The classroom level is particularly relevant, as studies indicate that streaming practices generate significant heterogeneity within the same school (Cervini, 2016; Cervini, 2009; Torche, 2005; Troncoso, 2019). Omitting this level can introduce biases in the decomposition of variance and in the estimation of fixed and random effects, compromising the interpretation of results regarding school effectiveness and differential effectiveness (Martínez, 2012; Santelices et al., 2015). Furthermore, empirical evidence suggests that a substantial portion of performance variability is associated with classroom composition, such as the average achievement of students, reinforcing the need to empirically distinguish between classroom- and school-level effects (De Fraine et al., 2002; Timmermans & Thomas, 2015). In this sense, the multilevel structure adopted allows for a more precise interpretation of the estimates presented below.

The random coefficients modeling approach is employed to capture differential effectiveness, reflected in the variation of the prior achievement effect across units, and social equity, reflected in the variation of the SES effect across schools. This modeling allows the slopes of these relationships to vary randomly across schools. The general specification of the model, adapted to its five-level structure – where students are denoted by i, classrooms by j, schools by k, municipalities by m, and states (or Federative Unit – UF) by l – is presented below:

$$y_{ijkml}^{Reading} = \beta_{0jkml} + \beta_{1kml}DP_{ijkml} + \beta_{2kml}SES_{ijkml} + \boldsymbol{\beta}'\boldsymbol{X}_{ijkml} + e_{ijkml} \tag{1}$$

$$\beta_{0jkml} = \beta_0 + v_{0l} + f_{0ml} + u_{0kml} + r_{0jkml} \tag{2}$$

$$\beta_{1kml} = \beta_1 + u_{1kml} \tag{3}$$

$$\beta_{2kml} = \beta_2 + u_{2kml} \tag{4}$$

In Equation (1), corresponding to Level 1 (student), the response variable, $y_{ijkml}^{Reading}$, represents proficiency in Reading, and the term β_{0ijkml} is the model intercept at the student level. The coefficients $\beta_{\ 1kml},\ \beta_{2kml}\ e\ \boldsymbol{\beta}$ denote, respectively, the slopes for prior achievement (DP), SES, and control variables (e.g., gender, race/ethnicity, employment status). These coefficients capture the controlled associations between the covariates and the educational outcome, conditional on the model specification. The matrix $\boldsymbol{\beta}'\boldsymbol{X}_{ijkml}$ represents the linear combination of student-level covariates weighted by their respective coefficients, and the term e_{ijkml} represents the random error at the student level, assumed to follow $e_{ijkml} \sim \mathrm{N}(0, \sigma_e^2)$.

Equations (2), (3), and (4) formalize the multilevel structure of the model, decomposing the coefficients into a fixed component (population mean) and a random component (variability across units), as established in the multilevel modeling literature (Goldstein, 2010). Equation (2) defines the intercept, composed of a fixed part representing the population mean of proficiency and random components associated with Levels 2, 3, 4, and 5 (classroom, school, municipality, and state), capturing unobserved heterogeneity across these units. Equations (3) and (4) define the slope coefficients, allowing the effects of prior achievement and SES to vary across schools through random terms, enabling the analysis of differential effectiveness and social equity. The random effects associated with the intercepts at levels classroom, municipality, and state are assumed to follow normal distributions with zero means and variances $\sigma^2_{r_0}$, $\sigma^2_{f_0}$ and $\sigma^2_{v_0}$, respectively. However, at Level 3 (school), the inclusion of random coefficients implies that the vector of random effects $\boldsymbol{u}_{kml} = (u_{0kml}, u_{1kml}, u_{2kml})'$ follows a multivariate normal distribution with mean zero and a variance-covariance matrix $\Omega_u = \begin{pmatrix} \sigma^2_{u0} & \sigma_{u01} & \sigma_{u02} \\ \sigma_{u01} & \sigma^2_{u1} & \sigma_{u12} \\ \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} \end{pmatrix}$.

In this matrix, the diagonal elements $(\sigma^2_{u_0}, \sigma^2_{u_1}, \sigma^2_{u_2})$ represent the variances of the school intercept, the prior achievement slope, and the SES slope, respectively. The off-diagonal elements $(\sigma_{u_{01}}, \sigma_{u_{02}}, \sigma_{u_{12}})$ capture the covariances between these school-level effects.

An analogous model will be estimated for Math proficiency as the response variable, y^{Math}_{ijkml}, following the same specification.

In a second specification, a binary model is estimated in which the dependent variable indicates student success, taking a value of 1 for students who were not held back in primary education and progressed regularly from 5th to 9th grade (2011–2015), and a value of 0 for those who experienced grade retention or interruptions in their schooling trajectory, including dropout, such that they did not advance regularly during this period. These estimates allow for the assessment of whether the patterns observed in the models with continuous dependent variables (proficiency in Reading and Math) hold, even when considering an earlier period and without assuming causal relationships with subsequent covariates.

To model this binary variable, a multilevel logistic model with a logit link function is employed, which linearizes the relationship between the probability of success and the explanatory variables, under the assumption that $Z_{ijkml} \sim \text{Bernoulli}(P_{ijkml})$. The model specification is given by:

$$\log\left(\frac{P_{ijkml}}{1 - P_{ijkml}}\right) = L_{ijkml} = \beta_{0jkml} + \beta_1 DP_{ijkml} + \beta_2 SES_{ijkml} + \boldsymbol{\beta}'\boldsymbol{X}_{ijkml} \qquad (5)$$

where L_{ijkml} is the log-odds of student success, and the individual intercept, β_{0jkml}, behaves analogously to Equation (2) on the logit scale, although the dependent variable is binary and the student-level variance is determined by $P_{ijkml}(1 - P_{ijkml})$.

Multilevel models will be estimated for each dependent variable, with the gradual inclusion of blocks of explanatory variables to analyze how the addition of different sets of covariates affects the estimated coefficients. Model 1 (M1) includes prior achievement and students' SES. Model 2 (M2) is the full model and adds individual student characteristics as well as variables related to schools and school leadership. Additionally, for robust purposes, Model 3 (M3) will be estimated, which corresponds to M2 without the classroom level. In these models, the hierarchical structure involves five levels for the national analysis and four levels for the states of Maranhão and Minas Gerais.

4.4.2 Data

The empirical analysis uses microdata from the Basic Education Assessment System (SAEB/Prova Brasil) for the years 2011 (5th grade of Elementary School) and 2015 (9th grade of Elementary School)[14], from public schools, paired to track the same students over time. The datasets were obtained from INEP and include test scores, as well as questionnaires administered to students, schools, and school principals. The 2011 dataset was complemented with the Student Socioeconomic Status Index, while the 2015 dataset was augmented with variable capturing differentiated school location, extracted from the School Census of the same year, as well as annual Fundeb funding values by municipality, used as socioeconomic and educational finance indicators. It is worth noting that SAEB and School Census microdata are publicly accessible upon request to INEP. All stages of microdata processing and analysis followed confidentiality guidelines and best practices for handling microdata, including the removal of direct identifiers from auxiliary datasets. Since the study relies exclusively on secondary and de-identified publicly accessible data, it is exempt from evaluation by the Research Ethics Committee of the Universidade Federal de Juiz de Fora (UFJF).

The analysis used two datasets. The first includes students who took the Prova Brasil in 2011 and 2015, with proficiency scores observed, without inputing missing values, and excluding those who experienced interruptions in their school trajectory[15]. The second dataset comprises students observed in 5th grade in 2011, including all cases with a recorded school

[14] The empirical analysis was initially planned to include more recent cohorts. However, administrative and scheduling constraints during the research visit to the secure data room at one of the Protected Data Access Service (SEDAP) of the INEP reduced the time available for processing restricted microdata, thereby limiting the scope of the empirical analysis.

[15] It is important to note that the Brazilian elementary education system (Ensino Fundamental) comprises a complete 9-year cycle. Therefore, whenever the terms 'academic trajectory', 'educational trajectory', or 'school trajectory' are used throughout this essay, they strictly refer to the observed period between the 5th (2011) and 9th grades (2015). This represents a critical, yet partial, segment of the students' foundational schooling.

trajectory in that year, regardless of subsequent interruptions. In this dataset, missing data were handled through multiple imputations[16], first conducted by UF and subsequently consolidated to create the national dataset. To assess the assumption that missing values did not follow a systematic pattern, Little's test (Little, 1988) was applied to different samples from Maranhão, Minas Gerais, and Brazil, indicating that the missing data mechanism was not Missing Completely at Random (MCAR), thus justifying the use of multiple imputation to reduce potential biases. Additionally, many variables had more than 5% missing data in the analyzed locations.

Table 4.1 – Variables included in the continuous and binary models

Type of variable	Continuous model	Binary model
Previous achievement	Reading and Math proficiency in 2011	Reading and Math proficiency in 2011
Student socioeconomic characteristics	SES (2011), sex, self-declared race/color, worker student in 2011 and 2015	SES (2011), sex, self-declared race/color, worker student in 2011 and 2015
Family background	Literate mother/father, live with mother/father	Literate mother/father, live with mother/father
School characteristics	Repetition primary education, administrative dependency, school location and differentiated location[(a)], class size, tutoring at school, tenured teachers, school and pedagogical infrastructure[(b)]	Administrative dependency, school location, class size, school and pedagogical infrastructure
School leadership	Director's sex and higher education	
Resources	Log of Fundeb per student	Log of Fundeb per student

Source: Own elaboration (2026).

Note: [(a)] The variable "differentiated location" was recoded as a binary indicator, taking the value of 1 for schools located in settlement areas, Indigenous lands, quilombola territories, or sustainable use conservation units, and 0 otherwise. [(b)] The school infrastructure and pedagogical infrastructure indices were constructed as simple averages of dichotomous variables capturing the presence of different physical facilities of the school building and pedagogical resources in 2011. Higher values indicate greater availability of these resources.

The variables included in the models differ with respect to the reference year and are listed in Table 4.1. In the continuous model, the reference year is 2015, although some 2011 variables are also considered; in the binary model, the reference year is 2011. Some variables were excluded from the binary model due to a high proportion of missing data. For example, the supplementary instruction variable has 97% missing values, and even with multiple imputations, the information might be insufficient to produce reliable estimates and robust

[16] Missing data were handled using multiple imputation by chained equations implemented in R with the *mice* package. Imputations were performed separately by UF, generating 20 datasets per subset, following recommendations in the literature for datasets with moderate missingness (Bodner, 2008; White et al., 2011). Continuous variables were imputed using predictive mean matching, categorical variables using logistic or polytomous regression, while identifiers and the outcome variable were not imputed.

interpretations. The continuous model was estimated using Stata, while the binary model was estimated in R, as attempts to fit the multilevel logistic model on imputed data in Stata were unsuccessful. In R, the binary model was estimated using the *glmer* function, which implements Generalized Linear Mixed Models for binary responses, from the *lme4* package (Linear Mixed-Effects Models); however, this package does not provide standard errors for random effects.

4.5 RESULTS

4.5.1 Descriptive statistics

To guide the analyses of educational effectiveness conducted in this study – aimed at estimating the effectiveness of Brazilian schools, with an emphasis on differential effectiveness and social equity, and comparing national performance with that of Maranhão (MA) and Minas Gerais (MG) – this section presents the descriptive statistics of the sample. Table 4.2 shows the distribution of students' demographic, family, and school-related variables across the three geographic contexts, providing an initial overview of the conditions that influence student performance and educational trajectories.

The distribution of student characteristics reveals distinct profiles across the contexts analyzed. Although there is a slight female predominance in all regions, the composition by race/ethnicity shows more pronounced differences: MA exhibits a higher proportion of Pardo students, whereas MG and Brazil have relatively more self-identified White students. Regarding early labor market participation, in 2011 MA records the highest proportion of working students, while in 2015 this percentage is higher in MG, with Brazil falling in between. These differences also appear to be reflected in educational trajectories, as MA shows repetition rates prior to 2011 that are significantly higher than those of MG and Brazil, indicating greater educational vulnerability. In terms of family characteristics, parental education is lower in MA, where the proportion of literate mothers and fathers is smaller than in MG and Brazil. Additionally, the proportion of students living with both parents is slightly lower in MA, which may suggest disparities in family resources related to learning. Taken together, these findings highlight the importance of accounting for student heterogeneity when interpreting the educational effectiveness results presented below.

School characteristics highlight regional inequalities. Compared to MG and Brazil as a whole, MA has a predominance of municipal schools and a higher presence of institutions located in rural areas, whereas MG and the national context are characterized by a predominantly state-run and urban school network. Although the proportion of schools in differentiated locations is low across all contexts, it is relatively higher in MA, reflecting additional regional challenges associated with serving specific populations. The provision of supplementary instruction is more common in MG and Brazil, while the proportion of tenured (civil service) teachers is higher in MA and nationally, indicating potentially more favorable institutional conditions in these contexts. Furthermore, many school principals hold higher education degrees, although MA shows slightly lower percentages than MG and Brazil; across all contexts, women predominate in school leadership positions.

Table 4.2 – Distribution of students by demographic, family, and school characteristics

Variables		Maranhão		Minas Gerais		Brazil	
		Frequency	**%**	**Frequency**	**%**	**Frequency**	**%**
Sex	Male	14,829	42.43	73,502	45,57	526,157	46.00
	Female	19,822	56.72	87,206	54.06	612,791	53.58
	N/A	299	0.86	595	0.37	4,809	0.42
Color/Race	White	5,076	14.52	42,796	26.53	340,164	29.74
	Pardo	21,864	62.56	78,563	48.71	539,089	47.13
	Black	3,626	10.37	20,240	12.55	123,330	10.78
	Asian-descendant	1,065	3.05	5,357	3.32	36,512	3.19
	Indigenous	688	1.97	3,618	2.24	23,345	2.04
	Don’t know	2,340	6.70	9,692	6.01	74,443	6.51
	N/A	291	0.83	1,037	0.64	6,874	.60
Worker student 2011	No	28,988	82.94	142,354	88.25	1,002,179	87.62
	Yes	3,733	10.68	13,726	8.51	102,488	8.96
	N/A	2,229	6.38	5,223	3.24	39,090	3.42
Worker student 2015	No	26,448	75.67	113,462	70.34	842,272	73.64
	Yes	3,366	9.63	21,001	13.02	121,668	10.64
	N/A	5,136	14.70	26,840	16.64	179,817	15.72
Literate mother	No	2,868	8.21	4,397	2.73	44,516	3.89
	Yes	31,487	90.09	155,454	96.37	1,087,793	95.11
	N/A	595	1.70	1,452	0.90	11,448	1.00
Literate father	No	5,092	14.57	8,737	5.42	82,474	7.21
	Yes	28,488	81.51	147,633	91.53	1,027,958	89.88
	N/A	1,370	3.92	4,933	3.06	33,325	2.91
Live with mother	No	3,090	8.84	7,826	4.85	62,965	5.51
	Yes	27,381	78.34	130,949	81.18	921,916	80.60
	N/A	4,479	12.82	22,528	13.97	158,876	13.89
Live with father	No	9,871	28.24	38,536	23.89	291,004	25.44
	Yes	20,752	59.38	98,090	60.81	691,280	60.44
	N/A	4,327	12.38	24,677	15.30	161,473	14.12
Grade repetition before 2011	No	25,161	71.99	133,369	82.68	917,046	80.18
	Yes	8,084	23.13	22,610	14.02	197,218	17.24
	N/A	1,705	4.88	5,324	3.30	29,493	2.58
Administrative dependency	Federal	39	0.11	189	0.12	1,103	0.10
	State	4,720	13.51	117,578	72.89	693,725	60.65
	Municipal	30,191	86.38	43,536	26.99	448,929	39.25
Location of school	Urban	28,612	81.87	156,512	97.03	1,079,989	94.42
	Rural	6,338	18.13	4,791	2.97	63,768	5.58
Differentiated location	No	33,928	97.08	160,399	99.44	1,134,533	99.19
	Yes	1,022	2.92	904	0.56	9,224	0.81
Tutoring at school	No	5,811	16.63	18,690	11.59	118,927	10.40
	Yes	29,099	83.26	142,153	88.13	1,022,464	89.40
	N/A	40	0.11	460	0.29	2,366	0.21
Tenured teachers	≤50%	7,766	22.22	90,152	55.89	293,900	25.70
	>50%	26,908	76.99	70,076	43.44	843,185	73.72
	N/A	276	0.79	1,075	0.67	6,672	0.58
Director's sex	Male	9,231	26.41	38,625	23.95	293,900	25.70
	Female	25,593	73.23	122,042	75.66	843,185	73.72
	N/A	126	0.36	636	0.39	6,672	0.58
Higher education of the director	No	1,995	5.71	1,276	0.79	16,808	1.47
	Yes	32,123	91.91	157,709	97.77	1,075,980	94.07
	N/A	832	2.38	2,318	1.44	50,969	4.46
Total		39,950	100.00	161,303	100.00	1,143,757	100.00

Source: Own elaboration (2026).
Note: N/A = Not Available.

The mean and standard deviation statistics (Tables C1, C2, and C3 in Appendix C) corroborate the patterns identified in the frequency analysis, showing that students in more vulnerable conditions — such as repeaters, working students, those with lower parental education, or those enrolled in rural schools — systematically exhibit lower average proficiency levels. This pattern is consistent across the three contexts analyzed, both in 2011 and 2015, indicating that socioeconomic, family, and institutional inequalities persist throughout the educational trajectory. In MA, these disparities are more pronounced, reflecting lower overall performance and greater sensitivity to out-of-school conditions; in MG, group differences are less marked and average proficiency levels are higher. The results for Brazil generally align with those observed in MG, although mean values tend to be slightly lower. Overall, the findings indicate that the contrasts observed in 5th grade tend to persist into 9th grade, reflecting learning trajectories that vary in intensity across states but show consistent patterns at the national level.

Taken together, these elements indicate that students in MA are embedded in more unequal family and school contexts, highlighting the importance of considering structural differences between states when analyzing school effectiveness, differential effectiveness, and social equity. In this regard, MA and MG are examined as contexts positioned below and above the national average, respectively, allowing for a comparison of distinct educational realities. This initial characterization provides the contextual framework for interpreting the estimates presented in the following section, where the multilevel models estimated at the national and state levels are compared.

4.5.2 Multilevel regression model estimates for Brazil

Aiming to decompose learning inequalities across different levels of the educational system and to analyze patterns of educational effectiveness, differential effectiveness, and social equity, this section presents the results of the multilevel regression models estimated for Brazil. Table 4.3 shows the estimates of fixed and random effects for the traditional VA model (Model 1) and a model augmented with student sociodemographic and school- and municipality-level contextual variables (Model 2) for Brazil, with Reading and Math proficiency as the dependent variables. Both models respect the nested structure of the data, allowing the decomposition of performance variance across different levels and the identification of the specific contribution of each level.

Table 4.3 – Estimates of the fixed and random effects in the multilevel regression model: Brazil (Model 1 and 2)

(continues)

	Model 1		Model 2	
	Reading	**Math**	**Reading**	**Math**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Fixed effects				
Constant	0.6312***	0.4216***	0.7307***	0.7670***
	(0.0136)	(0.0117)	(0.0720)	(0.0686)
Proficiency 2011	0.6397***	0.5719***	0.5930***	0.5465***
	(0.0010)	(0.0010)	(0.0010)	(0.0011)
SES 2011	0.0251***	0.0400***	0.0240***	0.0283***
	(0.0010)	(0.0009)	(0.0011)	(0.0010)
Female vs. Male			0.1323***	-0.0987***
			(0.0014)	(0.0013)
Pardo vs. White			-0.0378***	-0.0314***
			(0.0017)	(0.0016)
Black vs. White			-0.0828***	-0.0757***
			(0.0025)	(0.0024)
Asian-descendant vs. White			-0.0077*	-0.0077**
			(0.0038)	(0.0036)
Indigenous vs. White			-0.0343***	-0.0362***
			(0.0051)	(0.0047)
Don't know vs. White			-0.1196***	-0.0799***
			(0.0032)	(0.0029)
Worker student 2011 (Yes vs. No)			-0.0916***	-0.0295***
			(0.0025)	(0.0023)
Worker student 2015 (Yes vs. No)			-0.1192***	-0.0550***
			(0.0022)	(0.0020)
Literate mother (Yes vs. No)			0.0381***	0.0303***
			(0.0035)	(0.0033)
Literate father (Yes vs. No)			0.0476***	0.0384***
			(0.0027)	(0.0025)
Live with mother (Yes vs. No)			0.0466***	0.0422***
			(0.0029)	(0.0027)
Live with father (Yes vs. No)			0.0112***	0.0180***
			(0.0015)	(0.0014)
Grade repetition (Yes vs. No)			-0.1808***	-0.1511***
			(0.0020)	(0.0018)
State school vs. Federal			-0.3892***	-0.5607***
			(0.0385)	(0.0401)
Municipal school vs. Federal			-0.3229***	-0.5047***
			(0.0385)	(0.0401)
Rural vs. Urban			0.0024	0.0305***
			(0.0052)	(0.0045)
School in differentiated location (Yes vs. No)			0.0008	0.0157
			(0.0126)	(0.0108)
Class size			0.0005**	0.0005***
			(0.0002)	(0.0002)
Tutoring at school (Yes vs. No)			0.0183***	0.0125***
			(0.0042)	(0.0037)
Tenured teachers (>50% vs. ≤50%)			0.0136***	0.0093***
			(0.0032)	(0.0028)
School infrastructure			0.0449***	0.0291***
			(0.0069)	(0.0060)
Pedagogical infrastructure			0.0086	0.0196***
			(0.0076)	(0.0066)

(conclusion)

	Model 1		Model 2	
	Reading	Math	Reading	Math
	Coefficient	Coefficient	Coefficient	Coefficient
School head: female vs. male			0.0185***	0.0103***
			(0.0030)	(0.0026)
				(conclusion)
School head: higher education (Yes vs. No)			0.0169*	0.0059
			(0.0097)	(0.0083)
Log of Fundeb per student			0.0007	0.0090***
			(0.0049)	(0.0045)
Random Parameters				
Level: States				
Var(Cons)	0.0045	0.0033	0.0046	0.0028
	(0.0014)	(0.0010)	(0.0013)	(0.0008)
Level: Municipality				
Var(Cons)	0.0076	0.0077	0.0074	0.0074
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Level: School				
Var(Cons)	0.0171	0.0252	0.0151	0.0225
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Var(Proficiency 2011)	0.0030	0.0055	0.0024	0.0052
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Var(SES)	0.0009	0.0010	0.0007	0.0008
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Covar(Proficiency 2011-SES)	-0.0005	-0.0001	-0.0004	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Covar(Proficiency 2011-Cons)	0.0017	0.0106	0.0010	0.0097
	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Covar(SES-Cons)	0.0011	0.0004	0.0011	0.0003
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Level: Class				
Var(Cons)	0.0215	0.0120	0.0189	0.0114
	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Level: Student				
Var(Cons)	0.3966	0.3345	0.3783	0.3275
	(0.0006)	(0.0005)	(0.0006)	(0.0005)
Nº of states	27	27	27	27
N° of municipalities	5,159	5,159	5,073	5,073
N° of schools	27,718	27,718	26,450	26,449
N° of classes	70,390	70,390	66,614	66,614
N° of students	962,575	962,575	831,852	831,842
Deviance	1,900,382	1,730,809	1,602,049	1,478,022
AIC	1,900408	1,730,836	1,602,125	1,478,099
BIC	1,900,561	1,730,989	1,602,567	1,478,541

Source: Own elaboration (2026).

Note: Standard errors in parentheses. ** $p < 0.10$, **** $p < 0.05$, *** $p < 0.01$. Deviance = −2 × log-likelihood; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion. Model 1 includes prior achievement and SES. Model 2 corresponds to the full model, adding student characteristics and school- and leadership-level variables.

The fixed effects show the average effects of individual and contextual characteristics on student performance. As expected, 2011 proficiency exerts a positive influence on 2015 scores, indicating persistence in learning trajectories, while SES remains significantly associated, albeit to a lesser extent in both subjects. In Model 2, the inclusion of sociodemographic variables reveals patterns well documented in the educational inequality literature, such as differences by gender, race, child labor, family composition, and parental education, while also highlighting the importance of schools' pedagogical and structural resources. Furthermore, this contextualized model improves overall model fit, with substantial reductions in deviance and in the AIC and BIC criteria, indicating that part of the observed educational inequalities can be explained by observable individual, family, and school characteristics. Given this superior fit, subsequent analyses are based on this model.

The random effects allow for quantifying the residual variance at different levels of the educational system, even after controlling individual and school characteristics. At the state (0.0046 and 0.0028 for Reading and Math, respectively) and municipality (0.0074 for both subjects) levels, the relatively low intercept variances indicate that, although regional influence exists, its contribution to performance inequality is limited. At the school level, however, variances are higher (0.02 for both subjects), reflecting more substantive heterogeneity among institutions. Additionally, the presence of non-negligible variance at the classroom level (0.02 and 0.01) indicates heterogeneity between classes within the same school, possibly associated with streaming practices, which continue to influence student performance. Finally, student-level variance (0.38 and 0.33) exceeds the sum of the higher-level variances, highlighting the strong effect of individual and family factors on student achievement. Thus, approximately 91% of the variance in Reading proficiency and 93% in Math occurs within schools.

In addition, the model includes random slopes for prior proficiency and SES at the school level, allowing for testing whether the relationship between these variables and performance varies across schools. Although the variances associated with these slopes are statistically different from zero, their magnitudes are small, indicating that between-school variation is limited. Therefore, rather than reflecting substantive differences among institutions, these results suggest that, in Brazil, differential effectiveness (0.0024 for Reading and 0.0052 for Math) and school equity (0.0007 for Reading and 0.0008 for Math) tend to operate moderately, with only marginal variation in schools' sensitivity to prior performance and SES. The covariances between intercepts and slopes follow the same pattern: although statistically different from zero, their small magnitude indicates that the school's average performance level is weakly associated with differences in the relationships with prior proficiency (0.0010 for Reading and 0.0097 for Math) or SES (0.001 for Reading and 0.0003 for Math), as well as with a weak inverse association between sensitivity to prior proficiency and to SES (-0.0004 for Reading and -0.0002 for Math). Overall, Reading and Math exhibit very similar patterns in both fixed and random effects, differing primarily in the magnitude of the random intercept variances – higher at the school level in Math and more concentrated at the student level in Reading – differences that, however, do not alter the substantive interpretation of the model.

This structure allows for a direct examination of the study's three central dimensions. School-level mean effectiveness is represented by the random intercept, with variances indicating differences between institutions, although reduced after controlling for student characteristics. Differential effectiveness is captured by the random slopes for prior proficiency: while statistically significant, their small variances suggest limited variation in schools' contribution to students with different initial performance levels. Social equity, differential effectiveness based on socioeconomic composition, is assessed through both the fixed effects of demographic and socioeconomic variables, highlighting persistent inequalities, and the random slope for SES, whose small variance indicates that schools differ only modestly in their ability to moderate the effect of SES on performance. Overall, most variability is concentrated within schools, particularly at the student and classroom levels, whereas differences between schools, municipalities, and states are substantially smaller.

These results are consistent with the approach used by Ferrão (2022), although with four levels and a more limited set of independent variables. In Ferrão's study, both differential effectiveness and social equity exhibited very slight variation at the school level and were statistically null at the municipal and state levels, indicating that these dimensions contribute little to differences across territories. As in the findings of the present study, the largest share of variability was attributed to factors internal to the school, reinforcing that learning inequalities primarily emerge within schools.

The national findings thus establish a reference point from which to assess the extent to which state-level contexts reproduce or diverge from this overall pattern. Although the low variances at the state and municipal levels indicate that most inequalities occur within schools, this does not diminish the importance of analyzing specific state contexts. Each federative unit organizes and operates its educational system under its own institutional and socioeconomic conditions, which may generate distinct learning and inequality dynamics. Examining MA and MG separately allows for the identification of internal structures of variability, effectiveness, and equity that remain hidden in the national model, while also enabling a comparison of their results with the national pattern to identify potential divergences. The next subsection delves deeper into this analysis, presenting the models estimated for both states and discussing the extent to which their patterns align with or depart from the national framework.

4.5.3 Multilevel regression model estimates for Maranhão and Minas Gerais

Table 4.4 highlights substantive differences between MA and MG in how individual, family, and school characteristics are associated with performance in Reading and Math. In both states, the fixed effects of prior proficiency and SES follow patterns very similar to those observed for Brazil as a whole (see Table 4.3): high, consistent, and statistically significant coefficients, reaffirming the strong persistence of educational trajectories. However, the magnitudes differ. In MG, the effects of prior proficiency are, on average, higher than in MA, suggesting greater persistence in learning trajectories. This trend already reflects a first regional distinction: even in contexts with higher average performance and more structured school infrastructure, such as MG, students' past performance remains more strongly associated with future outcomes. These results are also consistent with those obtained from the more parsimonious model without contextual variables, in Table C4 of Appendix C.

Regarding sociodemographic variables, MA exhibits a larger disadvantage for girls in Math, whereas MG shows a greater advantage for girls in Reading. Concerning race/ethnicity, coefficients in MA tend to be smaller and, in some cases, not statistically significant, likely reflecting the greater racial homogeneity of the student population, which is predominantly self-identified as Pardo (see Table 4.2). In MG, which combines greater racial diversity with higher overall academic performance, penalties for self-identified Black, Pardo, and Indigenous students remain more pronounced and statistically significant, indicating that racial inequalities are evident even in contexts with higher average performance. These patterns suggest that the magnitude of racial disparities varies according to each state's educational context.

Table 4.4 – Estimates of the fixed and random effects in the multilevel regression model for Maranhão and Minas Gerais, controlling for contextual variables (Model 2)

(continues)

	Maranhão		Minas Gerais	
	Reading	**Math**	**Reading**	**Math**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Fixed effects				
Constant	1.4028***	0.8861***	-0.1396	0.1120
	(0.3304)	(0.3142)	(0.1842)	(0.1777)
Proficiency 2011	0.6063***	0.4873***	0.5927***	0.5683***
	(0.0063)	(0.0070)	(0.0027)	(0.0026)
SES 2011	0.0160***	0.0356***	0.0151***	0.0316***
	(0.0058)	(0.0053)	(0.0029)	(0.0027)
Female vs. Male	0.0849***	-0.1450***	0.1851***	-0.0427***
	(0.0080)	(0.0075)	(0.0039)	(0.0036)
Pardo vs. White	-0.0040	0.0305***	-0.0367***	-0.0335***
	(0.0115)	(0.0107)	(0.0046)	(0.0043)
Black vs. White	-0.0164	0.0124	-0.0771***	-0.0824***
	(0.0159)	(0.0148)	(0.0065)	(0.0061)
Asian-descendant vs. White	0.0275	0.0487**	-0.0192*	-0.0281***
	(0.0229)	(0.0214)	(0.0103)	(0.0096)
Indigenous vs. White	0.0482	0.0313	-0.0654***	-0.0489***
	(0.0297)	(0.0276)	(0.0132)	(0.0123)
Don’t know vs. White	-0.0623***	-0.0046	-0.1138***	-0.0750***
	(0.0190)	(0.0177)	(0.0090)	(0.0084)
Worker student 2011 (Yes vs. No)	-0.0998***	-0.0352***	-0.0791***	-0.0335***
	(0.0127)	(0.0118)	(0.0069)	(0.0064)
Worker student 2015 (Yes vs. No)	-0.1416***	-0.0606***	-0.1054***	-0.0506***
	(0.0126)	(0.0117)	(0.0053)	(0.0050)
Literate mother (Yes vs. No)	0.0196	0.0227*	0.0187*	0.0446***
	(0.0142)	(0.0132)	(0.0113)	(0.0105)
Literate father (Yes vs. No)	0.0465***	0.0252**	0.0553***	0.0482***
	(0.0109)	(0.0102)	(0.0081)	(0.0075)
Live with mother (Yes vs. No)	0.0288**	0.0173	0.0589***	0.0382***
	(0.0131)	(0.0122)	(0.0082)	(0.0076)
Live with father (Yes vs. No)	-0.0130	-0.0091	0.0223***	0.0224***
	(0.0085)	(0.0079)	(0.0042)	(0.0039)
Grade repetition (Yes vs. No)	-0.1539***	-0.1316***	-0.2017***	-0.1799***
	(0.0097)	(0.0090)	(0.0058)	(0.0054)
State school vs. Federal	-0.6233***	-0.5237**	-0.3378***	-0.6286***
	(0.1898)	(0.2213)	(0.0941)	(0.0987)
Municipal school vs. Federal	-0.6257***	-0.4958**	-0.1961**	-0.5143***
	(0.1900)	(0.2215)	(0.0938)	(0.0984)
Rural vs. Urban	-0.0575***	-0.0339**	0.0138	0.0431***
	(0.0171)	(0.0145)	(0.0182)	(0.0164)
School in differentiated location (Yes vs. No)	0.0597*	0.0904***	0.0081	-0.0165
	(0.0342)	(0.0290)	(0.0418)	(0.0371)
Class size	-0.0005	0.0006	0.0039***	0.0038***
	(0.0009)	(0.0007)	(0.0006)	(0.0006)
Tutoring at school (Yes vs. No)	0.0027	0.0266*	0.0113	0.0197*
	(0.0166)	(0.0140)	(0.0117)	(0.0106)
Tenured teachers (>50% vs. ≤50%)	0.0271*	0.0024	-0.0019	0.0103
	(0.0149)	(0.0126)	(0.0083)	(0.0076)
School infrastructure	0.0829***	0.0388*	0.0284*	0.0174
	(0.0263)	(0.0221)	(0.0168)	(0.0153)
Pedagogical infrastructure	-0.1052***	-0.0482	0.0027	0.0182
	(0.0363)	(0.0305)	(0.0218)	(0.0198)

(conclusion)

	Maranhão		Minas Gerais	
	Reading	**Math**	**Reading**	**Math**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
School head: female vs. male	-0.0055	-0.0233**	0.0158*	0.0108
	(0.0140)	(0.0118)	(0.0086)	(0.0078)
				(conclusion)
School head: higher education (Yes vs. No)	0.0363	0.0107	0.0017	0.0301
	(0.0251)	(0.0213)	(0.0373)	(0.0339)
Log of total Fundeb per student	-0.0243	-0.0074	0.0667***	0.0588***
	(0.0215)	(0.0177)	(0.0136)	(0.0127)
Random effects				
Level: Municipality				
Var(Cons)	0.0110	0.0067	0.0055	0.0054
	(0.0023)	(.)	(0.0008)	(.)
Level: School				
Var(Cons)	0.0213	0.0438	0.0117	0.0163
	(0.0046)	(.)	(0.0012)	(.)
Var(Proficiency 2011)	0.0039	0.0134	0.0018	0.0028
	(0.0015)	(.)	(0.0005)	(.)
Var(SES)	0.0015	0.0004	0.0009	0.0009
	(0.0013)	(.)	(0.0005)	(.)
Covar(Proficiency 2011-SES)	-0.0004	-0.0012	0.0000	0.0008
	(0.0010)	(.)	(0.0003)	(.)
Covar(Proficiency 2011-Cons)	0.0065	0.0238	-0.0006	0.0067
	(0.0023)	(.)	(0.0005)	(.)
Covar(SES-Cons)	0.0004	-0.0015	0.0023	0.0022
	(0.0017)	(.)	(0.0005)	(.)
Level: Class				
Var(Cons)	0.0125	0.0083	0.0270	0.0209
	(0.0018)	(.)	(0.0011)	(.)
Level: Student				
Var(Cons)	0.3461	0.3016	0.3868	0.3354
	(0.0034)	(.)	(0.0017)	(.)
N° of municipalities	214	214	822	822
N° of schools	1,214	1,214	3,282	3,282
N° of classes	2,495	2,495	8,402	8,402
N° of students	24,701	24,701	116,654	116,654
Deviance	45,333	41,845	227,811	211,023
AIC	45,407	41,901	227,885	211,079
BIC	45,707	42,128	228,242	211,350

Source: Own elaboration (2026).

Note: Standard errors in parentheses. ** $p < 0.10$, **** $p < 0.05$, *** $p < 0.01$. Deviance = −2 × log-likelihood; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion. Standard errors marked as (.) indicate variances too small or unstable to estimate precision. Model 2 corresponds to the full specification, including student characteristics as well as school- and leadership-level variables.

The divergences are particularly pronounced in the random effects. In Maranhão, municipal-level variance is higher, whereas in Minas Gerais it appears lower; for Math in MG, the model did not provide a standard error, indicating limited precision. At the school level, MA exhibits greater heterogeneity (0.04 in Math) compared to MG (0.02). Classroom-level variance is more pronounced in MG (0.03 in Reading and 0.02 in Math), indicating stronger within-class heterogeneity. At the student level, MG shows higher residual variance, particularly in Reading, highlighting the substantial contribution of individual and family factors. In both states, student-level variance exceeds the sum of higher-level variances, with most proficiency variance occurring within schools: 91% in Reading and 86% in Math in MA, and 96% and 94% in MG. These results are also consistent with those obtained from the more parsimonious model without contextual variables in Table C4 of Appendix C.

With respect to average school effectiveness, Brazil and Minas Gerais display moderate levels of between-school variance, whereas Maranhão presents relatively higher – though still modest (0.04) – coefficients, suggesting homogeneous systems in terms of schools' average capacity to promote learning. In contrast, differential effectiveness is more heterogeneous in Maranhão (0.0039 for Reading and 0.0134 for Math), indicating greater school-level differentiation in the association between prior student achievement and subsequent learning, while Minas Gerais (0.0018 for Reading and 0.0028 for Math) shows comparatively lower variability. A similar pattern emerges in the dimension of social equity: Brazil as a whole shows sensitivity to students' SES, with moderate and comparable school-level variances in Reading and Math. Maranhão, however, appears as a context with relatively greater inequality, combining stronger fixed SES effects with a larger, but statistically marginal, variance in the random SES slope for Reading (0.0015). This indicates that the influence of socioeconomic background on learning varies across schools, particularly in this subject.

Taken together, these findings indicate that, although Brazil, MA, and MG exhibit similar trends in both subjects, MA is characterized by greater heterogeneity between schools in terms of differential effectiveness and social equity. Even so, the magnitude of this variance is modest across all contexts analyzed, suggesting that schools differ little in their capacity to moderate the effects of prior performance or SES on learning.

4.5.4 Binary multilevel model estimates

Complementing the proficiency analyses presented in Sections 4.2 and 4.3, a multilevel logistic model was estimated with school success as the dependent variable, as shown in Table 4.5, which reports both the estimated logit coefficients and the corresponding odds ratios. This analysis allows us to assess whether the factors that influence standardized test scores also affect the likelihood that a student progresses regularly through the school system. Descriptive statistics for this variable and the main covariates are presented in Appendix C (Table C5), highlighting relevant differences between MA, MG, and Brazil in both the distribution of school success and the students' sociodemographic and school *characteristics*, as well as the non-negligible presence of missing data, detailed in Table C6.

The comparison between school success and proficiency outcomes highlights structural convergences while also revealing important nuances regarding gender, race, and infrastructure. As in the proficiency models, the binary model reinforces that prior performance and socioeconomic status (SES) are the strongest determinants of educational trajectories. Reading and Mathematics proficiency in 2011 show a large multiplicative impact, indicating that success in 9th grade is strongly conditioned by the cognitive foundation established in 5th grade: an increase in prior performance in Reading and Mathematics more than doubles a student's odds of success (OR for Reading = 2.309 in MA, 2.056 in MG, and 2.052 in Brazil; OR for Mathematics = 2.106 in MA, 2.377 in MG, and 2.234 in Brazil). Similarly, SES and parental education significantly increase the probability of success (for example, having a literate mother increases the odds by nearly 90% in Maranhão, OR = 1.893), reaffirming the role of family background in the reproduction of educational inequalities. These patterns are already evident in the more parsimonious model (Table C7, Appendix C), indicating that the dependence of school success on cognitive background and social origin precedes the incorporation of school-level characteristics.

The binary model also reinforces the regional racial paradox: in Brazil and in MG, Black, Pardo, and Indigenous students face statistically significant penalties in the probability of school success (in MG, Black students have 24% lower odds of success, OR = 0.760; Pardo students 12.3% lower, OR = 0.877; and Indigenous students 24.6% lower, OR = 0.754, compared with White students). In MA, however, these variables are not statistically significant or do not have comparable magnitudes (OR for Black students = 0.860; OR for Pardo students = 1.014). This suggests that, in contexts of high vulnerability such as MA, race loses predictive power as a statistical discriminator relative to structural poverty, whereas in more established contexts such as MG, racial inequality manifests more clearly in students' progression through the school system.

Table 4.5 – Estimates of fixed and random effects from the multilevel binary regression model with contextual controls (Model 2)

(continues)

	Maranhão		Minas Gerais		Brazil	
	Coefficient	**Odds ratios**	**Coefficient**	**Odds ratios**	**Coefficient**	**Odds ratios**
Fixed effects						
Constant	-0.109	0.897	-2.229**	0.108	-3.265***	0.038
	(1.103)		(0.650)		(0.257)	
Reading 2011	0.837***	2.309	0.721***	2.056	0.719***	2.052
	(0.023)		(0.009)		(0.003)	
Math 2011	0.745***	2.106	0.866***	2.377	0.804***	2.234
	(0.024)		(0.009)		(0.003)	
SES 2011	0.265***	1.303	0.234***	1.264	0.256***	1.292
	(0.034)		(0.013)		(0.005)	
SES 2011 x Reading 2011	0.056**	1.058	0.054***	1.055	0.025***	1.025
	(0.021)		(0.011)		(0.004)	
SES 2011 x Math 2011	0.110***	1.116	0.064***	1.066	0.093***	1.097
	(0.021)		(0.011)		(0.004)	
Female vs. Male	0.690***	1.994	0.679***	1.972	0.625***	1.868
	(0.019)		(0.011)		(0.004)	
Pardo vs. White	0.014	1.014	-0.131***	0.877	-0.108***	0.898
	(0.023)		(0.013)		(0.004)	
Black vs. White	-0.151***	0.860	-0.275***	0.760	-0.260***	0.771
	(0.035)		(0.020)		(0.007)	
Asian-descendant vs. White	0.477**	1.611	0.449***	1.567	0.716***	2.046
	(0.192)		(0.113)		(0.040)	
Indigenous vs. White	-0.146**	0.864	-0.283***	0.754	-0.204***	0.815
	(0.063)		(0.034)		(0.012)	
Don't know vs. White	-0.020	0.980	-0.218***	0.804	-0.181***	0.834
	(0.035)		(0.020)		(0.006)	
Worker student 2011 (Yes vs. No)	-0.560***	0.571	-0.465***	0.628	-0.489***	0.613
	(0.027)		(0.016)		(0.006)	
Literate mother (Yes vs. No)	0.638***	1.893	0.558***	1.747	0.566***	1.761
	(0.036)		(0.030)		(0.009)	
Literate mother (Don't know vs. No)	0.263***	1.301	0.138**	1.148	0.142***	1.153
	(0.062)		(0.057)		(0.016)	
Literate father (Yes vs. No)	0.337***	1.401	0.383***	1.467	0.347***	1.415
	(0.030)		(0.026)		(0.008)	
Literate father (Don't know vs. No)	0.119**	1.126	0.014	1.014	0.015	1.015
	(0.042)		(0.036)		(0.011)	
Live with mother (Yes vs. No)	0.353***	1.423	0.459***	1.582	0.435***	1.545
	(0.043)		(0.030)		(0.010)	
Live with father (Yes vs. No)	0.197***	1.218	0.284***	1.328	0.246***	1.279
	(0.025)		(0.014)		(0.005)	

(conclusion)

	Maranhão		Minas Gerais		Brazil	
	Coefficient	**Odds ratios**	**Coefficient**	**Odds ratios**	**Coefficient**	**Odds ratios**
School infrastructure	-0.086	0.918	0.069	1.071	-0.065*	0.937
	(0.114)		(0.080)		(0.033)	
Pedagogical infrastructure	-0.019	0.981	-0.004	0.996	0.002	1.002
	(0.023)		(0.021)		(0.006)	
Class size	0.011***	1.011	0.030***	1.030	0.024***	1.024
	(0.002)		(0.002)		(0.001)	
State school vs. Federal	0.833	2.300	0.209	1.232	0.642***	1.900
	(0.570)		(0.328)		(0.125)	
Municipal school vs. Federal	0.679	1.972	0.002	1.002	0.540***	1.716
	(0.569)		(0.327)		(0.125)	
Rural vs. Urban	0.110**	1.116	0.258***	1.294	0.244***	1.276
	(0.037)		(0.041)		(0.011)	
Log of Fundeb per student	-0.055	0.946	0.132**	1.141	0.175***	1.191
	(0.083)		(0.049)		(0.018)	
Random effects						
Level: State (variance)	-		-		0.202	
Level: Municipality (variance)	0.168		0.311		0.268	
Level: School (variance)	0.088		0.175		0.122	
Level: Class (variance)	0.242		0.160		0.220	
Level: Student (variance)	Binomial		Binomial		Binomial	

Source: Own elaboration (2026).
Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Model 2 corresponds to the full specification.

Changing the dependent variable from test scores to school success reveals patterns that were not evident in the linear models. In the proficiency analyses, girls showed a disadvantage in Mathematics, yet in the binary model they have a significant advantage over boys in all contexts, with nearly twice the odds of regular progression (OR = 1.994 in MA; 1.972 in MG; and 1.868 in Brazil). This suggests that, although they obtain lower scores in Mathematics tests, they employ behaviors or strategies that result in higher promotion rates. Similarly, the relationship between school infrastructure and student progression differs: whereas infrastructure positively affected learning in the linear models, in the school success model the odds ratios indicate a non-significant or even negative effect (OR = 0.918 in MA and 0.937 in Brazil), suggesting that schools with better physical infrastructure may apply stricter evaluation criteria, decoupling infrastructure quality from automatic student progression. The analysis also highlights the detrimental effect of child labor, which drastically reduces the probability of school success (odds decrease by about 40%, reflected in OR = 0.571 in MA; 0.628 in MG; and

0.613 in Brazil), emphasizing the conflict between early entry into the labor market and educational achievement.

In MA, school success appears to be strongly conditioned by the classroom level, whereas in MG inequalities in progression take on a more institutional character, concentrating at the municipality and school levels. At the national level, despite regional heterogeneity, within-school factors remain central. The stability of the variance estimates compared with the models without controls (Table C7) reinforces the reliability of the estimated random effects. With regard to social equity, the odds ratios indicate that SES exerts a positive direct effect on the probability of success in all contexts, increasing the odds by between 26% and 30% (OR = 1.303 in MA; 1.264 in MG; and 1.292 in Brazil), even after controlling for prior performance. In addition, the interaction terms between SES and prior proficiency show a consistently positive association (for example, the SES × Mathematics interaction has OR = 1.116 in MA and 1.097 in Brazil), indicating that the relationship between prior and current performance is stronger among students from more advantaged socioeconomic backgrounds. This effect is particularly pronounced in Mathematics—especially in MA and at the national level—reinforcing the interpretation that this subject functions as a more selective mechanism capable of amplifying cumulative trajectories of advantage and disadvantage.

In terms of differential effectiveness, the variance decomposition highlights a substantive contribution from intermediate institutional levels, particularly schools and classrooms, even in a model that extensively controls individual and contextual characteristics. The magnitude of variance at these levels indicates that opportunities for success are not distributed homogeneously within the educational system, pointing to the existence of schools and classrooms that are systematically more effective in converting prior skills into academic progression. Taken together, these findings reinforce the notion that reducing educational inequalities depends not only on strengthening initial competencies but also on the capacity of schools to provide equally effective learning conditions for students from diverse social backgrounds.

In summary, incorporating the binary multilevel model enhances the understanding of educational inequalities by showing that the factors associated with learning and school success do not entirely overlap. While prior performance and SES remain the most robust factors across both outcomes, the school success model reveals additional mechanisms of selection and progression that are not captured by standardized test scores. The presence of positive moderating effects between SES and prior proficiency, particularly in Math, indicates that social inequalities not only persist but are amplified throughout the educational trajectory, creating a pattern of cumulative advantage in student progression. At the same time, the variance decomposition shows that, although the largest share of heterogeneity is concentrated at the student and classroom levels, institutional levels – especially schools and municipalities – continue to play a relevant role in shaping opportunities for success. These findings underscore that policies aimed at reducing educational inequalities should combine early interventions on learning with institutional strategies capable of mitigating the effects of social background on school progression. In this way, the joint analysis of proficiency and school success provides a more comprehensive view of effectiveness and equity dynamics within the Brazilian educational system, establishing the empirical foundation for the discussions and implications presented in the following section.

4.6 DISCUSSION

This study aimed to estimate the contribution of schools to student academic progress in Brazilian elementary education, analyzing differential school effectiveness and social equity through a comparative approach across Brazil, MA, and MG. Overall, the results indicate that school effectiveness exists but operates within clear structural limits: most of the variability in student performance is concentrated within schools, particularly at the student and classroom levels, while differences between institutions remain relatively modest after controlling for individual and contextual characteristics. The persistence of educational trajectories and the limited capacity of schools to mitigate the effects of SES suggest that learning inequalities (continuous model) and student school success (binary model) largely stem from the social composition of students, with variations associated with the institutional context.

Regarding the fixed effects, the results align consistently with literature and allow for an interpretation of how educational inequalities are structured throughout elementary education. The fact that prior performance emerges as the key predictor of subsequent learning is consistent with the school effectiveness literature (Ferrão, 2018, 2022a, 2022b; Ferrão & Couto, 2013; Jerrim et al., 2019; Muñoz-Chereau & Thomas, 2016; Taut et al., 2016), supporting Hypothesis 1. Complementarily, the persistence of the effects of SES, student labor participation, parental education, and living with guardians reinforces the interpretation that schools operate within deep structural inequalities, in which social origin is a key factor

(Coleman, 1968; Alves et al., 2007; Bonamino et al., 2010; Soares et al., 2017; Breen et al., 2009).

The gender inequalities and racial penalties observed reinforce evidence already established in the literature on school effectiveness and educational inequalities (Mortimore et al., 1988; Sammons et al., 1997; Alves et al., 2007; Soares et al., 2017; Ferrão, 2022a, 2022b). In the proficiency models, girls show lower performance in Math, while racial disparities tend to be more pronounced in contexts marked by higher school stratification and internal segregation, such as MG, in line with patterns identified at the national level by Ferrão (2022a). However, the results of the binary model (Table 4.5) indicate that girls' lower test performance does not automatically translate into reduced school success. The higher likelihood of school success observed across all contexts reinforces the distinction between cognitive achievement and regular student progression, aligning with Gomes-Neto et al. (1997) and Ferrão (2022b), who conceptualize school performance as a multidimensional process encompassing both learning outcomes and students' trajectories within the educational system. Thus, patterns observed in standardized tests are reflected, albeit partially, in the probability of school success.

Moreover, in the continuous model, grade repetition has a negative effect on student performance, regardless of the context analyzed, consistent with evidence documented by Gomes-Neto et al. (1997) and Ferrão (2022b). This aligns with the literature on educational inequalities in contrasting institutional contexts and extends to the Brazilian case the argument of Breen et al. (2009), originally formulated based on European evidence, that educational development tends to redefine, rather than eliminate, the mechanisms reproducing inequalities.

Previous studies emphasize the importance of school infrastructure as a foundational element of school effectiveness (Soares et al., 2017; Blaskó et al., 2022; Guilherme et al., 2024). Consistent with this literature, the results of this study show that infrastructure-related effects are stronger in contexts of higher vulnerability, such as MA, underscoring the central role of basic material conditions in shaping educational outcomes. However, when considering the availability of pedagogical infrastructure and its association with Reading proficiency, the results reveal negative effects on student performance in this same context, suggesting challenges in resource allocation and effective use (Gomes-Neto et al., 1997; Singh, 2020; Thieme et al., 2016), as well as limitations in institutional and managerial capacity to convert available pedagogical resources into learning gains (Hanushek, 1971; Muñoz & Prem, 2024). Furthermore, although the school effectiveness literature highlights the importance of leadership (e.g., Valle & Lillejord, 2023), the findings suggest that the limited set of observable characteristics of school leaders used in this study does not translate into substantive effects on student learning in the contexts analyzed. Considering a broader set of leadership characteristics, complemented with qualitative evidence and the perspectives of school actors, could provide a more complete understanding of leadership processes and their influence on educational outcomes.

It is also observed that the marginal effect of state and municipal schools is lower than that of federal schools across all contexts analyzed, supporting Ferrão's (2022a) findings that the practices of these schools should be investigated to inform strategies applicable to other types of schools. As for financial resources, their effectiveness depends on the institutional and local context. In MA, the resource showed no statistically significant association. These results suggest that simply providing financial resources does not automatically improve student learning, as their effectiveness relies on the institutional context and schools' managerial capacity (Silveira et al., 2017; Muñoz & Prem, 2024).

As previously shown in Section 4.5, in the continuous model, most of the variance in student performance is concentrated within schools and is attributed to the student level, while average differences between schools, municipalities, and states are relatively small, corroborating national and international findings (Chetty et al., 2014; Ferrão, 2022a; Hill et al., 2011; Koedel et al., 2015). Heterogeneity between classrooms is limited but may partly reflect grouping practices or class composition (Cervini, 2009, 2016; Torche, 2005; Troncoso, 2019; Timmermans & Thomas, 2015), and it remains stable regardless of the level of development of the educational system. Thus, individual and family factors continue to account for a large share of the observed heterogeneity (Alves et al., 2007; Bernstein, 1970; Breen et al., 2009; Coleman et al., 1966; Ferrão, 2022a, 2022b; Jencks et al., 1972; Mortimore et al., 1988; Sammons et al., 1997; Shavit & Blossfeld, 1993; Soares et al., 2017; Strand, 2010), which also supports Hypothesis 1. In the model in which the student's probability of school success constitutes the dependent variable, this pattern persists, with the combined variance at the classroom and school levels exceeding that of the other levels.

The analysis of random slopes (continuous model) indicates that schools' capacity to mitigate the effects of prior proficiency and SES on student performance exhibits little heterogeneity across institutions (Ferrão, 2022a; Strand, 2010, 2016; Sammons et al., 1997; Mortimore et al., 1988), addressing Hypothesis 2, which investigated whether greater heterogeneity would emerge between contexts. From a theoretical perspective, this finding is consistent with Ferrer-Esteban's (2016) argument that educational inequalities are driven less by substantial differences in school effectiveness and more by processes of social sorting that concentrate socioeconomically disadvantaged students in specific schools or classrooms.

In the binary model, social equity manifests more prominently, as SES has a direct positive effect on the probability of school success and moderates the return to prior performance, particularly in Math, corroborating the effect identified by Ferrão (2022b). The variance decomposition reinforces the author's findings, showing that schools display greater variance compared to other levels, as does the classroom level, in all contexts analyzed. Schools exhibit higher differential effectiveness than in the proficiency model, indicating that their capacity to contribute to students' school success and promote social equity is more evident in this type of analysis. It is also observed that the variance coefficient at the school level is lower in Maranhão and higher in Minas Gerais and at the national level, partially corroborating Hypothesis 2. These differences refer exclusively to school-level heterogeneity and do not contradict the evidence of limited average school effects across contexts. Overall, these results do not support Hypothesis 3, as the patterns of differential effectiveness and social equity observed in standardized test outcomes do not replicate in students' school success, indicating that the effects detected in test scores are not fully reflected in this alternative indicator of academic performance.

The results should be interpreted in light of limitations associated with the dataset and the observational nature of the analysis, which apply to both estimated models. The models do not account for socioemotional, psychosocial, or motivational factors because these data are unavailable in the analyzed datasets, but prior studies have demonstrated their influence on student learning outcomes (e.g., Clark et al., 2021; Durlak et al., 2011; Greenberg et al., 2003; Loeb et al., 2019; Morando & Platt, 2022). This implies that part of the individual heterogeneity associated with educational trajectories may not be fully captured by the models. Furthermore, the data refers to the 2011-2015 period, so caution is warranted when generalizing the results, as recent changes in educational policies, socioeconomic conditions, and the organization of the school system may have altered the mechanisms identified.

From a methodological standpoint, there are specific limitations associated with each modeling strategy. In the continuous model, the analysis of performance is restricted to students who did not repeat a grade and progressed regularly from the 5th to the 9th grade, which may introduce selection bias, particularly in contexts with high repetition rates. This limitation is mitigated in the binary model, which explicitly incorporates grade repetition when analyzing the probability of regular school progression, allowing for a more comprehensive assessment of the factors associated with educational trajectories. Moreover, in this model, the *glmer* function from R's *lme4* package does not provide standard errors for random effects.

Additionally, in the binary model, random slopes for prior performance and SES at the school level were not included due to issues of instability and convergence, preventing an assessment of whether the returns to these characteristics vary across institutions. Finally, the estimation of models with multiple levels, parameters, and interactions can increase the likelihood of Type I errors (rejecting the null hypothesis when it is true), requiring caution when interpreting statistically significant effects, particularly those of smaller magnitude.

4.7 CONCLUSION

This study analyzes school effectiveness in Brazil from a perspective focused on differential effectiveness and social equity, addressing the methodological limitations identified in the systematic review of the previous essay. The adoption of a five-level multilevel model (student, classroom, school, municipality, and state), combined with the incorporation of school success indicators related to student progression, allowed for a more precise variance decomposition than that offered by the traditional two-level models prevalent in regional literature. Based on longitudinal data from 2011 to 2015, the analysis contrasts the national scenario with two institutionally distinct state contexts – MA and MG – and shows that, although school effectiveness exists, it operates within clear structural limits and is insufficient on its own to offset deep socioeconomic inequalities.

The empirical evidence from the matched 2011-2015 cohort suggests that, within the period analyzed, learning inequalities remain persistent across contexts with different levels of educational system development, providing a basis for evaluating the formulated hypotheses. Hypothesis 1 was supported, showing that the effects of prior performance, socioeconomic background, gender, and race persist consistently across all contexts analyzed, reaffirming that these inequalities are largely independent of the institutional development of the education system. Hypothesis 2 was only partially supported: in the continuous model, schools' capacity to mitigate the effects of prior performance and SES exhibits little heterogeneity across institutions, contrary to expectations, whereas in the binary model, differential effectiveness and social equity are more pronounced. In this regard, Hypothesis 3 was not supported, as the patterns of differential effectiveness and social equity observed in standardized test outcomes do not replicate in students' probability of school success.

By comparing different institutional contexts within the same country, this study advances the Brazilian literature on school effectiveness by showing that, although patterns in the reproduction of inequality are similar, the underlying mechanisms vary across contexts. Contrary to the empirically tested expectation that schools in more structured systems might act as equalizers, the findings indicate that both in MG and MA the education system exhibits low differential effectiveness and limited social equity, suggesting the presence of a common "effectiveness threshold" among Brazilian schools. This apparent stability may reflect structural features of the Brazilian schooling system, including limited pedagogical autonomy, curricular and evaluative homogeneity, and restricted school discretion. Furthermore, modest gains associated with concentrations of advantaged students are offset by larger losses in schools serving disadvantaged populations, so that the low variation in school effects reflects institutional constraints rather than an absence of school impact. The explicit incorporation of the classroom level – still little explored in the Ibero-American literature – revealed substantial within-school heterogeneity.

The evidence indicates that analyses based solely on school averages or school rankings tend to obscure this within-school heterogeneity. By incorporating models of proficiency and school success, the study shows that different dimensions of student outcomes are influenced by partially distinct sets of factors, indicating that those associated with learning do not always coincide with those associated with school success. This finding highlights relevant paradoxes, such as the coexistence of lower performance by girls in Math with higher probabilities of school success, suggesting that unidimensional assessments provide an incomplete view of how the education system functions. One possible explanation is that the Brazilian school system may prioritize continuity, attendance, proper conduct, and adherence to institutional rules alongside purely cognitive achievement, which could help account for this paradox.

Regarding differential effectiveness, the results show that, in aggregate terms, schools differ little in their capacity to mitigate the effects of prior performance, although relatively greater variability is observed in MA. Continuous models show modest variation across schools, indicating weak social equity: schools differ little in their ability to offset the effects of socioeconomic status, providing little evidence of equity, which would require actively reducing learning inequalities. In contrast, the binary success model highlights a positive interaction between SES and prior proficiency, suggesting that more advantaged students are better able to convert their initial knowledge into subsequent educational success than disadvantaged peers with similar prior achievement. In both models, most of the observed variability occurs at the student and classroom levels rather than between schools, reinforcing that the reproduction of inequalities predominantly takes place within schools.

Regarding public policy implications, the results indicate that interventions focused solely on the school as an average unit are insufficient considering the high variance concentrated at the student and classroom levels. These findings point to the need for policies aimed at reducing educational inequalities that focus on schools' internal organization and student-level factors, considering heterogeneous learning trajectories and socioeconomic conditions. Comparisons across contexts further indicate that, although the reproduction of inequalities is structural, its mechanisms are strongly conditioned by the level of institutional vulnerability: in contexts such as MA, improvements in physical infrastructure are necessary, but their effects depend on the institutional capacity to allocate and utilize resources effectively; in more structured contexts, such as MG and at the national level, racial penalties and gender inequalities persist, requiring targeted and compensatory policies beyond the mere provision of material inputs. The positive effect of SES on the returns to prior performance underscores the need for redistributive policies capable of ensuring that students from less advantaged backgrounds can translate their initial skills into effective learning and school progression. Finally, the joint analysis of proficiency and school success suggests that policies and assessment systems should incorporate multiple dimensions of performance, avoiding an exclusive focus on standardized tests or progression indicators that may obscure internal inequalities and limit the potential for more effective interventions.

By highlighting the structural limits of school effectiveness and the central role of within-school processes in reproducing inequalities, this study contributes to the debate on school effectiveness in Brazil. The findings show that, while schools do make a difference, their ability to counteract socioeconomic disparities is limited, with most inequality reproduced through processes within schools rather than between them. These insights provide a useful analytical framework for researchers examining educational systems in other unequal contexts.

The temporal scope of the study constrains the analysis of recent changes in the educational system, opening opportunities for future studies using more recent datasets, particularly to examine the effects of the COVID-19 pandemic on student performance and progression, as well as the effects of implementing the National Common Curricular Base (BNCC) in Brazil. Although the data covers 2011–2015, the structural nature of the intra-school inequality mechanisms identified suggests persistence. From a methodological perspective, future investigations could also employ alternative computational approaches to overcome the limitations of the R lme4 package regarding the estimation of standard errors for random effects in binary models, thereby enabling a formal statistical assessment of the precision of variance estimates. Furthermore, subsequent research could further investigate the within-school processes associated with classroom-level variance, including pedagogical practices and class assignment criteria, and expand the scope to comparative international analyses.

5. GENERAL CONCLUSION

This thesis aimed to critically analyze the use of Value-Added (VA) models in educational evaluation, integrating conceptual foundations, empirical evidence from the literature, and an empirical application in the Brazilian context, focusing on indicators of differential effectiveness and social equity. Starting from the recognition that educational inequalities persist even after expanded school access, the study sought to understand the extent to which, and under what conditions, schools contribute to students' academic progress, considering their prior trajectories and the institutional contexts in which they are embedded.

The analytical trajectory of the thesis was structured around three complementary and progressive essays. The first essay aimed to systematize the conceptual and methodological development of VA models in educational research, from their origins in the economics of education to their incorporation into studies of school effectiveness. The essay highlighted both the potential of VA models to isolate the relative contribution of schools and the critical debates surrounding the validity, stability, and interpretation of their estimates, particularly when applied in contexts marked by social and institutional inequality. Furthermore, the study identified key limitations in literature, such as the scarcity of research in developing countries, the limited exploration of differential effectiveness and social equity, and the insufficient incorporation of longitudinal data, contextual and socioemotional variables, and causal approaches.

The second essay advanced this agenda by conducting a systematic review of the empirical application of VA models in Ibero-American countries. The results revealed a still nascent body of research, concentrated in a few countries, characterized by high methodological heterogeneity and the predominance of simplified multilevel models, typically limited to two levels (student and school). The review also highlighted the centrality of standardized tests as the most used measure of performance, as well as the limited attention given to differential effectiveness, social equity, and within-school heterogeneity. These findings underscore that, despite recent progress, regional literature still offers a partial view of how educational systems function and their role in reproducing or mitigating inequalities, pointing to the need for more sophisticated empirical approaches that are sensitive to institutional and regional contexts and capable of capturing the hierarchical complexity of education systems.

The third essay directly addresses this gap by conducting an empirical investigation of school effectiveness in Brazil, explicitly examining differential effectiveness and social equity across contrasting institutional contexts. The use of five-level multilevel models (student, classroom, school, municipality, and state), combined with longitudinal data and the incorporation of a complementary indicator of school success, helped address recurring limitations in the regional literature by allowing a detailed decomposition of variance across educational levels. The results of the continuous models show that differential school effectiveness is limited, with most of the variance in achievement concentrated at the student and classroom levels and only modest variation between schools in their capacity to mitigate the effects of prior achievement and SES. Evidence of social equity is therefore weak, as schools differ little in their ability to reduce learning inequalities. In the binary model of school success, school-level heterogeneity is more pronounced, and SES positively moderates the return to prior achievement, revealing a cumulative advantage pattern rather than compensatory effects. Taken together, these findings indicate that, while schools do matter, their capacity to compensate for socioeconomic inequalities is limited, and the reproduction of inequalities occurs predominantly through intra-school processes, rather than through differentiation between schools.

The main contributions of this thesis can be summarized across three dimensions. Conceptually, the work reinforces that VA models are not neutral measurement tools, but analytical constructions whose results depend on their underlying assumptions, methodological choices, and the institutional context in which they are applied. Methodologically, the thesis advances the field by demonstrating the importance of more complex hierarchical structures, modeling differential effectiveness and social equity, and incorporating multiple dimensions of school performance. The comparison between proficiency and school success indicators shows that different dimensions of performance respond to partially distinct sets of factors, revealing important paradoxes – for example, girls' lower performance in Math coupled with higher probabilities of regular progression. This suggests that evaluations based solely on average standardized test scores provide an incomplete picture of the functioning of the educational system and may obscure substantive internal inequalities. Empirically, the results indicate that, in the Brazilian context, most of the variance in student performance and school success is concentrated at the student and classroom levels rather than between schools, suggesting that the reproduction of educational inequalities occurs predominantly through intra-school processes.

From a public policy perspective, the thesis calls for caution in using VA models as high-stakes accountability tools, particularly in contexts characterized by high social and institutional inequality, such as in Ibero-American countries. While useful as diagnostic instruments, the findings indicate that policies focused solely on comparing school averages or performance rankings tend to obscure intra-school inequalities and overestimate the capacity of schools to compensate for students' disadvantaged backgrounds. The evidence accumulated across the three essays highlights the need for more integrated policies that combine context-sensitive assessments, interventions at the student and classroom levels, strengthened school leadership, and actions that promote not only redistribution but also the allocative efficiency of resources, ensuring that socially disadvantaged students can translate input and initial skills into effective learning. In this sense, the use of VA models should be understood as part of a broader system of educational assessment and management, aimed not at punishment but at reducing inequalities and promoting equality.

The findings of this thesis should be interpreted considering limitations that manifest across different analytical levels. From a conceptual perspective, as discussed in the first essay, VA models rely on strong assumptions, particularly regarding selection, stability, and the measurement of achievement, which constrain their interpretation as strictly causal measures. At the empirical-comparative level, the systematic review highlighted the geographical and methodological concentration of Ibero-American research and the predominance of models with simplified hierarchical structures, limiting the generalizability of regional findings and pointing to structural gaps in data availability and the institutionalization of educational assessment. Finally, at the empirical level, the analyses are based on observational data from the Brazilian public school system, predating recent transformations such as the COVID-19 pandemic and the implementation of the BNCC, and do not incorporate non-cognitive dimensions, school climate, or family mechanisms, which limits the explanation of the processes underlying observed inequalities and the unexplained variance at the classroom level. Although the data cover 2011–2015, the identified intra-school inequality mechanisms appear persistent.

In this regard, the thesis highlights important avenues for future research, including the incorporation of more recent data, the expanded use of non-cognitive measures, the investigation of the role of school leadership in promoting VA, a deeper analysis of intra-school mechanisms associated with classroom-level variance, and the development of international comparisons among Ibero-American countries. The use of school success as a complementary metric of effectiveness also appears promising for more comprehensive evaluations in contexts characterized by high rates of grade repetition and school dropout.

In sum, this thesis addresses its central research question by demonstrating that VA models, from their conceptual foundations to their applications in Ibero-American contexts, constitute analytical tools capable of diagnosing schools' differential effectiveness and social equity under specific methodological and institutional conditions. By integrating conceptual reflection, critical systematization of literature, and applied empirical analysis, the thesis contributes to the debate on evaluating school effectiveness. Beyond advocating for the use of VA models as measurement instruments, the study highlights their potential as analytical tools to understand the limits and possibilities of school action in promoting learning and reproducing or mitigating educational inequalities, provided they are employed in a critical, context-sensitive, and methodologically rigorous manner.

REFERENCES

Ackerman, B. P., Brown, E. D., & Izard, C. E. (2004). The relations between contextual risk, earned income, and the school adjustment of children from economically disadvantaged families. *Developmental Psychology*, *40*(2), 204.

AERA Statement on use of Value-Added Models (VAM) for the evaluation of educators and educator preparation programs. (2015). *Educational Researcher*, *44*(8), 448–452. https://doi.org/10.3102/0013189X15618385

Agostinelli, F., Doepke, M., Sorrenti, G., & Zilibotti, F. (2020). *When the great equalizer shuts down: Schools, peers, and parents in pandemic times*. CEPR Discussion Papers.

Albernaz, Â., Ferreira, F. H. G., & Franco, C. (2002). Qualidade e eqüidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico (PPE)*. https://repositorio.ipea.gov.br/handle/11058/4340

Alves, F., Ortigão, I., & Franco, C. (2007). Origem social e risco de repetência: Interação raça-capital econômico. *Cadernos de Pesquisa*, *37*(130), 161–180. https://doi.org/10.1590/S0100-15742007000100008

Alves, M. T. G., & Franco, C. (2008). A pesquisa em eficácia escolar no Brasil: Evidências sobre o efeito das escolas e fatores associados à eficácia escolar. In N. Brooke & J. F. Soares (Eds.), *Pesquisa em eficácia escolar: Origem e trajetórias* (Vol. 22, pp. 482–500). Editora UFMG.

Alves, M. T. G., & Ferrrão, M. E. (2020). Uma década da Prova Brasil: Evolução do desempenho e da aprovação. *Estudos Em Avaliação Educacional*, *30*(75), 688. https://doi.org/10.18222/eae.v0ix.6298

Amrein-Beardsley, A. (2023). A Florida teacher wrongfully terminated? Alleged (mis)uses of Value-Added Model (VAM) estimates for high-stakes teacher evaluation decisions. *Educational Assessment, Evaluation and Accountability*, *35*(3), 313–352. https://doi.org/10.1007/s11092-023-09411-8

Amrein-Beardsley, A., Collins, C., Polasky, S. A., & Sloat, E. F. (2013). Value-Added Model (VAM) research for educational policy: Framing the issue. *Education Policy Analysis Archives*, *21*, 4. https://doi.org/10.14507/epaa.v21n4.2013

Amrein-Beardsley, A., Pivovarova, M., & Geiger, T. J. (2016). Value-added models: What the experts say. *Phi Delta Kappan*, *98*(2), 35–40. https://doi.org/10.1177/0031721716671904

Amrein-Beardsley, A., Ryan Lavery, M., Holloway, J., Pivovarova, M., & L. Hahs-Vaughn, D. (2023). Evaluating the validity evidence surrounding use of value-added models to evaluate teachers: A systematic review. *Education Policy Analysis Archives*, *31*. https://doi.org/10.14507/epaa.31.8201

Andrabi, T., Das, J., Ijaz Khwaja, A., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, *3*(3), 29–54. https://doi.org/10.1257/app.3.3.29

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018). The impact of school SES on student achievement: Evidence from U.S. statewide achievement data. *Educational Evaluation and Policy Analysis*, *40*(4), 613–630. https://doi.org/10.3102/0162373718787917

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. *EPI Briefing Paper*, *278*.

Baker, R. S. J. d, Merceron, A., & Pavlik, P. I. (2010). Proceedings of the international conference on Educational Data Mining (EDM). *International Working Group on Educational Data Mining*.

Barbosa, M. E. F., & Fernandes, C. (2001). A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4ª série. *Promoção, Ciclos e Avaliação Educacional. Porto Alegre: ArtMed*, 155-172.

**Barrios-Fernández, A., & Riudavets-Barcons, M. (2024). Teacher value-added and gender gaps in educational outcomes. *Economics of Education Review*, *100*, 102541. https://doi.org/10.1016/j.econedurev.2024.102541

**Bartholo, T. L., Koslinski, M. C., da Costa, M., & Barcellos, T. (2019). What do children know upon entry to pre-school in Rio de Janeiro? *Ensaio*, *28*(107), 292–313. https://doi.org/10.1590/S0104-40362019002702071

Bartholo, T. L., Koslinski, M. C., De Andrade, F. M., & De Castro, D. L. (2020). Segregação escolar e desigualdades educacionais no início da escolarização no Brasil. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, *18*(4). https://doi.org/10.15366/reice2020.18.4.003

**Bartholo, T. L., Koslinski, M. C., Tymms, P., & Castro, D. L. (2023). Learning loss and learning inequality during the Covid-19 pandemic. *Ensaio*, *31*(119). https://doi.org/10.1590/S0104-40362022003003776

*Bartolucci, F., Pennoni, F., & Vittadini, G. (2011). Assessment of school performance through a multilevel latent Markov Rasch model. *Journal of Educational and Behavioral Statistics*, *36*(4), 491–522. https://doi.org/10.3102/1076998610381396

*Bartolucci, F., Pennoni, F., & Vittadini, G. (2023). A causal latent transition model with multivariate outcomes and unobserved heterogeneity: Application to human capital development. *Journal of Educational and Behavioral Statistics*, *48*(4), 387–419. https://doi.org/10.3102/10769986221150033

Bernstein, B. (1970). *Education cannot compensate for society*. *38*(New Society), 344–347.

**Berthelon, M., Contreras, D., Kruger, D., & Palma, M. (2020). Harsh parenting during early childhood and child development. *Economics & Human Biology*, *36*. https://doi.org/10.1016/j.ehb.2019.100831

**Bertoni, E., Elacqua, G., Mendez, C., & Santos, H. (2024). Teacher selection instruments and teacher value-added: Evidence from Peru. *Educational Evaluation and Policy Analysis*, *46*(1), 53–81. https://doi.org/10.3102/01623737221149417

Blaskó, Z., Costa, P. da, & Schnepf, S. V. (2022). Learning losses and educational inequalities in Europe: Mapping the potential consequences of the COVID-19 crisis. *Journal of European Social Policy*, *32*(4), 361–375. https://doi.org/10.1177/09589287221091687

Bodner, Todd E. What Improves with Increased Missing Data Imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, v. 15, n. 4, p. 651–675, 22 out. 2008.

Bonamino, A., Alves, F., Franco, C., & Cazelli, S. (2010). Os efeitos das diferentes formas de capital no desempenho escolar: Um estudo à luz de Bourdieu e de Coleman. *Revista Brasileira de Educação*, *15*(45), 487–499. https://doi.org/10.1590/S1413-24782010000300007

Branco, J. R., Carr, M., Fidalgo, C., Bigotte de Almeida, M. E., Santos, V., Murphy, E., & Ní Fhloinn, E. (2015). Mathematics diagnostic testing in engineering: An international comparison between ireland and Portugal. *European Journal of Engineering Education*, *40*(5), 546–556.

Braun, H. (2013). Value-added modeling and the power of magical thinking. *Ensaio: Avaliação e Políticas Públicas Em Educação*, *21*(78), 115–130. https://doi.org/10.1590/S0104-40362013000100007

Braunack-Mayer, A. J., Street, J. M., Tooher, R., Feng, X., & Scharling-Gamba, K. (2020). Student and staff perspectives on the use of Big Data in the tertiary education sector: A scoping review and reflection on the ethical issues. *Review of Educational Research*, *90*(6), 788–823. https://doi.org/10.3102/0034654320960213

Brauw, A. de, Gilligan, D. O., Hoddinott, J., & Roy, S. (2015). The impact of Bolsa Família on schooling. *World Development*, *70*, 303–316. https://doi.org/10.1016/j.worlddev.2015.02.001

Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2009). Nonpersistent inequality in educational attainment: Evidence from eight European countries. *American Journal of Sociology*, *114*(5), 1475–1521.

**Brooke, N., da Silva Fernandes, N. S., de Miranda, S. P. H., & Soares, T. M. (2014). Modeling of the growth of learning in the early years with longitudinal data of GERES research. *Educacao e Pesquisa*, *40*(1), 77–94. https://doi.org/10.1590/S1517-97022014000100006

Bryk, A. S., & Weisberg, H. I. (1976). Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics*, *1*(2), 127. https://doi.org/10.2307/1164980

*Bryk, A. S., & Weisberg, H. I. (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, *84*(5).

Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility*, *44*, 54–67.

Burger, K. (2019). The socio-spatial dimension of educational inequality: A comparative european analysis. *Studies in Educational Evaluation*, *62*, 171–186. https://doi.org/10.1016/j.stueduc.2019.03.009

Cao, W., Gnana Sanga Mithra, S., & B R, A. (2024). Unraveling the factors shaping academic success: A structural equation modeling approach for college students. *Heliyon*, *10*(4), e25775. https://doi.org/10.1016/j.heliyon.2024.e25775

Casillas, B. B. (2006). Murillo Torrecilla, F. Javier (2005). La investigación sobre eficacia escolar. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*. Barcelona: Octaedro. 320 págs.

**Castro-Morera, M., de Miguel, C. R., & López-Martín, E. (2009). Growth basic form at value-added models: Supressing regression effect. *Revista de Educacion*, *348*, 111–136.

**Castro-Morera, M., García-Medina, A. M., Pedroza, L. H., & Niebla, J. (2015). High and low value-added schools. Differential profiles of secondary schools in Baja California (Mexico). *Education Policy Analysis Archives*, *23*. https://doi.org/10.14507/epaa.v23.1917

Cervini, R. (2016). Comparando la inequidad en los logros escolares de la educación primaria y secundaria de Argentina: Un estudio multinivel comparativo. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, *7*(1). https://doi.org/10.15366/reice2009.7.1.001

Cervini, R. A. (2009). Class, school, municipal, and state effects on mathematics achievement in Argentina: A multilevel analysis. *School Effectiveness and School Improvement*, *20*(3), 319–340. https://doi.org/10.1080/09243450802664404

César, C. C., & Soares, J. F. (2001). Desigualdades acadêmicas induzidas pelo contexto escolar. *Revista Brasileira de Estudos de População*, *18*(1/2), 97–100.

Chan, T. J., & Dai, M. (2023). Factors influencing academic achievement of university students. *Journal of Communication, Language and Culture*, *3*(2), 14–26. https://doi.org/10.33093/jclc.2023.3.2.2

Charalambous, C. Y., Kyriakides, E., Kyriakides, L., & Tsangaridou, N. (2019). Are teachers consistently effective across subject matters? Revisiting the issue of differential teacher effectiveness. *School Effectiveness and School Improvement*, *30*(4), 353–379.

Charalambous, E., Kyriakides, L., & Creemers, B. P. M. (2018). Promoting quality and equity in socially disadvantaged schools: A group- randomisation study. *Studies in Educational Evaluation*, *57*, 42–52. https://doi.org/10.1016/j.stueduc.2016.06.001

*Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). 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

*Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Clark, A. E., D'Ambrosio, C., & Barazzetta, M. (2021). Childhood circumstances and young adulthood outcomes: The role of mothers' financial problems. *Health Economics*, *30*(2), 342–357. https://doi.org/10.1002/hec.4194

Closs, L., Weiss, D., Pilla, B. S., Saraiva, M., & Salerno, G. (2024). *Fatores que influenciam a eficácia escolar: Uma revisão sistemática da literatura no brasil no período de 2014 a 2023*. SciELO Preprints. https://doi.org/10.1590/SciELOPreprints.9893

Coleman, J. S. (1968). Equality of educational opportunity. *Integrated Education*, *6*(5), 19–28.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Constituição da República Federativa do Brasil (1988). Promulgada em 5 de outubro de 1988. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Creemers, B. P. (2005). Educational effectiveness the development of the field. *Keynote Address Presented at the First International Conference on School Effectiveness and School Improvement*.

Cumpston, M., Li, T., Page, M. J., Chandler, J., Welch, V. A., Higgins, J. P., & Thomas, J. (2019). Updated guidance for trusted systematic reviews: A new edition of the Cochrane Handbook for systematic reviews of interventions. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.ED000142

De Fraine, B., Van Damme, J., & Onghena, P. (2002). Accountability of schools and teachers: What should be taken into account? *European Educational Research Journal*, *1*(3), 403–428. https://doi.org/10.2304/eerj.2002.1.3.2

Decreto nº 6.425 (2008). Dispõe sobre o censo anual da educação. https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6425.htm

Decreto nº 9.432, art. 5° (2018). Regulamenta a Política Nacional de Avaliação e Exames da Educação Básica. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/d9432.htm

Dumay, X., Coe, R., & Anumendem, D. N. (2014). Stability over time of different methods of estimating school performance. *School Effectiveness and School Improvement*, *25*(1), 64–82.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15-24.

**Eigbiremolen, G. O., Ogbuabor, J. E., & Nwambe, C. S. (2020). Estimating private school effects for school children in Peru: Evidence from individual-level panel data. *Journal of International Development*, *32*(2), 131–149. https://doi.org/10.1002/jid.3438

Emenda Constitucional No. 108 (2020). Altera a Constituição Federal para estabelecer critérios de distribuição da cota municipal do Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS), para disciplinar a disponibilização de dados contábeis pelos entes federados, para tratar do planejamento na ordem social e para dispor sobre o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb); altera o Ato das Disposições Constitucionais Transitórias; e dá outras providências. https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc108.htm

Emenda Constitucional nº 53, art. 2°, No. 53 (2006). Dá nova redação aos arts. 7º, 23, 30, 206, 208, 211 e 212 da Constituição Federal e ao art. 60 do Ato das Disposições Constitucionais Transitórias. https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc53.htm

*Emslander, V., Levy, J., Scherer, R., & Fischbach, A. (2022). Value-added scores show limited stability over time in primary school. *PLOS ONE*, *17*(12), e0279255. https://doi.org/10.1371/journal.pone.0279255

Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research*, *87*(1), 35–70. https://doi.org/10.3102/0034654316637199

Fernandes, R., & Gremaud, A. P. (2009). Qualidade da educação: Avaliação, indicadores e metas. *Educação Básica No Brasil: Construindo o País Do Futuro. Rio de Janeiro: Elsevier*, *1*, 213–238.

**Ferrão, M. E. (2009). Sensitivity of value-added model specifications: Measuring socio-economic status. *Revista de Educacion*, *348*, 137–152.

Ferrão, M. E. (2012). Avaliação educacional e modelos de valor acrescentado: Tópicos de reflexão. *Educação & Sociedade*, *33*(119), 455–469. https://doi.org/10.1590/S0101-73302012000200007

**Ferrão, M. E. (2014). School effectiveness research findings in the portuguese speaking countries: Brazil and Portugal. *Educational Research for Policy and Practice*, *13*(1), 3–24. https://doi.org/10.1007/s10671-013-9151-7

Ferrão, M. E. (2018). Educational statistics and public policy: On value added models. *Educacao e Sociedade*, *39*(142), 19–38. https://doi.org/10.1590/ES0101-73302017176230

**Ferrão, M. E. (2022a). Estudo longitudinal sobre eficácia diferencial e equidade social no Brasil. *REICE - Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *20*(1). https://doi.org/10.15366/reice2022.20.1.004

Ferrão, M. E. (2022b). The evaluation of students' progression in lower secondary education in Brazil: Exploring the path for equity. *Studies in Educational Evaluation*, *75*. https://doi.org/10.1016/j.stueduc.2022.101220

Ferrão, M. E. (2023). Differential effect of university entrance scores on graduates' performance: The case of degree completion on time in Portugal. *Assessment & Evaluation in Higher Education*, *48*(1), 95–106. https://doi.org/10.1080/02602938.2022.2052799

Ferrão, M. E. (2025). Valor añadido en la educación: Mapeando las contribuciones de Portugal y España. *Revista de Educación*, *409*, 441–476. https://doi.org/10.4438/1988-592X-RE-2025-409-700

Ferrão, M. E., & Alves, M. T. G. (2023). The relationship between students' socio-demographics and the probability of grade repetition in Brazilian primary education: Is it decreasing over time? *Large-Scale Assessments in Education*, *11*(1), 12. https://doi.org/10.1186/s40536-023-00160-0

**Ferrão, M. E., & Couto, A. (2013). Value-added indicator and topics on consistency and stability: An application to Brazil. *Ensaio*, *21*(78), 131–164. https://doi.org/10.1590/S0104-40362013000100008

**Ferrão, M. E., & Couto, A. P. (2014). The use of a school value-added model for educational improvement: A case study from the Portuguese primary education system. *School Effectiveness and School Improvement*, *25*(1), 174–190. https://doi.org/10.1080/09243453.2013.785436

**Ferrão, M. E., & Goldstein, H. (2009). Adjusting for measurement error in the value added model: Evidence from Portugal. *Quality and Quantity*, *43*(6), 951–963. https://doi.org/10.1007/s11135-008-9171-1

Ferrão, M. E., Barros, G. T. de F., Bof, A. M., & Oliveira, A. S. de. (2018). Estudo longitudinal sobre eficácia educacional no Brasil: Comparação entre resultados contextualizados e valor acrescentado. *Dados*, *61*, 265–300.

Ferrão, M. E., Beltrão, K. I., Fernandes, C., Santos, D., Suárez, M., & do Couto Andrade, A. (2001). O SAEB-Sistema Nacional de Avaliação da Educação Básica: Objetivos, características e contribuições na investigação da escola eficaz. *Revista Brasileira de Estudos de População*, *18*(1/2), 111–130.

Ferrer-Esteban, G. (2016). Trade-Off between Effectiveness and Equity? An Analysis of Social Sorting between Classrooms and between Schools. *Comparative Education Review*, *60*(1), 151–183. https://doi.org/10.1086/684490

**Filho, G. A. S. (2019). Efeito da formação docente sobre proficiência no início do ensino fundamental. *Revista Brasileira de Economia*, *73*(3), 385–411. https://doi.org/10.5935/0034-7140.20190018

García-Pérez, J. I., Hidalgo-Hidalgo, M., & Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, *46*(12), 1373–1392. https://doi.org/10.1080/00036846.2013.872761

Goldstein, H. (2003). Multilevel modelling of educational data. In D. Courgeau (Ed.), *Methodology and Epistemology of Multilevel Analysis* (pp. 25–42). Springer Netherlands. https://doi.org/10.1007/978-1-4020-4675-9_2

Goldstein, H. (2010). *Multilevel statistical models*. John Wiley & Sons.

** Gomes-Neto, J. B., Hanushek, E. A., Leite, R. H., & Frota-Bezzera, R. C. (1997). Health and schooling: Evidence and policy implications for developing countries. *Economics of Education Review*, *16*(3), 271–282. https://doi.org/10.1016/S0272-7757(97)00001-0

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, *58*(6–7), 466–474. https://doi.org/10.1037/0003-066X.58.6-7.466

Guilherme, A. A., de Araujo, J. M., Silva, L., & de Oliveira Brito, R. (2024). Two ‘Brazils’: Socioeconomic status and education performance in Brazil. *International Journal of Educational Research*, *123*, 102287. https://doi.org/10.1016/j.ijer.2023.102287

Gustafsson, J.-E., Hansen, K. Y., & Rosén, M. (2011). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. *Timss and Pirls*, *4*, 181–287.

Hancock, K. J., Lawrence, D., Shepherd, C. C. J., Mitrou, F., & Zubrick, S. R. (2017). Associations between school absence and academic achievement: Do socioeconomics matter? *British Educational Research Journal*, *43*(3), 415–440. https://doi.org/10.1002/berj.3267

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, *61*(2), 280–288.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Hanushek, E. A. (2019). Testing, accountability, and the american economy. *The Annals of the American Academy of Political and Social Science*, *683*(1), 110–128. https://doi.org/10.1177/0002716219841299

*Hanushek, E. A., & Rivkin, S. G. (2010). The quality and distribution of teachers under the No Child Left Behind Act. *Journal of Economic Perspectives*, *24*(3), 133–150. https://doi.org/10.1257/jep.24.3.133

*Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, *4*(1), 131–157. https://doi.org/10.1146/annurev-economics-080511-111001

*Hanushek, E. A., & Woessmann, L. (2011). How much do educational outcomes matter in OECD countries? *Economic Policy*, *26*(67), 427–491. https://doi.org/10.1111/j.1468-0327.2011.00265.x

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2022). Long-run trends in the US SES—Achievement gap. *Education Finance and Policy*, *17*(4), 608–640.

*Hanushek, E. A., Rivkin, S. G., & Schiman, J. C. (2016). Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review*, *55*, 132–148. https://doi.org/10.1016/j.econedurev.2016.08.004

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005). The market for teacher quality. *NBER - National Bureau of Economic Research*, *11154*. https://doi.org/10.3386/w11154

*Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). An exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, *23*(3), 5. https://doi.org/10.2307/1177220

**Hernández, L., & Olariaga, L. (2009). Analysis of dimensionality at value-added models: A study of math test using non parametric methods based on item response theory. *Revista de Educacion*, *348*, 175–194.

Higgins, J., & Thomas, J. (2023). Cochrane Handbook for systematic reviews of interventions. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.4*. https://training.cochrane.org/handbook/current

Hoffman, D. M. (2009). Reflecting on social emotional learning: A critical perspective on trends in the United States. *Review of Educational Research*, *79*(2), 533–556. https://doi.org/10.3102/0034654308325184

Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

IBGE. (2018). *Instituto Brasileiro de Geografia e Estatística* (No. Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC)) [Dataset]. https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/20979-norte-e-nordeste-convivem-com-restricoes-no-acesso-a-saneamento-basico

*Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), 801–825. https://doi.org/10.1146/annurev-economics-080213-040845

Jackson, M. (2013). *Determined to succeed?: Performance versus choice in educational attainment*. Stanford University Press.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). Inequality: A reassessment of the effect of the family and schooling in America. In D. B. Grusky (Ed.), *Social stratification: Class, race, and gender in sociological perspective* (4th edition). Routledge. https://doi.org/10.4324/9780429494642

Jensen, B., Giorguli Saucedo, S., & Hernández Padilla, E. (2018). International migration and the academic performance of mexican adolescents. *International Migration Review*, *52*(2), 559–596. https://doi.org/10.1111/imre.12307

**Jerrim, J., Lopez–Agudo, L. A., & Marcenaro-Gutiérrez, O. D. (2019). The relationship between homework and the academic progress of children in Spain during compulsory elementary education: A twin fixed-effects approach. *British Educational Research Journal*, *45*(5), 1021–1049. https://doi.org/10.1002/berj.3549

*Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631. https://doi.org/10.1016/j.econedurev.2007.05.005

Karino, C. A., & Laros, J. A. (2017). Estudos brasileiros sobre eficácia escolar: Uma revisão de literatura. *Examen: Política, Gestão e Avaliação Da Educação*, *1*(1), 32–32.

Kersting, N. B., Chen, M.-K., & Stigler, J. W. (2013). Value-added teacher estimates as art of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Education Policy Analysis Archives*, *21*(7). https://doi.org/10.14507/epaa.v21n7.2013

Koedel, C., Mihaly, K., & Rockoff, 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

*Konstantopoulos, S., & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal*, *48*(2), 361–386. https://doi.org/10.3102/0002831210382888

Krüger, N. (2020). Efectos compañero en contextos escolares altamente segregados. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, *18*(4), 171–196. https://doi.org/10.15366/reice2020.18.4.007

Kuzmanic, D., Meneses, F., Valenzuela, J. P., Rodríguez, P., & Claro, S. (2024). Inequalities in learning loss during the COVID-19 pandemic in Chile: The significance of school effectiveness. *School Effectiveness and School Improvement*, *35*(2), 95–115. https://doi.org/10.1080/09243453.2024.2328016

Kyriakides, L. (2004). Differential school effectiveness in relation to sex and social cslass: Some implications for policy evaluation. *Educational Research and Evaluation*, *10*(2), 141–161. https://doi.org/10.1076/edre.10.2.141.27907

Kyriakides, L., Charalambous, E., Creemers, H. P. M. (Bert), & Dimosthenous, A. (2019). Improving quality and equity in schools in socially disadvantaged areas. *Educational Research*, *61*(3), 274–301. https://doi.org/10.1080/00131881.2019.1642121

Larrea, V. A., Rojas, I. B., & Hernández, L. L. (2024). Mejora escolar y coeducación en centros de secundaria de la comunidad autónoma vasca: Implicaciones para la orientación educativa. *REOP - Revista Española de Orientación y Psicopedagogía*, *35*(1), 45–62. https://doi.org/10.5944/reop.vol.35.num.1.2024.40754

Lavy, V. (2016). What makes an effective teacher? Quasi-experimental evidence. *CESifo Economic Studies*, *62*(1), 88–125. https://doi.org/10.1093/cesifo/ifv001

Lei No. 14.113 (2020). Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), de que trata o art. 212-A da Constituição Federal; revoga dispositivos da Lei nº 11.494, de 20 de junho de 2007; e dá outras providências. Brasília, DF. https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14113.htm

*Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2019). Methodological issues in value-added modeling: An international review from 26 countries. *Educational Assessment, Evaluation and Accountability*, *31*(3), 257–287. https://doi.org/10.1007/s11092-019-09303-w

*Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2023). How sensitive are the evaluations of a school's effectiveness to the selection of covariates in the applied value-added model? *Educational Assessment, Evaluation and Accountability*, *35*(1), 129–164. https://doi.org/10.1007/s11092-022-09386-y

*Levy, J., Mussack, D., Brunner, M., Keller, U., Cardoso-Leite, P., & Fischbach, A. (2020). Contrasting classical and machine learning approaches in the estimation of value-added scores in large-scale educational data. *Frontiers in Psychology*, *11*, 2190. https://doi.org/10.3389/fpsyg.2020.02190

Libâneo, J. C. (2016). Políticas educacionais no Brasil: Desfiguramento da escola e do conhecimento escolar. *Cadernos de Pesquisa*, *46*(159), 38–62. https://doi.org/10.1590/198053143572

Loeb, S., Christian, M. S., Hough, H., Meyer, R. H., Rice, A. B., & West, M. R. (2019). School differences in social-emotional learning gains: Findings from the first large-scale panel survey of students. *Journal of Educational and Behavioral Statistics*, *44*(5), 507–542. https://doi.org/10.3102/1076998619845162

**López-Martin, E., Kuosmanen, T., & Gaviria, J. (2014). Linear and nonlinear growth models for value-added assessment: An application to Spanish primary and secondary schools' progress in reading comprehension. *Educational Assessment Evaluation and Accountability*, *26*(4), 361–391. https://doi.org/10.1007/s11092-014-9194-1

Ma, Y., Hou, X., Huang, J., Wang, W., Li, Y., Zhou, X., & Du, X. (2018). Educational inequality and achievement disparity: An empirical study of migrant children in China. *Children and Youth Services Review*, *87*, 145–153. https://doi.org/10.1016/j.childyouth.2018.02.026

Marder, M. (2012). Measuring teacher quality with value-added modeling. *Kappa Delta Pi Record*, *48*(4), 156–161.

*Marder, M., David, B., & Hamrock, C. (2020). Math and science outcomes for students of teachers from standard and alternative pathways in Texas. *Education Policy Analysis Archives*, *28*, 27. https://doi.org/10.14507/epaa.28.4863

*Marder, M., Horn, C., Stephens, S., & Rhodes, A. (2022). Student learning and teacher retention for graduates of Texas Noyce programs. *Education Policy Analysis Archives*, *30*. https://doi.org/10.14507/epaa.30.7254

Markowitz, J. B., Hebbeler, K., Larson, J. C., Cooper, J. A., & Edmister, P. (1991). Using value-added analysis to examine short-term effects of early intervention. *Journal of Early Intervention*, *15*(4), 377–389. https://doi.org/10.1177/105381519101500406

Marks, G. N. (2017). Is SES really that important for educational outcomes in Australia? A review and some recent evidence. *The Australian Educational Researcher*, *44*(2), 191–211.

Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, *12*(02), 105–128.

Martin, A. J. (2011). Personal Best (PB) Approaches to Academic Development: Implications for Motivation and Assessment. *Educational Practice and Theory*, *33*(1), 93–99.

Martínez, J. F. (2012). Consequences of omitting the classroom in multilevel models of schooling: An illustration using opportunity to learn and reading achievement. *School Effectiveness and School Improvement*, *23*(3), 305–326. https://doi.org/10.1080/09243453.2012.678864

Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, *66*, 100875. https://doi.org/10.1016/j.stueduc.2020.100875

McLaughlin, M. J. (2010). Evolving interpretations of educational equity and students with disabilities. *Exceptional Children*, *76*(3), 265–278. https://doi.org/10.1177/001440291007600302

Medeiros, M., & Oliveira, L. F. B. D. (2014). Desigualdades regionais em educação: Potencial de convergência. *Sociedade e Estado*, *29*(2), 561–585. https://doi.org/10.1590/S0102-69922014000200012

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Slot, P. L., Broekhuizen, M., Leseman, P., Broekhuisen, M. (2015). A review of research on the effects of Early Childhood Education and Care (ECEC) upon child development. *EU CARE project*.

Miranda, M. G., & Santos, S. V. (2012). Propostas de tempo integral: A que se destina a ampliação do tempo escolar? *Perspectiva*, *30*(3), 1073–1098. https://doi.org/10.5007/2175-795X.2012v30n3p1073

Morais, A. M. D. P., Betarelli, A. A., Vasconcelos, S. P., & Vieira, M. D. T. (2021). Universities, socioeconomic standards and inclusion policies: Assessing the effects on the performance of Brazilian undergraduates. *Studies in Educational Evaluation*, *70*, 100996. https://doi.org/10.1016/j.stueduc.2021.100996

Morando, G., & Platt, L. (2022). The impact of centre-based childcare on non-cognitive skills of young children. *Economica*, *89*(356), 908–946. https://doi.org/10.1111/ecca.12440

*Morera, M., Martínez Abad, F., Lizasoain Hernández, L., & Joaristi Olariaga, L. M. (2009). Selección de escuelas de alta y baja eficacia en Baja California (México). *Revista Electrónica de Investigación Educativa*, *19*(2), 38–53. https://doi.org/10.24320/redie.2017.19.2.960

Morganstein, D., & Wasserstein, R. (2014). ASA Statement on value-added models. *Statistics and Public Policy*, *1*(1), 108–110. https://doi.org/10.1080/2330443X.2014.956906

Mortimore, P. (2014). *The road to improvement*. Routledge.

Mortimore, P., Sammons, P., Stoll, L., Ecob, R., & Lewis, D. (1988). The effects of school membership on pupils' educational outcomes. *Research Papers in Education*, *3*(1), 3–26.

Mortimore, P., & Whitty, G. (2000). *Can school improvement overcome the effects of disadvantage?* (Revised ed). Institute of Education.

**Muñoz, P., & Prem, M. (2024). Managers' productivity and recruitment in the public sector. *American Economic Journal: Economic Policy*, *16*(4), 223–253. https://doi.org/10.1257/pol.20220732

**Muñoz-Chereau, B. (2019). Exploring gender gap and school differential effects in mathematics in chilean primary schools. *School Effectiveness and School Improvement*, *30*(2), 83–103. https://doi.org/10.1080/09243453.2018.1503604

**Muñoz-Chereau, B., & Thomas, S. M. (2016). Educational effectiveness in chilean secondary education: Comparing different "value added" approaches to evaluate schools. *Assessment in Education: Principles, Policy & Practice*, *23*(1), 26–52. https://doi.org/10.1080/0969594X.2015.1066307

Murillo, F. J., & Martínez-Garrido, C. (2019). Una Mirada a la investigación educativa en América Latina a partir de sus artículos. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *17*(2), 5. https://doi.org/10.15366/reice2019.17.2.001

Muthen, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*(4).

Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, *18*(23). https://doi.org/10.14507/epaa.v18n23.2010

Nurse, L., & Melhuish, E. (2021). Comparative perspectives on educational inequalities in Europe: An overview of the old and emergent inequalities from a bottom-up perspective. *Contemporary Social Science*, *16*(4), 417–431. https://doi.org/10.1080/21582041.2021.1948095

Nuttall, D. L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, *13*(7), 769–776. https://doi.org/10.1016/0883-0355(89)90027-X

*Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257. https://doi.org/10.3102/01623737026003237

OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. OECD. https://doi.org/10.1787/9789264073234-en

**Ortega, L., Malmberg, L., & Sammons, P. (2018). School effects on chilean children's achievement growth in language and mathematics: An accelerated growth curve model. *School Effectiveness and School Improvement*, *29*(2), 308–337. https://doi.org/10.1080/09243453.2018.1443945

Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. *Sociology of Education*, *91*(1), 1–27.

**Page, G. L., San Martín, E., Orellana, J., & González, J. (2017). Exploring complete school effectiveness via quantile value added. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *180*(1), 315–340. https://doi.org/10.1111/rssa.12195

**Page, G. L., San Martín, E., Torres-Irribarra, D. T., & Bellegem, S. V. (2024). Temporally dynamic, cohort-varying value-added models. *Psychometrika*, *89*(3), 1074–1103. https://doi.org/10.1007/s11336-024-09979-0

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, *19*(1), 21–49. https://doi.org/10.1080/09243450801936845

Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations.

American Educational Research Journal, *51*(2), 328–362. https://doi.org/10.3102/0002831213508299

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, *5*(4), 371–385. https://doi.org/10.1002/jrsm.1123

Pickett, K. (2014). Addressing health inequalities through greater social equality at a local level: Implement a living wage policy. *IF YOU COULD DO ONE THING...*, 22.

PNUD, IPEA, & FJP. (2019). *Radar IDHM: evolução do IDHM e de seus índices componentes no período de 2012 a 2017*. Ipea.

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451. https://doi.org/10.1016/j.lindif.2010.05.001

Qi, D., & Wu, Y. (2019). Comparing the extent and levels of child poverty by the income and multidimensional deprivation approach in China. *Child Indicators Research*, *12*(2), 627–645. https://doi.org/10.1007/s12187-018-9544-5

Reardon, S. F., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, *40*(1), 199–218.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 197–230. https://doi.org/10.1080/09243453.2014.885450

Reynolds, D., Teddlie, C. (2002). An introduction to school effectiveness research. In *The international handbook of school effectiveness research* (pp. 17–39). Routledge.

Ribeiro, C. (2018). Contínuo racial, mobilidade social e “embranquecimento.” *Revista Brasileira de Ciências Sociais*, *32*(95), 1–25. https://doi.org/10.17666/329503/2017

*Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, *4*(4), 537–571. https://doi.org/10.1162/edfp.2009.4.4.537

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, *125*(1), 175–214. https://doi.org/10.1162/qjec.2010.125.1.175

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*(1), 103–116. https://doi.org/10.3102/10769986029001103

Rutter, M. (Ed.). (1980). *Fifteen thousand hours: Secondary schools and their effects on children* (Reprint). Open Books Publ.

Ryu, H., Helfand, S. M., & Moreira, R. B. (2020). Starting early and staying longer: The effects of a Brazilian primary schooling reform on student performance. *World Development*, *130*, 104924. https://doi.org/10.1016/j.worlddev.2020.104924

Sammons, P. (2007). School effectiveness and equity: Making connections. *Reading: CfBT*, 1–22.

Sammons, P., Hillman, J., & Mortimore, P. (1995). Key characteristics of effective schools: A review of school effectiveness research. *A Review of School Effectiveness Research*.

Sammons, P., Thomas, S., Mortimore, P., & Walker, A. (1997). *Forging links: Effective schools and effective departments*. P. Chapman Pub.

**Santelices, M. V., Galleguillos, P., González, J., & Taut, S. (2015). A study about teacher quality in Chile: The role of the context where teachers work and value-added measures. *Psykhe*, *24*(1). https://doi.org/10.7764/psykhe.23.2.673

**Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2017). Two teacher quality measures and the role of context: Evidence from Chile. *Educational Assessment, Evaluation and Accountability*, *29*(2), 111–146. https://doi.org/10.1007/s11092-016-9247-8

Saunders, L. (1999). A brief history of educational "value added": How did we get to where we are? *School Effectiveness and School Improvement*, *10*(2), 233–256. https://doi.org/10.1076/sesi.10.2.233.3507

Saunders, L. (2000). Understanding schools' use of 'value added' data: The psychology and sociology of numbers. *Research Papers in Education*, *15*(3), 241–258. https://doi.org/10.1080/02671520050128740

Scheerens, J., & Bosker, R. (1998). *The foundations of educational effectiveness*. Pergamon Press.

Schneeweis, N. (2011). Educational institutions and the integration of migrants. *Journal of Population Economics*, *24*(4), 1281–1308.

Shavit, Y., & Blossfeld, H.-P. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries. Social Inequality Series.* ERIC.

Silveira, I. M., Lima, J. E., Teixeira, E. C., & Silva, R. G. (2017). Avaliação do efeito do Fundeb sobre o desempenho dos alunos do ensino médio no Brasil. *Pesquisa e Planejamento Econômico*.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*(4), 323–355. https://doi.org/10.3102/10769986023004323

**Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, *18*(4), 1770–1813. https://doi.org/10.1093/jeea/jvz033

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

**Soares, T. M., de Bonamino, A., Brooke, N., & da Silva Fernandes, N. S. (2017). Value-added models for the measurement of the effectiveness of GERES schools. *Ensaio*, *25*(94), 59–89. https://doi.org/10.1590/S0104-40362017000100003

*Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97–118. https://doi.org/10.1257/jep.24.3.97

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, *21*(3), 289–314. https://doi.org/10.1080/09243451003732651

Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal*, *37*(2), 197–229. https://doi.org/10.1080/01411920903540664

Strand, S. (2012). The White British-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal*, *38*(1), 75–101. https://doi.org/10.1080/01411926.2010.526702

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education*, *4*(2), 107–144. https://doi.org/10.1002/rev3.3054

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The effective provision of pre-school education (EPPE) project: Findings from pre-school to end of key Stage1 research brief*.

**Taut, S., Valencia, E., Palacios, D., Santelices, M. V., Jiménez, D., & Manzi, J. (2016). Teacher performance and student learning: Linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, *23*(1), 53–74. https://doi.org/10.1080/0969594X.2014.961406

Teddlie, C., & Reynolds, D. (2001). Countering the Critics: Responses to Recent Criticisms of School Effectiveness Research. *School Effectiveness and School Improvement*, *12*(1), 41–82. https://doi.org/10.1076/sesi.12.1.41.3458

Teddlie, C., Reynolds, D., & Sammons, P. (2002). The Methodology and Scientific Properties of School Effectiveness Research. In D. Reynolds & C. Teddlie (Eds.), *The International Handbook of School Effectiveness Research*. Taylor and Francis.

**Thieme, C., Prior, D., Tortosa-Ausina, E., & Gempp, R. (2016). Value added, educational accountability approaches and their effects on schools' rankings: Evidence from Chile. *European Journal of Operational Research*, *253*(2), 456–471. https://doi.org/10.1016/j.ejor.2016.01.023

Thomas, S., Peng, W. J., & Gray, J. (2007). Modelling patterns of improvement over time: Value added trends in Enlish secondary school performance across ten cohorts. *Oxford Review of Education*, *33*(3), 261–295. https://doi.org/10.1080/03054980701366116

Tilak, J. B. G. (2002). Education and poverty. *Journal of Human Development*, *3*(2), 191–207. https://doi.org/10.1080/14649880220147301

Timmermans, A. C., & Thomas, S. M. (2015). The Impact of student composition on schools' value-added performance: A comparison of seven empirical studies. *School Effectiveness and School Improvement*, *26*(3), 487–498.

Todd, Petra E.; Wolpin, Kenneth I. On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, v. 113, n. 485, p. F3–F33, 1 fev. 2003.

Todd, Petra E.; Wolpin, Kenneth I. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, v. 1, n. 1, p. 91–136, dez. 2007.

Torche, F. (2005). Privatization reform and inequality of educational Opportunity: The Case of Chile. *Sociology of Education*, *78*(4), 316–343. https://doi.org/10.1177/003804070507800403

**Torres, R. (2018). Tackling inequality? Teacher effects and the socioeconomic gap in educational achievement. Evidence from Chile. *School Effectiveness and School Improvement*, *29*(3), 383–417. https://doi.org/10.1080/09243453.2018.1443143

**Troncoso, P. (2019). A two-fold indicator of school performance and the cost of ignoring it. *International Journal of Educational Research*, *95*, 153–167. https://doi.org/10.1016/j.ijer.2019.02.002

**Troncoso, P., Pampaka, M., & Olsen, W. (2016). Beyond traditional school value-added models: A multilevel analysis of complex school effects in Chile. *School Effectiveness and School Improvement*, *27*(3), 293–314. https://doi.org/10.1080/09243453.2015.1084010

Twersky, S. E. (2022). Do state laws reduce uptake of Medicaid/CHIP by U.S. citizen children in immigrant families: Evaluating evidence for a chilling effect. *International Journal for Equity in Health*, *21*(1), 50. https://doi.org/10.1186/s12939-022-01651-2

UNESCO. (2024). *Relatório GEM 2024/5: Liderança na educação: liderar para a aprendizagem* (p. 8). GEM Report UNESCO. https://doi.org/10.54676/XTNJ1654

Valle, R., & Lillejord, S. (2023). A new regime of understanding. School leadership in Norwegian education policy (1990–2017). *Nordic Journal of Studies in Educational Policy*, *9*(2), 144–160. https://doi.org/10.1080/20020317.2023.2208258

**Vivanco, E. F. (2013). When change matters identifying score gains school determinants in Mexico: An intra-cohort value-added approach. *Economia Mexicana, Nueva Epoca*, *22*(4), 361–405.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

White, Ian R.; Royston, Patrick; Wood, Angela M. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, v. 30, n. 4, p. 377–399, 20 fev. 2011.

APPENDIX A

Table A1 – Representativeness of academic journals citing Hanushek (1971) or Bryk & Weisberg (1976)

Journal	**Number of publications citing**		
	Hanushek (1971)	**Bryk & Weisberg (1976)**	**SJR[a]**
American Economic Review	2	-	22.344
American Educational Research Journal	1	-	2.232
Annual Review of Economics	2	-	8.925
Econometrica	1	-	17.701
Economic Policy	1	-	2.229
Economics of Education Review	2	-	1.059
Education Policy Analysis Archives	3	-	0.313
Educational Assessment, Evaluation and Accountability	2	-	1.611
Educational Evaluation and Policy Analysis	1	-	1.782
Educational Researcher	1	-	3.956
Frontiers in Psychology	1	-	0.800
Journal of Economic Perspectives	2	-	7.496
Journal of Human Resources	1	-	4.524
Early Childhood Research Quarterly	-	1	1.569
Estudios Sobre Educacion	-	1	0.305
Evaluation Review	-	1	0.297
Journal of Applied Behavior Analysis	-	1	1.218
Journal of Early Intervention	-	1	0.556
Journal of Educational and Behavioral Statistics	-	2	1.336
Psychological Bulletin	-	1	8.412
Psychological Methods	-	2	4.235
Revista Electronica de Investigacion Educativa	-	1	0.232

Source: Own elaboration (2026)

Note: [a] The SCImago Journal Rank (SJR) of a journal indicates the average weighted citations received per document over the last three years, with higher values indicating greater prestige and allowing the comparison of impact between journals within the same field.

Table A2 – Characterization of selected articles citing Hanushek (1971)

(continues)

Author, year, and cluster [(a)]	Objective	Methodology	Main results
Hedges et al. (1994) [K]	Examine whether Hanushek's claims about the lack of correlation between educational resources, particularly per-pupil expenditure (PPE), and student performance hold up under more advanced statistical analysis.	Meta-analysis	Re-analysis of the data shows that financial investment in schools is significantly associated with better educational outcomes, challenging the notion that money does not influence educational quality. These findings underscore the importance of school resources for students' academic success.
Nye et al. (2004) [K]	Estimate the impact of teaching staff on student performance using data from a four-year experiment where both teachers and students were randomly assigned to classes.	Multilevel linear model	Teacher effects were larger in mathematics than in reading, with experience having a significant effect only in reading at grade 2 and mathematics at grade 3. The variability of the effects was higher in schools of low socio-economic status than in those of high status.
Rivkin et al. (2005) [H]	The objective is to analyze and disentangle the impact of schools and teachers on academic performance, with a particular focus on potential challenges related to omitted or poorly measured variables, as well as the selection of students and schools.	Fixed effects model	The study highlights the importance of teacher and school quality for student achievement, with significant gains in the first year of a teacher's experience and modest effects of class size. It emphasizes the need for effective hiring, firing and mentoring practices, especially for disadvantaged students.
Kane et al. (2008) [R]	Determine whether initial certification status has a significant impact on student achievement and whether changes in teacher composition can improve this achievement.	Fixed effects model	Initial teacher certification has a small effect on student achievement, but there are large and persistent differences in effectiveness between teachers with the same experience and certification. Teacher turnover has a negative impact on student achievement, suggesting that these teachers would need to become slightly more effective each year to compensate.
Hanushek & Rivkin (2010) [H]	Analyze the impact of the No Child Left Behind Act on schools, focusing on teacher incentives and the distribution of qualified teachers, and how these changes affect teacher quality and student achievement.	Fixed effects model	Accountability systems such as No Child Left Behind influence teacher distribution and teacher composition, highlighting the importance of principals' ability to differentiate between less and more effective teachers. Teacher turnover can be positive in high-poverty schools, and alternative accountability models, such as across-the-board salary increases, performance-based salary variation, and VA models that assess student growth, can significantly alter the distribution of teacher quality.
Staiger & Rockoff (2010) [R]	Evaluate the impact of using available performance information to select teachers more effectively, with the goal of improving academic achievement.	Multilevel linear model	The study highlights the significant heterogeneity in teacher effects on student achievement and the importance of pre-hiring information in selecting effective teachers. More reliable measures, such as classroom observations, can improve teacher selection. Strategies involving a broad sample of candidates, even if only a small percentage are stable, have been shown to be effective in improving student achievement, highlighting the importance of rigorous selection and retention processes.
Hanushek & Woessmann (2011) [H]	Analyze the importance of educational outcomes in OECD countries, focusing on the impact of students' cognitive skills, as measured by tests such as PISA, on long-term economic performance, and how educational policies influence these skills and economic growth.	Endogenous and neoclassical growth model	Cognitive skills, as measured by tests such as PISA, have a significant impact on long-term economic performance in OECD countries, over and above the influence of economic institutions. The study shows that improvements in education that increase students' cognitive skills can yield substantial economic gains. Different model structures have been explored, showing that the differences between them are not so significant.

(continues)

Author, year, and cluster [(a)]	**Objective**	**Methodology**	**Main results**
Konstantopoulos & Chung (2011) [K]	Examine the persistence of early teacher effects on student achievement over seven years, using data from Project STAR and the Lasting Benefits Study.	Multilevel linear model	The results show that effective teachers have a positive and persistent impact on students' performance in mathematics, reading and science up to grade 6. Having effective teachers continuously in the early grades brings significant benefits, almost half a standard deviation, highlighting the importance of identifying and studying the characteristics of these teachers.
Hanushek & Rivkin (2012) [H]	Analyse the distribution of teacher quality and its implications for education policy.	Review and fixed effects model	The study highlights that teacher quality has a significant impact on students' academic progress, independent of other factors. The introduction of VA measures for teachers in evaluations and staffing decisions is central, sparking debate and controversy in education and highlighting the importance of literature reviews on the topic to inform effective education policy.
Jackson et al. (2014) [R]	To examine how teacher effectiveness affects student outcomes and how this information can guide educational policy and practice to improve the quality of education.	Review	The study highlights the variation in teacher effectiveness and recommends the use of VA-based assessments for selection, mentoring and compensation. Replacing 5 to 10% of the least effective teachers with those of average effectiveness can significantly improve student outcomes. Informing principals about teacher effectiveness can increase the turnover of low-performing teachers and modestly improve student outcomes.
Chetty et al. (2014a) [R]	Investigate whether teacher impact on student test scores, also known as VA, is an effective measure of teacher quality.	Fixed effects model	VA models that control for students' prior performance provide unbiased estimates of the impact of teachers on student achievement. By using unobserved parental characteristics and a quasi-experimental design based on teacher turnover, the study validates the effectiveness of VA policies and confirms the ability of VA models to predict the impact of teachers on student performance while minimizing prediction bias.
Chetty et al. (2014b) [R]	Assess whether VA measures provide unbiased estimates of the impact of teachers on student achievement, and examine the validity of VA as an indicator of teacher quality.	Fixed effects model	VA measures are effective in assessing the impact of teachers on student achievement. Students with high VA teachers are more likely to attend college, earn higher salaries, and avoid teenage pregnancy. Replacing a low VA teacher with an average VA teacher can significantly increase students' lifetime earnings. However, using VA to evaluate teachers can lead to undesirable behaviors, such as teaching to the test.
Hanushek et al. (2016) [H]	Analyze how teacher turnover affects teacher quality and student achievement over time, and identify factors that influence these effects to inform education policy and school management practice.	Fixed effects model	Teacher turnover has a negative impact on low-performing schools. Teachers who change campuses or districts tend to be more effective, especially those early in their careers. In addition, teachers who leave the school system are less effective, and measuring effectiveness in the year before the change helps to identify persistent differences in teacher productivity.

(conclusion)

Author, year, and cluster [(a)]	**Objective**	**Methodology**	**Main results**
Levy et al. (2019) [L]	Identify key factors for improving VA modelling, particularly with regard to the inclusion of covariates, adjustments and model diagnostics, as well as clarity and transparency in the presentation of results.	Systematic review	The majority of studies in the review used linear regression or multilevel models that included prior achievement as a covariate, but few included non-cognitive predictors of achievement. Many studies did not use statistical adjustments or model diagnostics, suggesting that research could be improved by including covariates, adjustments and diagnostics, as well as increasing transparency in reporting.
Marder et al. (2020) [M]	Compare the performance and retention of mathematics and science teachers prepared through traditional university programs and alternative certification routes in Texas, focusing on their retention since 2008 and their impact on Algebra 1 and Biology assessments from 2012 to 2018.	Multilevel linear model	Teachers with college degrees tend to stay in the classroom longer, and students learn more in Algebra I and Biology when taught by these teachers. There has also been a shift in teacher experience, with more professionals having 10 to 20 years' experience. This underscores the significant impact of teacher training and experience on student outcomes in high school science and mathematics.
Marder et al. (2022) [M]	Evaluate the National Science Foundation's Robert Noyce Teacher Scholarship Program, analyzing the retention of scholarship recipients, their teaching locations, and their impact on student learning, to make recommendations for expanding and improving the program.	Multilevel linear model	Key findings from the study of the Robert Noyce Teacher Scholarship Program indicate that Noyce Scholars were more likely to teach marginalised students and that their students achieved higher VA scores in mathematics than students of other teachers in the same schools. However, Noyce Scholars did not stay in teaching as long as other STEM teachers from their universities and were more likely to leave schools with low-income students. In addition, the Noyce program is currently operating well below the scale needed to address the national STEM teacher shortage.
Emslander et al. (2022) [L]	Examine whether VA scores have shown stability over time and whether there has been variation across different outcome domains.	Multilevel linear model	The results show moderate stability in primary school VA scores, with 34-38% of schools maintaining consistent scores over two years, and moderate correlations suggest that this stability is similar across subjects. VA scores should not be the sole measure of educational accountability, and complementary sources of data are recommended for making appropriate educational decisions. However, VA models may be useful for identifying effective teaching practices, especially in heterogeneous populations where educational inequalities are a major concern.
Levy et al. (2020) [L]	Compare the results of classical statistical methods and machine learning techniques in estimating school VA, and assess when and how these techniques can complement or replace traditional methods.	Multilevel linear model, regressions, and machine learning	The main findings show that multilevel models outperformed linear regression, polynomial models and various machine learning models in estimating VA scores. Despite the high correlation between VA scores generated by different types of models, the percentage differences between multilevel models and other models were not trivial, which could have significant implications for schools depending on the model used. In addition, the inclusion of covariates related to students' prior academic achievement and socio-demographic context improves the accuracy of VA estimates.
Levy et al. (2023) [L]	Analyze how the selection of covariates affects the assessment of school effectiveness in VA models for mathematics and language, providing insights for educational evaluation.	Multilevel linear model	The study highlights the importance of including additional covariates, such as students' socio-demographic and motivational characteristics, in VA models to accurately assess school effectiveness. The variation in results shows that careful selection of covariates is crucial to avoid misleading conclusions about school effectiveness.

Source: Own elaboration (2026)

Note: [(a)] Each cluster is identified by the term in brackets, where: "K" is the cluster led by Konstantopoulos; "H" is the cluster led by Hanushek; "R" is the cluster led by Rockoff; "L" is the cluster led by Levy; and "M" is the cluster led by Marder.

Table A3 - Characterization of selected articles citing Bryk & Weisberg (1976)

Author, year, and cluster [(a)]	Objective	Methodology	Main results
Bryk & Weisberg (1977) [Br]	Evaluate the adequacy of statistical techniques used to adjust for post-test comparisons in experimental programs, particularly in non-randomized contexts, using pre- and post-test data.	Linear Growth Model	Adjustment strategies are sensitive to the nature of individual growth, recommending additional pre-intervention growth data and interrupted time-series designs. Moreover, the results emphasize the need for new statistical techniques that align with natural growth processes and highlight the complexity of analyzing changes in evolving contexts.
Markowitz et al. (1991) [C]	Assess the short-term effects of early intervention on children who have received special education services before the age of 5.	Value Added Analysis	The results indicate that children made progress exceeding what would be expected from maturation alone, with documented improvements in various domains for children with speech and language impairments and multiple disabilities, and cognitive benefits for those with speech and language impairments.
Muthen & Curran (1997) [M]	Demonstrate the generalizability of latent variable modeling for analyzing individual differences in development over time, with a focus on randomized intervention studies.	Random Coefficients Growth Model	The study proposes an approach to estimate treatment effects as deviations from normative development within the control group. It also suggests a methodology for calculating the statistical power necessary to detect these effects, considering factors such as sample size, number of measurements, and effect sizes, with illustrative examples using both real and synthetic data.
Morera et al. (2009) [Mo]	Describe the key aspects of VA models in education, which enable the simultaneous assessment of students' current learning levels and the rate of growth in their performance resulting from school interventions.	Multilevel linear model	Key findings of the study emphasize the importance of measuring student progress over time, the effectiveness of multivariate models for this analysis, and the need for longitudinal assessments with at least three measures. It also underscores the significance of VA models in educational evaluation for guiding school improvement.
Bartolucci et al. (2011) [B]	Propose a multilevel extension of the Latent Markov Rasch (LM Rasch) model to analyze longitudinal binary test data administered repeatedly to mathematics students in public and private schools in the Lombardy region of Italy, accounting for school-level and individual-level covariates.	Latent Markov Rasch Multilevel Model	The study demonstrates that mathematics achievement is influenced by both individual factors and school quality. High-quality schools can mitigate the impact of a disadvantaged family background. The analysis identified levels of mathematics achievement consistent with international standards, highlighting the need to consider both individual and contextual factors in assessing academic performance.
Morera et al. (2009) [Mo]	Identify and select high- and low-efficiency schools in Baja California based on ENLACE scores and contextual tests, using multilevel linear models to calculate residuals at the school level.	Multilevel linear model	The results indicate that the distributions of the selected schools are similar to the reference population, for both high and low-performing schools. The study confirms the robustness of the selection process used and emphasizes its importance in identifying best practices in high-efficiency schools.
Bartolucci et al. (2023) [B]	Propose a transition model to assess the effect of policies or treatments on multivariate outcomes, accounting for unobserved heterogeneity. This model extends the differences-in-differences method to express causal effects in terms of transition probabilities.	Latent Causal Transition Model	The study showed that educational programs benefited students with higher cognitive abilities. The proposed methodology proved effective in analyzing data with multiple outcomes and heterogeneous populations, and can be generalized and extended to include school or class effects. Comparisons with the differences-in-differences (DiD) method offered additional insights into the effects of the programs.

Source: Own elaboration (2026).

Note: [(a)] Each cluster is identified by the term in brackets, with “Br” referring to the cluster led by Bryk; “C” indicating the cluster led by Cooper; “M” representing the cluster led by Muthén; “Mo” denoting the cluster led by Morera; and “B” signifying the cluster led by Bartolucci.

APPENDIX B

Table B1 – General search string adapted for each database

(continues)

	Database	nº of studies
	ERIC	
String 1	abstract:(educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models") AND (Andorra OR Argentina OR Bolivia OR Brazil OR Chile OR Colombia OR "Costa Rica" OR Cuba OR Ecuador OR "El Salvador" OR Spain OR "Equatorial Guinea" OR Guatemala OR Honduras OR Mexico OR Nicaragua OR Panama OR Paraguay OR Peru OR Portugal OR "Dominican Republic" OR Uruguay OR Venezuela OR Brasil OR España OR "Guinea Ecuatorial" OR México OR Panamá OR Perú OR "República Dominicana" OR Bolívia OR Colômbia OR Equador OR Espanha OR "Guiné Equatorial" OR Nicarágua OR Panamá OR Paraguai OR Uruguai OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica")	36
String 2	abstract:(educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models") AND location:(Andorra OR Argentina OR Bolivia OR Brazil OR Chile OR Colombia OR "Costa Rica" OR Cuba OR Ecuador OR "El Salvador" OR Spain OR "Equatorial Guinea" OR Guatemala OR Honduras OR Mexico OR Nicaragua OR Panama OR Paraguay OR Peru OR Portugal OR "Dominican Republic" OR Uruguay OR Venezuela OR Brasil OR España OR "Guinea Ecuatorial" OR México OR Panamá OR Perú OR "República Dominicana" OR Bolívia OR Colômbia OR Equador OR Espanha OR "Guiné Equatorial" OR Nicarágua OR Panamá OR Paraguai OR Uruguai OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica")	42
	Scopus	
String 1	TITLE-ABS-KEY((educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models") AND (Andorra OR Argentina OR Bolivia OR Brazil OR Chile OR Colombia OR "Costa Rica" OR Cuba OR Ecuador OR "El Salvador" OR Spain OR "Equatorial Guinea" OR Guatemala OR Honduras OR Mexico OR Nicaragua OR Panama OR Paraguay OR Peru OR Portugal OR "Dominican Republic" OR Uruguay OR Venezuela OR Brasil OR España OR "Guinea Ecuatorial" OR México OR Panamá OR Perú OR "República Dominicana" OR Bolívia OR Colômbia OR Equador OR Espanha OR "Guiné Equatorial" OR Nicarágua OR Panamá OR Paraguai OR Uruguai OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica"))	27

(conclusion)

	Database	nº of studies
String 2	TITLE-ABS-KEY((educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models")) AND (LIMIT-TO (AFFILCOUNTRY, "Andorra") OR LIMIT-TO (AFFILCOUNTRY, "Argentina") OR LIMIT-TO (AFFILCOUNTRY, "Bolivia") OR LIMIT-TO (AFFILCOUNTRY, "Brazil") OR LIMIT-TO(AFFILCOUNTRY,"Chile") OR LIMIT-TO (AFFILCOUNTRY, "Colombia") OR LIMIT-TO (AFFILCOUNTRY, "Costa Rica") OR LIMIT-TO(AFFILCOUNTRY,"Cuba") OR LIMIT-TO (AFFILCOUNTRY, "Ecuador") OR LIMIT-TO (AFFILCOUNTRY, "El Salvador") OR LIMIT-TO (AFFILCOUNTRY, "Spain") OR LIMIT-TO (AFFILCOUNTRY, "Equatorial Guinea") OR LIMIT-TO (AFFILCOUNTRY, "Guatemala") OR LIMIT-TO (AFFILCOUNTRY, "Honduras") OR LIMIT-TO (AFFILCOUNTRY, "Mexico") OR LIMIT-TO (AFFILCOUNTRY, "Nicaragua") OR LIMIT-TO (AFFILCOUNTRY, "Panama") OR LIMIT-TO (AFFILCOUNTRY, "Paraguay") OR LIMIT-TO (AFFILCOUNTRY, "Peru") OR LIMIT-TO (AFFILCOUNTRY, "Portugal") OR LIMIT-TO (AFFILCOUNTRY, "Dominican Republic") OR LIMIT-TO (AFFILCOUNTRY, "Uruguay") OR LIMIT-TO (AFFILCOUNTRY, "Venezuela") OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica")	33
	Web of Science	
String 1	TS=((educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models") AND (Andorra OR Argentina OR Bolivia OR Brazil OR Chile OR Colombia OR "Costa Rica" OR Cuba OR Ecuador OR "El Salvador" OR Spain OR "Equatorial Guinea" OR Guatemala OR Honduras OR Mexico OR Nicaragua OR Panama OR Paraguay OR Peru OR Portugal OR "Dominican Republic" OR Uruguay OR Venezuela OR Brasil OR España OR "Guinea Ecuatorial" OR México OR Panamá OR Perú OR "República Dominicana" OR Bolívia OR Colômbia OR Equador OR Espanha OR "Guiné Equatorial" OR Nicarágua OR Panamá OR Paraguai OR Uruguai OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica"))	26
String 2	TS=((educa** OR escola** OR professor** OR estudante** OR escuela** OR estudiante** OR educat** OR school** OR teach** OR student**) AND ("modelo de valor agregado" OR "modelos de valor agregado" OR "modelo de valor adicionado" OR "modelos de valor adicionado" OR "modelo de VA" OR "modelos de VA" OR "modelos de valor añadido" OR "value-added model" OR "value-added models" OR "value added model" OR "value added models" OR "VA model" OR "VA models")) AND CU=(("Andorra" OR "Argentina" OR "Bolivia" OR "Brazil" OR "Chile" OR "Colombia" OR "Costa Rica" OR "Cuba" OR "Ecuador" OR "El Salvador" OR "Spain" OR "Equatorial Guinea" OR "Guatemala" OR "Honduras" OR "Mexico" OR "Nicaragua" OR "Panama" OR "Paraguay" OR "Peru" OR "Portugal" OR "Dominican Republic" OR "Uruguay" OR "Venezuela" OR "Ibero-Americ**" OR Iberoameric** OR "Ibero-American" OR "Ibero-América" OR "Iberoamérica"))	26

Source: Own elaboration (2026).

Table B2 – Summary of the reviewed studies

(continues)

Authors and year	Journal	Statistical model	Analyzed location	Objective
Gomes-Neto et al. (1997)	Economics of Education Review	Multiple regression	Rural Northeast - Brazil	Investigate the complementarities between health, academic performance, and cognitive performance, highlighting the impact of visual acuity and nutrition.
Ferrão (2009)	Revista de Educacion	Multilevel model	Cova da Beira - Portugal	Assess how different measures of socioeconomic status (SES), or their exclusion, affect estimates of school Value Added (VA) in a variance components model.
Ferrão & Goldstein (2009)	Quality & Quantity	Multilevel model	Portugal	Adjust for measurement error in multilevel VA models to evaluate its impact on school effect estimates.
Hernández & Olariaga (2009)	Revista de Educacion	Regression combined with Item Response Theory (IRT)	Madrid - Spain	Analyze the dimensional structure of mathematics tests administered in the Community of Madrid (2005-06 and 2006-07) across different educational cycles, evaluating their essential unidimensionality and structural complexity using non-parametric IRT-based procedures.
Castro-Morera et al. (2009)	Revista de Educacion	Multilevel model	Madrid - Spain	Examine the key elements that affect school growth models in a longitudinal design, considering the student's initial knowledge status as a predictor of growth rate, and evaluate the impact of the statistical regression effect on school rankings.
Vivanco (2013)	Economia Mexicana, Nueva Epoca	Multilevel model and spatial techniques	Mexico	Analyze the impact of school characteristics on students' educational performance in Mexico over time, using fixed effects models, multilevel models, and spatial techniques to address endogeneity and provide insights for educational policies.
Ferrão & Couto (2013)	Ensaio	Multilevel model	Brazil	Discuss the use of the VA indicator and the choice of statistical model in the literature on school effectiveness in Brazil, evaluating the consistency and stability of the variance components model applied to GERES 2005 data.
Brooke et al. (2014)	Educação e Pesquisa	Multilevel model	Brazil	Compare two VA model approaches applied to the GERES 2005 longitudinal study data, identifying their advantages and limitations in analyzing school performance and student progress overtime.
Ferrão (2014)	Educational Research for Policy and Practice	Multilevel model	Brazil and Portugal	The goal is to analyze regional disparities in educational outcomes and evaluate the stability and consistency of VA estimates across models, curricular content, and over time.
Ferrão & Couto (2014)	School Effectiveness and School Improvement	Multilevel model	Cova da Beira - Portugal	Promote school improvement in Portugal by applying a VA approach to longitudinal data from primary schools to analyze the stability of VA estimates over time and identify schools with consistently underperforming outcomes.
López-Martín et al. (2014)	Educational Assessment, Evaluation and Accountability	Multilevel model and its nonlinear version	Madrid - Spain	Estimate the VA of schools for reading comprehension using nonlinear growth models. The study aims to demonstrate the benefits of using nonlinear growth trajectories and incorporating student and family characteristics for more accurate VA estimates.
Castro-Morera et al. (2015)	Education Policy Analysis Archives	Multilevel model	Baja California - Mexico	The goal is to identify effective educational practices by comparing schools with high and low VA, controlling for socioeconomic context, and emphasizing features such as democratic coexistence and teacher support.
Santelices et al. (2015)	Psykhe	Multilevel model	Chile	The study aims to analyze teacher quality in Chile by estimating the impact of teachers on learning and predicting student performance based on contextual variables and teacher characteristics, using multilevel models.
Muñoz-Chereau & Thomas (2016)	Assessment in Education: Principles, Policy & Practice	Multilevel efficiency model with analysis of variance	Chile	Analyze the accuracy of two-level models in assessing the effectiveness of secondary schools in Chile and how student characteristics and school context influence performance.
Thieme et al. (2016)	European Journal of Operational Research	Frontier model	Chile	Propose a robust frontier model to estimate the contextual VA of schools in Chile and assess their effectiveness, controlling for contextual variables and analyzing performance differences between public, subsidized private, and fee-paying private schools.

(continues)

Troncoso et al. (2016)	School Effectiveness and School Improvement	Multilevel model	Chile	Implement a four-level contextual VA model to study student progress in mathematics in Chile, considering class and locality effects that are masked by traditional models.
Taut et al. (2016)	Assessment in Education: Principles, Policy & Practice	Multilevel model	Chile	Investigate the validity of a national teacher evaluation program in Chile by examining the relationship between teacher evaluation outcomes and student progress.
Page et al. (2017)	Journal of the Royal Statistical Society Series A-Statistics in Society	Quantile regression model	Chile	Develop a VA definition based on the quantiles of student score distributions to provide a more comprehensive picture of school effectiveness and apply the methodology to standardized test data in Chile.
Santelices et al. (2017)	Educational Assessment, Evaluation and Accountability	Multilevel model	Chile	Examine the relationship between two measures of teacher quality—one based on professional standards and the other on VA estimates—and the impact of teacher characteristics and school context on student performance in Chile.
Soares et al. (2017)	Ensaio	Multilevel model and status models	Brazil	Compare statistical models of varying complexity to determine the effectiveness of primary schools, assessing whether a more complex model improves accuracy, consistency, and stability in representing school performance.
Ortega et al. (2018)	School Effectiveness and School Improvement	Multilevel model - accelerated growth curve model	Chile	Investigate the effects of schools on student performance in language and mathematics over time by analyzing variability across schools and their influence on student growth from grades 3 to 8.
Torres (2018)	School Effectiveness and School Improvement	Multilevel model	Chile	Examine socioeconomic inequality in teacher effects by analyzing how variations in teacher performance may explain disparities in student achievement in Chilean secondary schools.
Bartholo et al. (2019)	Ensaio	Multivariate regression models	Rio de Janeiro - Brazil	Identify family factors and prior daycare experiences that correlate with children's cognitive development upon entering compulsory education, using multivariate regression models.
Filho (2019)	Revista Brasileira de Economia	Multilevel model	Brazil	Assess the impact of teachers' education on student performance in public schools, analyzing whether specialized training in mathematics or languages significantly affects academic achievement.
Jerrim et al. (2019)	British Educational Research Journal	Twin fixed-effects model	Spain	Investigate the relationship between time spent on homework and elementary school students' academic performance using longitudinal data and VA models to control for fixed effects, particularly in the Spanish context.
Muñoz-Chereau (2019)	School Effectiveness and School Improvement	Multilevel model	Chile	Analyze within-school and between-school variations in Chilean schools regarding the promotion of academic performance among boys and girls, investigating the role of schools in explaining gender disparities using multilevel modeling.
Troncoso (2019)	International Journal of Educational Research	Multilevel model	Chile	Propose a five-level bivariate statistical model to assess schools' contributions to student progress in Mathematics and Language in Chile, addressing the limitations of traditional VA models by accounting for variations across classrooms, schools, and local authorities.
Berthelon et al. (2020)	Economics and Human Biology	Fixed effects model for panel data	Chile	Investigate the relationship between parents' use of strict disciplinary strategies and their impact on the cognitive and socioemotional development of young children, employing a VA model that controls various children, family, and maternal characteristics.
Eigbiremolen et al. (2020)	Journal of International Development	Dynamic OLS model or lagged value-added model	Peru	Evaluate the effect of private school education on student learning by comparing raw scores with VA estimates, demonstrating that private schools show no significant effects after controlling prior performance.
Singh (2020)	Journal of the European Economic Association	Multilevel model	Ethiopia, India, Peru and Vietnam	Investigate differences in primary school productivity across four developing countries (Ethiopia, India, Peru, and Vietnam) and analyze how these differences contribute to international variations in human capital.
Ferrão (2022)	REICE-Revista Iberoamericana Sobre Calidad Eficacia y Cambio en Educacion	Multilevel model	Brazil	Examine differential effectiveness and social equity in Brazilian schools from 5th to 9th grade, applying random coefficient models to assess how sociodemographic factors and prior knowledge influence academic proficiency and social equity.

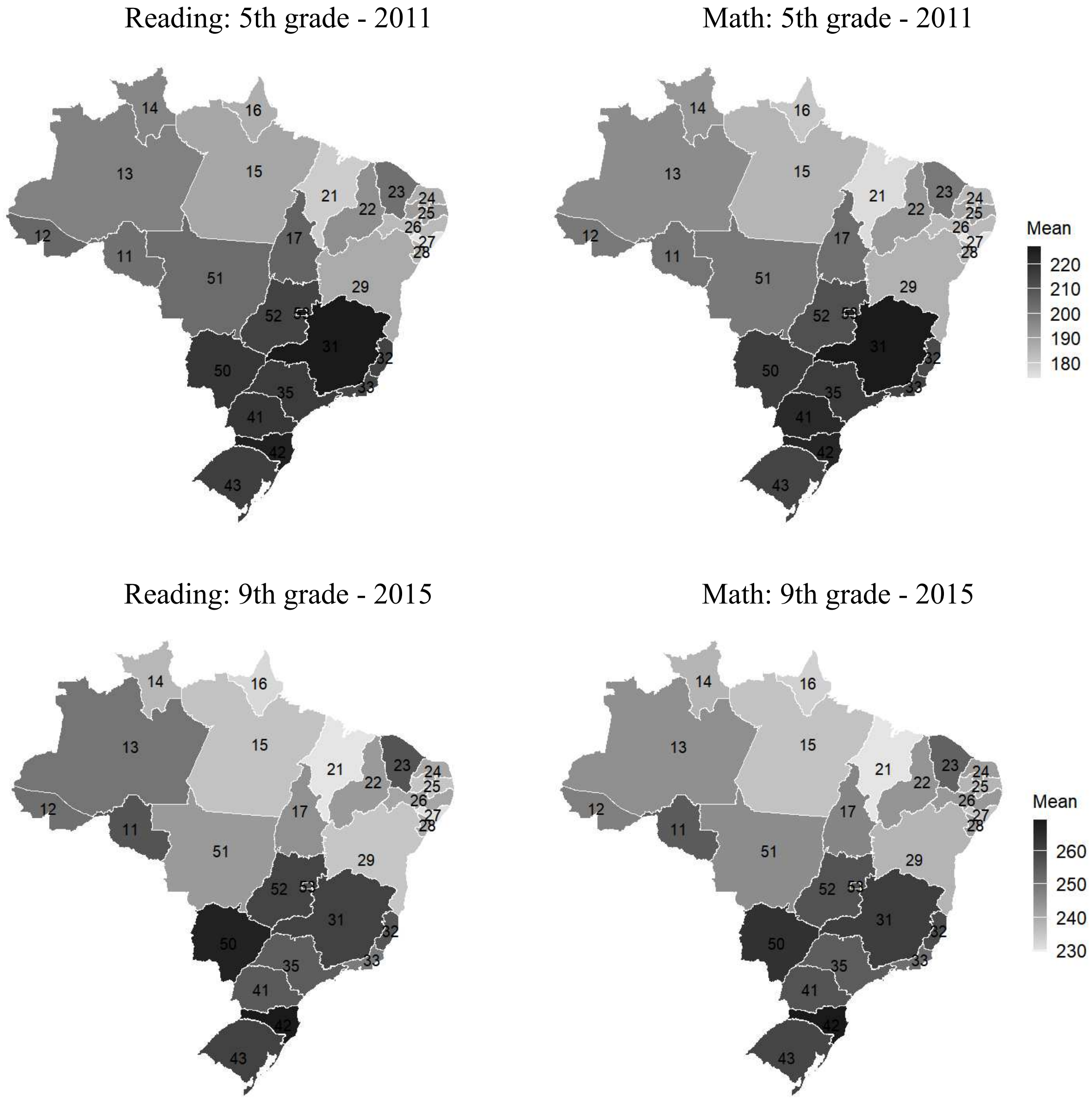
(conclusion)

Bartholo et al. (2023)	Ensaio	Multilevel model	Rio de Janeiro - Brazil	Assess the impact of school closures during the COVID-19 pandemic on learning loss and educational inequalities in Rio de Janeiro, using VA models to estimate effects on language and mathematics development, particularly for children from low socioeconomic status families.
Barrios-Fernández & Riudavets-Barcons (2024)	Educational Evaluation and Policy Analysis	Fixed effects model	Chile	Estimate Teacher Value Added (TVA) on test scores and educational attainment, examine gender differences in teacher effectiveness, and assess their contribution to gender gaps and associated teacher characteristics and practices.
Bertoni et al. (2024)	Educational Evaluation and Policy Analysis	Multilevel model	Peru	Evaluate whether teacher assessment instruments used in the teacher selection process in Peru predict teacher effectiveness by examining the relationship between teachers' VA measures and different components of the teacher evaluation process.
Muñoz & Prem (2024)	American Economic Journal: Economic Policy	Fixed effects model	Chile	Estimate the effectiveness of school principals and analyze the impact of a public sector reform aimed at more competitive and transparent principal selection on their effectiveness and on student outcomes.
Page et al. (2024)	Psychometrika	Multilevel model	Chile	Estimate the dynamic persistence of school effectiveness over time by incorporating temporal dependence into VA models to monitor school performance and persistence, with a focus on improving estimation efficiency.

Source: Own elaboration (2026).

APPENDIX C

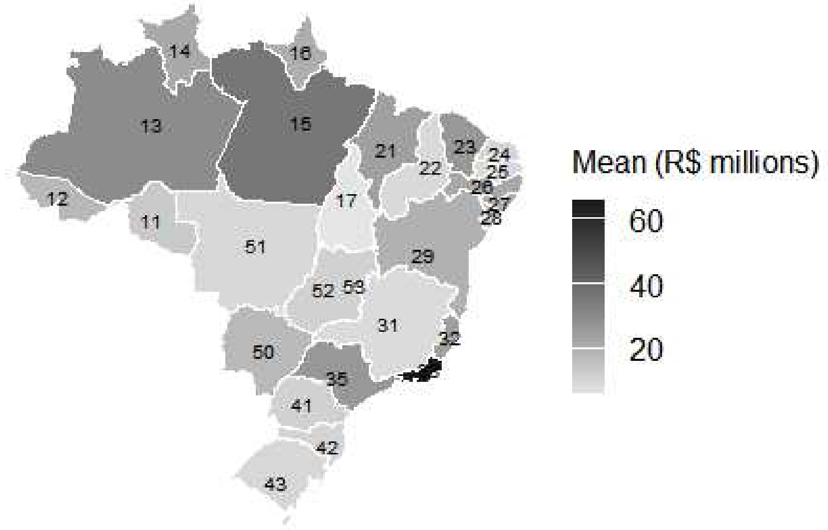
Figure C1 – Mean achievement in Reading and Math in 5th and 9th grades (2011 and 2015)



Source: Own elaboration (2026).

Note: 11 - Rondônia; 12 - Acre; 13 - Amazonas; 14 - Roraima; 15 - Pará; 16 - Amapá; 17 - Tocantins; **21 - Maranhão**; 22 - Piauí; 23 - Ceará; 24 - Rio Grande do Norte; 25 - Paraíba; 26 - Pernambuco; 27 - Alagoas; 28 - Sergipe; 29 - Bahia; **31 - Minas Gerais**; 32 - Espírito Santo; 33 - Rio de Janeiro; 35 - São Paulo; 41 - Paraná; 42 - Santa Catarina; 43 - Rio Grande do Sul; 50 - Mato Grosso do Sul; 51 - Mato Grosso; 52 - Goiás; 53 - Distrito Federal.

Figure C2 – Mean Fundeb resource allocation in 2015



Source: Own elaboration (2026).
Note: 11 - Rondônia; 12 - Acre; 13 - Amazonas; 14 - Roraima; 15 - Pará; 16 - Amapá; 17 - Tocantins; **21 - Maranhão**; 22 - Piauí; 23 - Ceará; 24 - Rio Grande do Norte; 25 - Paraíba; 26 - Pernambuco; 27 - Alagoas; 28 - Sergipe; 29 - Bahia; **31 - Minas Gerais**; 32 - Espírito Santo; 33 - Rio de Janeiro; 35 - São Paulo; 41 - Paraná; 42 - Santa Catarina; 43 - Rio Grande do Sul; 50 - Mato Grosso do Sul; 51 - Mato Grosso; 52 - Goiás; 53 - Distrito Federal.

Table C1 – Mean and standard deviation of Reading and Math proficiency by demographic, family, and school characteristics in Maranhão, 2011 and 2015

Variables		Proficiency							
		Reading 2011		Reading 2015		Math 2011		Math 2015	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sex	Male	171.76	38.70	230.24	44.48	190.48	40.66	239.99	40.86
	Female	178.72	39.68	241.78	42.92	183.61	37.11	229.74	36.89
	N/A	152.78	29.60	206.98	40.95	166.08	30.65	220.57	30.87
Self-declared color/race	White	176.26	40.83	238.35	46.35	187.33	40.05	234.61	39.22
	Pardo	176.40	39.08	237.46	43.49	187.00	38.76	234.62	39.01
	Black	175.66	40.11	236.17	43.92	186.54	38.23	233.86	38.52
	Asian-descendant	180.46	39.97	242.97	43.51	191.56	39.23	237.89	39.98
	Indigenous	174.00	40.46	237.18	40.66	185.03	36.50	233.76	37.74
	Don't know	165.71	35.93	224.35	42.58	177.61	36.16	227.01	37.70
	N/A	162.18	36.64	210.05	46.20	175.00	39.79	214.53	35.69
Worker student 2011	No	178.89	39.64	240.29	43.74	188.85	38.97	235.69	39.25
	Yes	160.01	33.82	219.09	40.48	176.57	36.19	227.34	36.82
	N/A	158.03	33.66	219.30	41.51	170.56	34.19	224.50	36.01
Worker student 2015	No	177.67	39.51	239.54	43.43	187.67	38.85	235.20	39.06
	Yes	167.10	37.60	221.81	42.95	183.87	38.12	229.80	38.04
	N/A	170.12	38.73	220.26	45.84	181.32	38.49	224.02	37.22
Literate mother	No	165.43	35.82	226.59	40.77	177.54	35.87	224.66	34.56
	Yes	176.76	39.60	237.84	44.15	187.43	38.99	235.12	39.26
	N/A	160.34	34.50	224.35	46.99	172.77	34.46	222.65	37.43
Literate father	No	166.68	35.69	227.68	41.29	178.63	35.89	226.76	36.01
	Yes	177.57	39.81	238.66	44.20	188.14	39.12	235.66	39.36
	N/A	166.41	38.00	229.18	48.24	178.34	37.99	227.63	38.60
Live with mother	No	174.32	37.61	233.92	41.98	184.69	37.80	231.65	37.43
	Yes	176.36	39.73	237.30	44.12	187.22	38.99	234.53	39.16
	N/A	171.40	38.26	228.55	45.65	182.33	38.02	229.13	37.22
Live with father	No	176.76	39.15	237.84	43.18	186.70	38.21	233.76	37.91
	Yes	175.73	39.62	236.45	44.27	186.95	39.10	234.38	39.45
	N/A	171.89	38.68	229.68	46.38	182.83	38.52	231.11	39.14
Grade repetition before 2011	No	181.26	39.86	243.11	43.67	191.46	39.41	238.60	39.63
	Yes	160.31	33.52	218.91	39.34	173.10	33.42	221.02	33.80
	N/A	163.58	36.42	223.72	43.47	174.19	35.79	227.25	36.57
Administrative dependency	Federal	236.66	32.04	305.57	33.87	253.35	31.94	298.07	39.40
	State	188.95	42.08	246.30	46.35	196.38	40.21	239.01	40.87
	Municipal	173.38	38.49	235.17	43.35	184.72	38.27	233.26	38.53
Location of school	Urban	177.58	39.60	238.75	43.97	188.30	39.08	235.66	39.26
	Rural	166.39	37.10	227.66	42.87	177.68	36.28	227.08	36.81
Differentiated location	No	176.01	39.43	237.17	43.98	186.84	38.82	234.36	39.04
	Yes	160.18	34.96	223.64	41.94	170.79	34.78	226.27	35.82
Tutoring at school	No	173.67	38.85	235.07	43.67	184.51	38.11	232.20	38.49
	Yes	175.91	39.49	237.07	44.04	186.73	38.92	234.50	39.05
	N/A	185.05	42.22	251.75	43.52	197.19	41.40	233.86	39.69
Tenured teachers	≤50%	173.81	39.59	234.16	44.16	184.58	38.28	232.34	39.33
	>50%	176.07	39.34	237.52	43.89	186.90	38.93	234.60	38.87
	N/A	173.11	38.49	234.42	45.91	185.17	38.84	237.08	36.91
Director's sex	Male	173.76	38.88	234.87	43.75	185.23	38.76	233.98	39.10
	Female	176.20	39.56	237.41	44.05	186.78	38.82	234.17	38.92
	N/A	174.25	40.03	240.37	43.20	187.50	36.42	232.56	38.52
Higher education of the director	No	167.69	36.92	228.28	43.11	179.31	36.71	228.07	36.82
	Yes	175.91	39.44	237.06	43.98	186.62	38.81	234.33	38.98
	N/A	180.17	41.44	244.70	43.36	193.59	41.18	240.00	41.61
Total		175.55	39.40	236.76	43.98	186.37	38.80	234.12	38.97

Source: Own elaboration (2026).
Note: N/A = Not Available.

Table C2 – Mean and standard deviation of Reading and Math proficiency by demographic, family, and school characteristics in 2011 and 2015 - Minas Gerais

Variables		Proficiency							
		Reading 2011		Reading 2015		Math 2011		Math 2015	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sex	Male	209.88	45.37	252.50	50.35	241.20	46.63	269.10	47.63
	Female	218.82	44.24	269.48	45.07	233.07	45.11	262.74	45.14
	N/A	176.37	41.38	225.61	44.97	197.89	43.99	226.84	39.96
Self-declared color/race	White	222.97	45.18	270.61	48.13	245.77	45.97	275.23	47.97
	Pardo	213.73	44.44	260.89	47.53	235.96	45.51	264.75	45.49
	Black	207.23	44.46	252.61	48.47	227.56	44.99	255.68	44.13
	Asian-descendant	216.00	44.47	263.91	47.15	236.41	44.59	264.13	45.02
	Indigenous	214.11	42.59	258.58	45.11	233.01	43.28	260.69	42.92
	Don't know	201.14	44.35	246.48	49.17	223.93	46.52	253.69	45.28
	N/A	198.38	45.34	245.08	53.74	221.49	47.01	251.92	48.63
Worker student 2011	No	217.45	44.35	264.35	47.63	238.77	45.46	267.28	46.27
	Yes	193.61	44.07	240.61	48.24	223.01	46.79	254.26	45.59
	N/A	192.00	44.90	241.27	49.18	214.64	47.78	249.01	44.98
Worker student 2015	No	217.21	44.62	264.57	47.38	238.22	45.68	266.82	46.24
	Yes	208.83	45.28	250.13	49.57	237.61	46.43	262.48	46.47
	N/A	208.04	45.48	249.39	52.15	229.25	46.52	255.89	47.39
Literate mother	No	194.61	42.48	243.85	45.38	214.73	45.23	244.29	41.68
	Yes	215.37	44.93	262.28	48.28	237.48	45.85	266.35	46.39
	N/A	192.17	42.97	228.68	50.99	213.46	46.41	239.51	44.98
Literate father	No	196.62	42.44	243.93	46.39	217.34	45.21	246.83	42.88
	Yes	215.99	44.90	262.94	48.18	238.20	45.75	266.98	46.36
	N/A	204.46	45.12	252.16	50.24	224.34	46.52	256.07	45.61
Live with mother	No	207.59	43.94	250.97	48.53	229.84	45.72	257.04	44.34
	Yes	216.13	44.90	262.56	48.12	238.35	45.86	266.39	46.42
	N/A	208.09	45.32	249.71	51.62	229.08	46.20	254.02	47.50
Live with father	No	213.40	44.58	258.69	47.91	233.92	45.50	261.01	45.27
	Yes	216.49	45.04	263.13	48.30	239.31	46.04	267.70	46.71
	N/A	208.90	45.15	255.65	50.27	230.28	46.02	260.54	46.40
Grade repetition before 2011	No	220.15	44.09	267.02	47.20	242.21	44.85	270.56	46.00
	Yes	185.99	38.45	231.59	43.46	208.17	41.03	237.76	38.52
	N/A	196.79	43.41	245.37	47.31	218.11	45.96	251.35	43.31
Administrative dependency	Federal	245.13	43.43	295.68	42.21	275.50	39.81	323.85	41.35
	State	215.16	45.08	260.50	48.73	237.59	46.15	264.73	46.61
	Municipal	212.92	44.82	264.62	47.04	233.91	45.58	267.78	45.66
Location of school	Urban	215.08	45.01	261.94	48.34	237.17	45.94	265.91	46.46
	Rural	198.79	42.87	252.74	46.64	219.58	45.96	256.56	43.92
Differentiated location	No	214.67	45.01	261.72	48.31	236.73	46.02	265.69	46.42
	Yes	200.17	46.89	250.61	49.44	220.59	46.93	254.99	44.64
Tutoring at school	No	216.15	44.91	262.17	48.58	237.96	45.84	265.54	46.29
	Yes	214.40	45.04	261.60	48.29	236.48	46.07	265.64	46.44
	N/A	210.63	44.95	260.52	46.37	233.82	44.45	264.51	43.29
Tenured teachers	≤50%	213.39	45.00	259.99	48.28	235.68	46.14	263.71	46.23
	>50%	216.12	45.01	263.81	48.26	237.89	45.89	268.10	46.50
	N/A	216.26	45.95	261.37	50.15	236.50	45.58	264.65	48.78
Director's sex	Male	215.73	44.86	261.57	48.62	238.27	45.83	266.07	46.58
	Female	214.26	45.07	261.69	48.23	236.16	46.11	265.49	46.37
	N/A	208.81	46.12	261.27	48.40	230.24	43.60	264.63	45.20
Higher education of the director	No	212.32	43.96	262.33	47.15	232.50	45.28	263.45	45.95
	Yes	214.65	45.03	261.69	48.32	236.73	46.04	265.70	46.42
	N/A	211.72	45.40	259.36	48.69	233.04	46.43	261.94	46.04
Total		214.59	45.03	261.66	48.32	236.64	46.04	265.63	46.41

Source: Own elaboration (2026).
Note: N/A = Not Available.

Table C3 – Mean and standard deviation of Reading and Math proficiency by demographic, family, and school characteristics in 2011 and 2015 - Brazil

Variables		**Proficiency**							
		Reading 2011		**Reading 2015**		**Math 2011**		**Math 2015**	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Sex	Male	195.65	44.60	246.11	48.93	224.20	47.61	260.96	45.79
	Female	204.32	44.48	260.09	45.19	215.72	45.21	251.54	43.34
	N/A	162.16	36.88	211.42	43.43	181.48	39.61	223.65	35.38
Self-declared color/race	White	209.28	45.63	262.91	47.79	229.66	47.36	265.06	46.34
	Pardo	197.86	43.84	251.76	46.36	216.82	45.69	253.73	43.68
	Black	192.90	43.54	243.88	47.34	211.62	44.38	247.20	42.10
	Asian-descendant	202.36	44.06	256.81	45.99	220.01	45.38	256.11	44.13
	Indigenous	199.79	42.68	251.78	45.03	216.48	44.27	252.26	42.17
	Don’t know	187.69	42.98	238.89	47.65	207.43	45.23	244.92	42.92
	N/A	183.25	45.08	231.69	51.62	202.36	47.73	240.62	44.45
Worker student 2011	No	203.29	44.30	256.53	46.83	221.95	46.16	257.55	44.72
	Yes	178.94	41.52	232.46	46.33	204.45	45.62	245.45	43.22
	N/A	175.51	42.57	230.78	47.40	195.50	45.07	238.89	41.94
Worker student 2015	No	202.47	44.54	256.08	46.71	220.96	46.38	256.81	44.63
	Yes	194.11	44.92	241.32	48.70	219.71	47.26	253.12	44.66
	N/A	193.41	44.78	237.87	51.15	212.35	46.28	244.40	45.28
Literate mother	No	181.23	40.58	236.39	44.31	198.03	42.85	238.57	40.43
	Yes	201.15	44.75	254.41	47.47	220.59	46.47	256.71	44.76
	N/A	179.13	41.97	230.00	48.31	197.29	43.93	236.29	43.18
Literate father	No	183.38	40.99	238.16	44.68	201.07	43.12	240.94	41.04
	Yes	201.84	44.76	255.05	47.45	221.32	46.47	257.29	44.80
	N/A	189.82	44.66	243.57	50.52	208.28	46.26	247.24	44.83
Live with mother	No	194.00	42.97	244.88	46.55	213.44	45.31	248.46	42.65
	Yes	201.63	44.84	254.43	47.39	221.04	46.64	256.60	44.81
	N/A	194.06	44.45	240.35	50.36	212.81	45.82	244.75	44.45
Live with father	No	199.40	44.14	251.50	46.81	217.64	45.54	252.36	43.48
	Yes	201.75	45.00	254.67	47.62	221.63	46.97	257.51	45.13
	N/A	194.71	44.55	246.32	50.50	213.58	46.01	250.54	45.73
Grade repetition before 2011	No	206.07	44.28	259.40	46.55	225.41	46.08	260.86	44.67
	Yes	175.46	37.68	227.58	42.56	194.79	39.69	233.48	37.65
	N/A	181.37	42.57	235.74	46.57	200.18	44.71	241.64	41.80
Administrative dependency	Federal	247.91	42.18	300.04	41.27	270.59	41.41	316.36	45.48
	State	202.15	44.75	254.10	47.74	222.42	46.41	256.51	44.48
	Municipal	196.96	44.61	252.65	47.04	214.81	46.36	254.74	45.01
Location of school	Urban	201.26	44.72	254.35	47.48	220.73	46.41	256.55	44.74
	Rural	181.46	41.69	240.69	45.67	198.20	44.01	244.73	43.21
Differentiated location	No	200.31	44.77	253.69	47.48	219.66	46.54	255.99	44.75
	Yes	181.34	42.65	238.02	46.36	196.79	43.90	241.60	41.46
Tutoring at school	No	199.41	44.73	251.36	47.61	218.21	46.41	253.78	44.29
	Yes	200.24	44.79	253.82	47.47	219.62	46.58	256.11	44.78
	N/A	199.48	44.77	254.07	47.88	219.24	46.98	257.44	46.45
Tenured teachers	≤50%	200.81	44.99	253.32	47.47	220.29	47.04	255.98	44.86
	>50%	199.93	44.71	253.66	47.49	219.19	46.39	255.83	44.70
	N/A	199.77	44.72	252.52	47.99	219.39	46.47	255.82	45.09
Director's sex	Male	199.02	44.77	252.12	47.38	218.11	46.63	254.53	44.57
	Female	200.56	44.78	254.07	47.51	219.96	46.53	256.34	44.79
	N/A	197.53	46.14	251.97	48.03	216.82	46.93	254.78	45.47
Higher education of the director	No	188.90	43.37	243.27	46.84	205.36	44.59	245.95	42.73
	Yes	200.34	44.77	253.71	47.45	219.62	46.54	256.00	44.76
	N/A	200.03	45.19	254.00	48.02	221.01	46.88	256.48	44.64
Total		200.16	44.79	253.57	47.49	219.48	46.56	255.87	44.74

Source: Own elaboration (2026).
Note: N/A = Not Available.

Table C4 – Estimates of fixed and random effects from the multilevel regression model without contextual controls for Maranhão and Minas Gerais (Model 1)

	Maranhão		**Minas Gerais**	
	Reading	**Math**	**Reading**	**Math**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Fixed effects				
Constant	0.6343^{***}	0.3123^{***}	0.6177^{***}	0.3936^{***}
	(0.0141)	(0.0129)	(0.0054)	(0.0051)
Proficiency 2011	0.6457^{***}	0.5107^{***}	0.6400^{***}	0.5947^{***}
	(0.0060)	(0.0066)	(0.0025)	(0.0024)
SES 2011	0.0221^{***}	0.0419^{***}	0.0129^{***}	0.0415^{***}
	(0.0053)	(0.0048)	(0.0027)	(0.0025)
Random effects				
Level: Municipality				
Var(Cons)	0.0119	0.0070	0.0049	0.0046
	(0.0024)	(.)	(0.0008)	(.)
Level: School				
Var(Cons)	0.0251	0.0485	0.0146	0.0207
	(0.0047)	(.)	(0.0013)	(.)
Var(Proficiency 2011)	0.0053	0.0141	0.0020	0.0027
	(0.0015)	(.)	(0.0004)	(.)
Var(SES)	0.0016	0.0004	0.0012	0.0013
	(0.0012)	(.)	(0.0005)	(.)
Covar(Proficiency 2011-SES)	-0.0014	-0.0014	0.0001	0.0008
	(0.0010)	(.)	(0.0003)	(.)
Covar(Proficiency 2011-Cons)	0.0079	0.0255	-0.0007	0.0075
	(0.0023)	(.)	(0.0005)	(.)
Covar(SES-Cons)	-0.0014	-0.0017	0.0025	0.0026
	(0.0016)	(.)	(0.0005)	(.)
Level: Class				
Var(Cons)	0.0133	0.0085	0.0322	0.0226
	(0.0018)	(.)	(0.0011)	(.)
Level: Student				
Var(Cons)	0.3599	0.3081	0.4097	0.3422
	(0.0032)	(.)	(0.0017)	(.)
N° of municipalities	216	216	831	831
N° of schools	1,255	1,255	3,402	3,402
N° of classes	2,588	2,588	8,710	8,710
N° of students	29,001	29,001	134,387	134,387
Deviance	54,444	49,756	270,53	245,996
AIC	54,467	49,763	270,555	246,002
BIC	54,567	49,788	270,672	246,031

Source: Own elaboration (2026).

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Deviance = -2*loglikelihood; AIC: Critério de Informação de Akaike; BIC: Critério Bayesiano de Schwarz.

Table C5 – Descriptive statistics

Variables	Maranhão	Minas Gerais	Brazil
	Percentage	**Percentage**	**Percentage**
Success			
No	49.5452	33.5159	42.1963
Yes	35.3657	56.3692	43.6353
N/A	15.0891	10.1149	14.1684
Sex			
Male	49.0253	49.6157	45.6226
Female	46.2881	48.6301	43.6931
N/A	4.6866	1.7542	10.6843
Self-declared race/color			
White	20.0019	27.0315	27.1445
Pardo	52.2361	48.1672	40.9094
Black	10.3776	9.5518	8.7066
Asian-descendant	0.1646	0.2833	0.2080
Indigenous	2.1899	2.5682	2.1332
Don’t know	9.3318	10.0051	9.5593
N/A	5.6981	2.3929	11.3390
Worker student 2011			
No	74.9621	84.4845	75.2704
Yes	15.9975	11.1903	11.7655
N/A	9.0404	4.3251	12.9641
Literate mother			
No	8.8556	3.1233	4.4345
Yes	85.2899	94.3728	84.0594
Don’t know	3.3269	1.1471	1.7043
N/A	2.5275	1.3568	9.8018
Literate father			
No	11.8486	4.2446	5.8457
Yes	73.5906	86.3019	76.0264
Don’t know	8.0787	3.8711	4.6656
N/A	6.4821	5.5825	13.4623
Live with mother			
No	5.4955	3.2474	3.2640
Yes	91.8846	95.2143	86.9526
N/A	2.6199	1.5383	9.7834
Live with father			
No	17.1925	18.4029	16.0963
Yes	77.6886	77.8949	72.1859
N/A	5.1189	3.7022	11.7178
Administrative dependency			
Federal	0.0640	0.0867	0.0587
State	12.2465	41.4947	24.1584
Municipal	87.6895	58.4186	67.3329
N/A	0	0	8.4499
Location of school			
Urban	75.5602	95.1788	82.8889
Rural	24.4398	4.8212	8.6611
N/A	0	0	8.4499

Source: Own elaboration (2026).
Note: N/A = Not Available.

Table C6 – Distribution of missings and valid cases by location

Variables	Maranhão			Minas Gerais			Brazil		
	Valid cases	Missings	% missings	Valid cases	Missings	% missings	Valid cases	Missings	% missings
School success	71,692	12,740	15.09	233,196	26,242	10.11	2,087,754	344,630	14.17
Reading 2011	84,334	98	0.12	255,928	3,510	1.35	2,212,175	220,209	9.05
Math 2011	84,427	5	0.01	259,436	2	0.00	2,226,156	206,228	8.48
SES 2011	83,224	1,208	1.43	257,256	2,182	0.84	2,205,335	227,049	9.33
Sex	80,475	3,957	4.69	254,887	4,551	1.75	2,172,501	259,883	10.68
Self-declared color/race	79,621	4,811	5.70	253,230	6,208	2.39	2,156,577	275,807	11.34
Worker student 2011	76,799	7,633	9.04	248,217	11,221	4.33	2,117,047	315,337	12.96
Literate mother	82,298	2,134	2.53	255,918	3,520	1.36	2,193,966	238,418	9.80
Literate father	78,959	5,473	6.48	244,955	14,483	5.58	2,104,928	327,456	13.46
Live with mother	82,220	2,212	2.62	255,447	3,991	1.54	2,194,415	237,969	9.78
Live with father	80,110	4,322	5.12	249,833	9,605	3.70	2,147,361	285,023	11.72
School administrative dependency	84,432	0	0.00	259,438	0	0.00	2,226,849	205,535	8.45
Location	84,432	0	0.00	259,438	0	0.00	2,226,849	205,535	8.45
Class size	84,432	0	0.00	259,438	0	0.00	2,226,849	205,535	8.45
School infrastructure	83,350	1,082	1.28	258,262	1,176	0.45	2,215,526	216,858	8.92
Pedagogical infrastructure	75,493	8,939	10.59	254,211	5,227	2.01	2,114,110	318,274	13.08
Log of Fundeb per student	84,432	0	0.00	259,438	26,242	9.19	2,226,933	205,451	8.45

Source: Own elaboration (2026).

Note: N/A = Not Available.

Table C7 – Estimates of fixed and random effects from the multilevel binary regression model without contextual controls (Model 1)

Variables	Maranhão		Minas Gerais		Brazil	
	Coefficient	**Odds ratios**	**Coefficient**	**Odds ratios**	**Coefficient**	**Odds ratios**
Fixed effects						
Constant	1.879*** (0.044)	6.547	1.856*** (0.025)	6,398	1.635*** (0.083)	5.129
Reading 2011	0.961*** (0.017)	2.614	0.871*** (0.009)	2,389	0.863*** (0.003)	2.370
Math 2011	0.588*** (0.017)	1.800	0.775*** (0.009)	2,171	0.705*** (0.003)	2.024
SES 2011	0.053*** (0.012)	1.054	0.163*** (0.007)	1,177	0.167*** (0.003)	1.182
Random effects						
Level: State (variance)	-		-		0.175	
Level: Municipality (variance)	0.173		0.306		0.258	
Level: School (variance)	0.078		0.154		0.111	
Level: Class (variance)	0.293		0.195		0.255	
Level: Student (variance)	Binomial		Binomial		Binomial	

Source: Own elaboration (2026).
Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.