



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E
CONTABILIDADE – FEAAC
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA - CAEN

DANIEL CATRIB DE AZEVEDO LEMOS

DETERMINANTES DO DESEMPENHO ESCOLAR NO ESTADO DO CEARÁ:
ABORDAGEM MULTINÍVEL

FORTALEZA

2020

DANIEL CATRIB DE AZEVEDO LEMOS

DETERMINANTES DO DESEMPENHO ESCOLAR NO ESTADO DO CEARÁ:
ABORDAGEM MULTINÍVEL

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Luiz Ivan de Melo Castelar

Coorientador: Prof. Dr. Frederico A. Gomes de Alencar.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

L576d Lemos, Daniel.

DETERMINANTES DO DESEMPENHO ESCOLAR NO ESTADO DO
CEARÁ: ABORDAGEM MULTINÍVEL / Daniel Lemos.

– 2021.

41 f.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de
Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em
Economia, Fortaleza, 2021.

Orientação: Prof. Dr. Prof. Dr. Luiz Ivan de Melo Castelar.

Coorientação: Prof. Dr. Prof. Dr. Frederico Augusto Gomes de Alencar.

1. Desempenho Escolar. 2. Modelo Multinível. 3. Escolas Públicas dos
Municípios Cearenses. 4. Prova Brasil. I. Título.

CDD 330

DANIEL CATRIB DE AZEVEDO LEMOS

DETERMINANTES DO DESEMPENHO ESCOLAR NO ESTADO DO CEARÁ:
ABORDAGEM MULTINÍVEL

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Luiz Ivan de Melo Castelar (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Brito Soares
Universidade Federal do Ceará (UFC)

Prof. Dr. Frederico Augusto Gomes de Alencar
Universidade Federal do Ceará (UFC)

RESUMO

Este trabalho investiga os fatores associados ao desempenho escolar dos alunos de escolas públicas dos municípios cearenses, mostrando especialmente a importância da turma, escola e município nos resultados da Prova Brasil de 2017 nas disciplinas de matemática e língua portuguesa. Para tanto foram utilizados dados do SAEB e Censo Escolar referentes aos alunos do 9º ano do ensino fundamental, através dos quais foi construído um modelo multinível que incorpora a natureza hierárquica dos dados, subdivididos nos níveis de aluno, turma, escola e município. Os resultados demonstram a importância que o /efeito-turma e o efeito-escola exercem nos resultados dos alunos, além da heterogeneidade dos escores entre os municípios cearenses. Fatores associados a defasagem escolar, como reprovação, foram os que mais impactaram as notas.

Palavras chave: Desempenho Escolar, Modelo Multinível, Escolas Públicas dos Municípios Cearenses, Prova Brasil.

ABSTRACT

This paper investigates the factors associated with the school performance of students from public schools in the municipalities of Ceará, showing especially the importance of the class, school and municipality in the results of Prova Brasil 2017 in the subjects of mathematics and Portuguese. For this purpose, data from the SAEB and Censo Escolar were used for students in the 9th grade of elementary school, through which a multilevel model was built that incorporates the hierarchical nature of the data, subdivided into the levels of student, class, school and municipality. The results demonstrate the importance that the class effect and the school effect exert on the students results, in addition to the heterogeneity of the scores between the cities of Ceará. Factors associated with school lag, such as failure, were the ones that most impacted the scores.

Keywords: School Performance, Multilevel Model, Public Schools in the Municipalities of Ceará, Prova Brasil.

LISTA DE GRÁFICOS

Gráfico 1 – Histograma da proficiência em matemática e língua portuguesa	21
--	----

LISTA DE TABELAS

Tabela 1 – Crescimento IDEB observado entre 2007 e 2017.....	13
Tabela 2 – Resumo de trabalhos que utilizaram modelos hierárquicos para avaliar os sistemas educacionais no Brasil	17
Tabela 3 – Número de observações	19
Tabela 4 – Variáveis independentes	19
Tabela 5 – Análise descritiva da variável dependente	20
Tabela 6 – Estatísticas descritivas das variáveis independentes	21
Tabela 7 – Índice de correlação interclasse para o modelo nulo	29
Tabela 8 – Proficiência em matemática no modelo nulo	30
Tabela 9 – Proficiência em língua portuguesa no modelo nulo	30
Tabela 10 – Partição da variância do modelo com variáveis explicativas no nível do aluno	31
Tabela 11 – Modelo 2 com variáveis independentes no nível do aluno	31
Tabela 12 – Partição da variância do Modelo 3	34
Tabela 13 – Modelo 3 com variáveis independentes no nível do aluno, turma e escola ..	35
Tabela 14 – Modelo 4 com variáveis independentes no nível do aluno, turma e escola, e efeitos aleatórios	36

LISTA DE ABREVIATURAS E SIGLAS

ABEP	Associação Brasileira de Empresas de Pesquisa
ICC	Índice de Correlação Intraclasse
IDEB	Índice de Desenvolvimento da Educação Básica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
PISA	Programa Internacional de Avaliação de Alunos
SAEB	Sistema Nacional de Avaliação da Educação Básica
SAERS	Sistema de Avaliação do Rendimento Escolar

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO DA LITERATURA	15
3	DADOS	18
3.1	Base de dados	18
3.2	Variáveis	19
3.3	Estatísticas descritivas	20
4	METODOLOGIA	23
4.1	Modelo multinível	23
4.2	Modelo econométrico	24
5	RESULTADOS	28
5.1	Modelo nulo	28
5.2	Modelos com variáveis no nível do aluno (efeitos fixos)	30
5.3	Modelos com variáveis no nível do aluno, da turma e da escola (efeitos fixos)	33
5.4	Modelos com variáveis no nível do aluno, da turma e da escola (efeitos fixos e aleatórios)	37
6	CONCLUSÃO	38
	REFERÊNCIAS	39

1 INTRODUÇÃO

A avaliação dos sistemas educacionais tem evoluído nas últimas décadas. Estudos econômicos estão se preocupando cada vez mais em avaliar esses sistemas, que são responsáveis diretos pelo crescimento econômico dos países. Conforme Hanushek (2002), a qualidade educacional tem um papel preponderante no crescimento econômico. A relação entre o resultado de avaliações internacionais padronizadas e o crescimento econômico são fortemente correlacionados, principalmente quando as estimativas são ponderadas por variáveis que captam as diferenças na qualidade do ensino nos diversos países (HANUSHEK; KIMKO, 2000).

Vários fatores podem influenciar o desempenho escolar do aluno. Dentre esses fatores podemos destacar o ambiente familiar e a infraestrutura escolar. Segundo Silva e Hasenbalg (2001), podem ser distinguidas três dimensões principais do ambiente familiar: a dos recursos econômicos ou capital econômico, a dos recursos educacionais ou capital cultural e a própria estrutura dos arranjos familiares. Estudos realizados no Brasil, que consideram a estrutura hierárquica dos dados, confirmaram a importância das características socioeconômicas dos alunos em sua trajetória escolar (RIANI, 2008).

Já a estrutura escolar diz respeito à disponibilidade da estrutura e sua qualidade. Como os fatores escolares são os mais propícios para a intervenção do gestor público, têm cada vez mais chamado a atenção de pesquisadores. Através desses fatores foram encontradas inúmeras evidências estatísticas que, mesmo controlado pelos fatores socioeconômicos e culturais dos alunos, o efeito da escola é significativo.

Nos países desenvolvidos, o impacto socioeconômico e cultural da família e o impacto da escola estão bem consolidados. Devido a uma maior variabilidade da qualidade das escolas e uma menor das diferenças socioeconômicas, estes países se concentram mais no debate em relação às escolas. Enquanto em países em desenvolvimento o debate sobre esses impactos é bem mais recente e mais focado nas diferenças socioeconômicas.

Além disso, existem evidências de que os sistemas educacionais podem ser usados para atenuar as desigualdades sociais. Assim, a qualidade educacional vem se tornando cada vez mais uma demanda da nossa sociedade. O que impulsionou o surgimento nos últimos anos de uma série de trabalhos que visam analisar o sistema educacional brasileiro, que não tem alcançado bons resultados nos principais sistemas de avaliação entre países, principalmente se focarmos apenas nas escolas públicas.

Ao analisar os resultados das últimas edições do PISA (Programa Internacional de Avaliação de Alunos) o resultado é preocupante. O desempenho escolar brasileiro se manteve estagnado nas últimas posições no exame apesar do crescimento dos investimentos em educação. De 2009 a 2018 houve um aumento de 116% nesses gastos, passando de R\$ 18 bilhões para R\$ 39 bilhões (INEP, 2018).

De acordo com os dados do PISA em 2018, o Brasil ficou com 57ª posição entre os 78 países participantes em leitura, atingindo os 413 pontos. Em ciências obteve a 66ª posição com 404 pontos. Já em matemática o rendimento foi ainda pior, 70ª posição com 384 pontos (INEP, 2018). Se comparado a 2015, houve uma discreta melhora. Naquele ano o país obteve 407 pontos em leitura, 401 em ciências e 377 em matemática, ficando com 59ª posição em leitura, 63ª em ciências e 66ª em matemática (INEP, 2015).

Diante desse cenário fica evidente que há muito espaço para a melhora dos sistemas escolares no Brasil. Além disso, verifica-se que o aumento de investimentos não necessariamente acarreta em melhora na qualidade. No caso brasileiro apesar de um aumento significativo da despesa, a qualidade escolar não evoluiu proporcionalmente.

Analisando o caso do Ceará, o estado se destaca no cenário nacional. Segundo dados do IDEB (Índice de Desenvolvimento da Educação Básica), em 2017 o Ceará atingiu a 3ª posição entre os estados para alunos no fim do ensino fundamental e a 5ª colocação para alunos no início do ensino fundamental, sendo um dos estados que mais cresceu nesses dois níveis (IDEB, 2018).

Tabela 1 – Crescimento IDEB observado entre 2007 e 2017

Estado/Região	5º ano	9º ano
Ceará	63%	46%
Nordeste	49%	33%
Brasil	40%	26%

Fonte: Inep

O objetivo deste trabalho é obter maior conhecimento do sistema educacional cearense e fornecer subsídios para medidas que visem melhorar o quadro, apontando os fatores responsáveis pelos escores dos alunos, identificando o percentual correspondente à contribuição da turma, da escola e do município em que o aluno está inserido. Utilizaremos, para isto, a relação entre o rendimento escolar dos alunos do 9º ano do ensino fundamental, medido pelo

escore padronizado de língua portuguesa e matemática, considerando as características socioeconômicas dos alunos, físicas das escolas, turmas e município.

O trabalho está dividido em outras cinco partes, além desta introdução. Na próxima parte será feita uma revisão da literatura focando na evolução dos trabalhos que analisam os fatores socioeconômicos, sociais e escolares no desempenho do aluno, e nos trabalhos que utilizam os modelos hierárquicos. Na terceira parte será feita uma breve análise descritiva dos dados. Na quarta parte será apresentada a metodologia. Na quinta analisamos os resultados obtidos. E na sexta e última parte a conclusão.

2 REVISÃO DA LITERATURA

Diversos fatores podem influenciar no desempenho escolar dos alunos. Fatores estes que são estudados há vários anos. O primeiro estudo apontado é o chamado *Relatório Coleman*, publicado em 1966 (COLEMAN et al., 1966), que analisou as causas para as diferenças de desempenho entre as escolas norte-americanas. Coleman concluiu que a origem social e o ambiente familiar determinavam o desempenho escolar e que as escolas pouco contribuíam para amenizar essas diferenças.

O relatório deu início a uma série de estudos sobre os possíveis efeitos das escolas nos rendimentos dos alunos. No entanto, novos estudos, como o *Fifteen Thousand Hours* (Rutter et al., 1979), questionam a pequena contribuição das escolas nos rendimentos dos alunos. Hanushek (2002) argumenta que o relatório e suas subsequentes interpretações confundem “mensurabilidade” e os verdadeiros efeitos das escolas sobre o aprendizado.

Estudos recentes mostram que a escola é responsável por uma variação considerável do rendimento dos alunos, ou seja, a escola tem impacto significativo na vida acadêmica dos alunos e melhores escolas tendem a ter melhores resultados. Essa constatação deu origem ao conceito de “escola eficaz”.

De acordo com Mortimore (1991, p.9):

Escola eficaz é aquela que o aluno progride além do esperado dadas as suas características ao serem admitidos.

As pesquisas sobre o efeito-escola passaram a defender a necessidade de entender os processos escolares associados ao desempenho dos alunos e identificar as práticas pedagógicas que levam alunos socioeconomicamente desfavorecidos a obterem melhores resultados acadêmicos (BRESSOUX, 2003).

Na literatura, os estudos utilizam várias metodologias para demonstrar os determinantes da proficiência dos alunos. Um avanço importante para os estudos sobre desempenho escolar foi a criação dos modelos hierárquicos. O modelo de Bryk e Raudesnbush (1992) leva em consideração a estruturação de agrupamento dos dados, refletindo no modelo. Segundo Goldstein (1995) o sistema educacional possui uma óbvia estrutura hierárquica, em que os alunos estão agrupados em turmas e as turmas em escolas. Estes modelos permitem separar a variabilidade dos resultados referentes às escolas da variabilidade dentro de cada escola.

No Brasil, após a adoção pelo SAEB da nova metodologia de construção do teste e análise dos resultados, que permitiu a comparabilidade entre os resultados ao longo do tempo, os trabalhos sobre eficácia escolar começaram a surgir com mais frequência. Alguns desses estudos utilizam o modelo hierárquico para analisar os dados da educação no país.

Barbosa e Fernandes (2000) utilizaram dados do SAEB coletados em 1997 referentes a alunos da 8ª série na região Sudeste. Foi utilizado um modelo de dois níveis, alunos e escolas, sendo a proficiência em matemática a variável dependente. Foram encontradas evidências de que as escolas fazem diferença nos resultados dos alunos e que a distribuição dos alunos pelas escolas não é aleatória, já que quando foram acrescentadas variáveis para controlar o nível socioeconômico do aluno diminuiu a variância entre escolas.

Dias, Navio e Ferrão (2006) testaram, através de um modelo de regressão multinível aplicados aos dados portugueses do PISA 2000, para testar as hipóteses de que alunos de estratos sociais desfavorecidos têm melhor desempenho em escolas públicas do que em escolas particulares com financiamento público ou sem financiamento público. Os resultados não foram significativos.

Em Soares e Mendonça (2003), através dos dados do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE) no ano 2000, foi realizada uma análise multinível com o objetivo de identificar a relação entre desempenho escolar, as características técnico-pedagógicas das escolas e o perfil socioeconômico dos alunos das 4ª séries do Ensino Fundamental. Foi identificado que a diferença entre as turmas e características das escolas afetam o rendimento dos alunos. Já em Palermo, Silva e Novellino (2014) foram utilizados modelos hierárquicos com três níveis, que permitiram avaliar características individuais dos alunos, turmas e escolas.

Em Soares (2003) também foi utilizado um modelo hierárquico com três níveis construído para explicar a proficiência dos alunos da 4ª série do ensino fundamental em língua portuguesa. Foram utilizados dados do Programa Mineiro de Avaliação da Educação Básica em 2002. Comprovou-se que um percentual da variação total é devido a diferenças entre turmas.

Outro trabalho que também utiliza um modelo hierárquico com três níveis foi desenvolvido por Almeida, Araújo Júnior e Ramalho (2016). Foram investigados os fatores associados ao desempenho de escolas públicas brasileiras, mostrando como o esforço empreendido na gestão escolar resulta em ganhos de aprendizagem para o aluno. Os resultados mostraram que a parcela referente aos alunos ainda explica maior parte da variabilidade das notas e que o esforço empreendido pela escola é relevante para os resultados dos alunos.

Moreira, Jacinto e Begolin (2017) investigaram os fatores associados ao desempenho escolar do estado do Rio Grande do Sul utilizando modelos hierárquicos e dados do Sistema de

Avaliação do Rendimento Escolar (SAERS). Ficou evidenciado que os fatores econômicos são fortemente associados ao rendimento escolar. Já as características das escolas pouco contribuem na explicação dos escores dos estudantes.

Tabela 2 – Resumo de trabalhos que utilizaram modelos hierárquicos para avaliar os sistemas educacionais no Brasil

AUTORES	DADOS	ANO	SERIE	LOCAL	NÍVEIS
Barbosa e Fernandes (2000)	SAEB	1997	8 ^a	Sudeste	2 (alunos e escolas)
Soares e Mendonça (2003)	SIMAVE	2000	4 ^a	Minas Gerais	3 (alunos, turmas e escolas)
Palermo, Silva e Novellino (2014)	SAEB	2007	5 ^a	Rio de Janeiro	3 (alunos, turmas e escolas)
Soares (2003)	SIMAVE	2002	4 ^a	Minas Gerais	3 (alunos, turmas e escolas)
Moreira, Jacinto e Begolin (2017)	SAERS	2007	5 ^a	Rio Grande do Sul	2 (alunos e escolas)
Almeida, Araújo Júnior e Ramalho (2016)	SAEB	2007,2009	5 ^a	Brasil	3 (alunos, escolas e municípios)

Fonte: Elaborada pelo autor

3. DADOS

3.1 Base de dados

Esta pesquisa utilizou dados transversais extraídos do Sistema Nacional de Avaliação da Educação Básica (SAEB) e do Censo Escolar. O SAEB é um conjunto de avaliações externas em larga escala, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que diagnostica a educação básica brasileira.

É aplicado de dois em dois anos e avalia os alunos do 9º ano do ensino fundamental e do 3º ano do ensino médio, nas disciplinas de matemática e língua portuguesa. Avalia a qualidade do ensino oferecido pelo sistema educacional brasileiro a partir de dois instrumentos: os testes e os questionários. Os testes buscam aferir o nível de conhecimento do aluno nas disciplinas e séries avaliadas. Os questionários buscam avaliar os fatores familiares e socioeconômicos relacionados aos alunos, bem como a qualificação de professores e diretores.

A população alvo é formada por alunos do 9º ano do ensino fundamental, matriculados em instituições públicas de ensino localizadas nos municípios do estado do Ceará, possibilitando analisar a diferença de impacto da turma, escola e município nos resultados dos alunos que estão no último ano do ensino fundamental.

Neste trabalho foram utilizadas as notas padronizadas dos alunos em língua portuguesa e matemática da Prova Brasil aplicada em 2017, assim como o questionário socioeconômico respondido por cada aluno, informações sobre as escolas através do questionário da escola e informações sobre diretor e professores, também através de questionário.

Já o Censo Escolar é o principal instrumento de coleta de informações da educação básica. É coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país (INEP, 2020). Utilizamos o Censo Escolar para complementar os dados referentes às escolas, visto que o mesmo não possui dados referentes a proficiência dos alunos, mas apresenta dados detalhados das escolas.

A grande dificuldade em relação a manipulação dos dados foi agrupar os questionários do SAEB (alunos, turmas, escolas, diretores e professores) e as tabelas do Censo Escolar (escola, turma, matrícula, docente), sem abrir mão de importantes variáveis.

Participaram da pesquisa 100.177 alunos, distribuído em 4.003 turmas, 2.236 escolas e 184 municípios, conforme Tabela 3.

Tabela 3 – Número de observações

Grupo de variáveis	Número de observações
Município	184
Escola	2.236
Turma	4.003
Aluno	100.177

Fonte: Elaborada pelo autor a partir de dados do Censo Escolar 2017 e SAEB 2017

3.2 Variáveis

As variáveis dependentes de interesse nesse estudo são a proficiência em língua portuguesa e a proficiência em matemática dos alunos. Partimos do princípio que as notas dos alunos nas duas disciplinas são influenciadas pelo papel da escola, pela turma e pelo município em que estão inseridos.

Já as variáveis independentes foram selecionadas conforme Soares (2003). Foi investigado o efeito de cada variável sobre as proficiências isoladamente e selecionadas as de maiores *deviance* entre as variáveis significantes. Para a análise empírica, as variáveis encontram-se agrupadas na Tabela 4. Estão ordenadas de acordo com o nível a qual pertencem. Primeiro aparecem as variáveis que indicam as características dos alunos, depois as que diferenciam características da turma e, por último, características do diretor e da escola na qual ele é o responsável.

Tabela 4 – Variáveis independentes

Variável	Descrição	Tipo	Valores
FEMININO	Sexo do aluno	Dummy	1 = feminino; 0 = masculino
NEGRO	Raça do aluno	Dummy	1 = negro; 0 = outras
ATRASO_IDADE	Aluno com idade acima da ideal para a série	Dummy	1 = sim; 0 = não
CLASSE_ABC ¹	Pertence à classe A, B ou C, de acordo com a renda familiar	Dummy	1 = sim; 0 = não
COMPUTADOR	Possui computador em casa	Dummy	1 = sim; 0 = não
DENS	Pessoa por quarto morando na casa do aluno	Numérica	
MORA_MAE	Mora com a mãe	Dummy	1 = sim; 0 = não
MAE_EM_COMP	Mãe completou o ensino médio	Dummy	1 = sim; 0 = não
MAE_ES_COMP	Mãe completou o ensino superior	Dummy	1 = sim; 0 = não
TRABALHA	Trabalha fora de casa	Dummy	1 = sim; 0 = não

¹ Indicador construído pela regra para classificação de domicílios a partir de dados de bens de consumo chamada Critério Brasil desenvolvido pela ABEP.

ENTROU_PRE_ESC	Entrou na escola até a pré-escola	Dummy	1 = sim; 0 = não
REPR	Já reprovou pelo menos uma vez	Dummy	1 = sim; 0 = não
GOSTA_EST_LP	Gosta de estudar língua portuguesa	Dummy	1 = sim; 0 = não
GOSTA_EST_MT	Gosta de estudar matemática	Dummy	1 = sim; 0 = não
DEVER_LP	Faz o dever de língua portuguesa	Dummy	1 = sim; 0 = não
DEVER_MT	Faz o dever de matemática	Dummy	1 = sim; 0 = não
MANHA	Estuda pelo turno da manhã	Dummy	1 = sim; 0 = não
MUNICIPAL	Escola com dependência administrativa municipal	Dummy	1 = sim; 0 = não
INSTALACOES	Escola com boas instalações	Dummy	1 = sim; 0 = não
NU_FUNCIONARIOS	Número de funcionários da escola	Numérica	
NU_EQUIP_AL	Quantidade de equipamentos que a escola possui por aluno do 9º ano	Numérica	
TEMPO_DIR	Diretor da escola exerce o cargo de diretor há mais de 5 anos	Dummy	1 = sim; 0 = não
TEMPO_DIR_ESC	Diretor da escola exerce o cargo de diretor na mesma escola há mais de 5 anos	Dummy	1 = sim; 0 = não
URBANA	Localização urbana da escola	Dummy	1 = sim; 0 = não

Fonte: Elaborada pelo autor

3.3 Estatísticas descritivas

A Tabela 5 apresenta as estatísticas descritivas referentes a variável dependente em ambas as proficiências. Podemos perceber que os alunos do estado do Ceará obtiveram uma melhor média e mediana na proficiência língua portuguesa. As médias em ambas as disciplinas são maiores que a média nacional se considerarmos apenas escolas públicas conforme utilizamos no trabalho, 251,68 em língua portuguesa e 250,07 em matemática (INEP, 2019). A amplitude das notas em matemática variou entre 124,99 a 429,64, enquanto a amplitude em língua portuguesa variou entre 127,01 e 374,36.

Tabela 5 – Análise descritiva da variável dependente

Proficiência	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Matemática	124,988	222,019	254,455	257,089	290,364	429,637
Língua Portuguesa	127,008	229,169	262,909	260,619	293,812	374,360

Fonte: Elaborada pelo autor

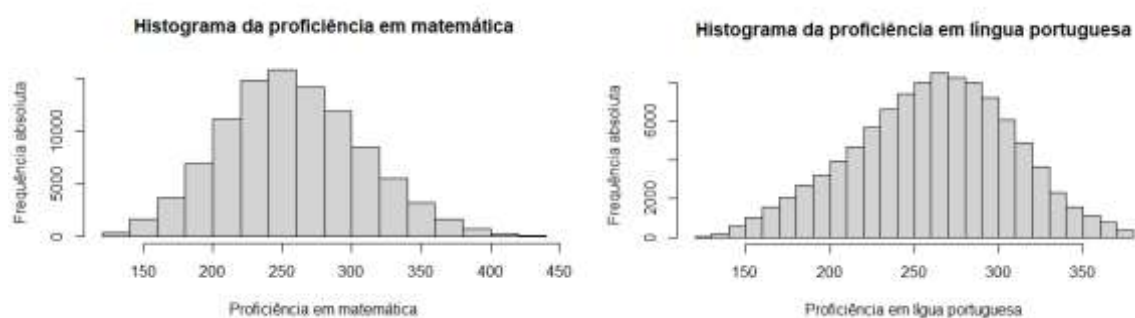


Gráfico 1: Histograma da proficiência em matemática e língua portuguesa

De acordo com a Figura 1 podemos observar que as notas de ambas as disciplinas possuem uma distribuição normal, sendo que a distribuição das notas em matemática apresenta uma leve assimetria positiva, já a de língua portuguesa apresenta assimetria negativa.

As variáveis independentes contendo características socioeconômicas e culturais dos alunos, físicas e qualitativas das escolas e individuais dos diretores apresentam estatísticas descritivas conforme a Tabela 6.

Tabela 6 – Estatísticas descritivas das variáveis independentes

VARIÁVEL	OBS	MÉDIA	DESVIO PADRÃO	MIN	MAX
PROFICIENCIA_MT_SAEB	100.177	257,089	50,173	124,988	429,638
PROFICIENCIA_LP_SAEB	100.177	260,619	46,826	127,008	374,360
FEMININO	96.276	0,508	0,500	0,000	1,000
NEGRO	96.230	0,088	0,283	0,000	1,000
ATRASO_IDADE	97.136	0,212	0,409	0,000	1,000
CLASSE_ABC ¹	81.652	0,541	0,498	0,000	1,000
COMPUTADOR	97.629	0,314	0,464	0,000	1,000
DENS	94.272	1,388	0,585	0,200	6,000
MORA_MAE_R	96.970	0,919	0,272	0,000	1,000
MAE_EM_COMP	96.836	0,191	0,393	0,000	1,000
MAE_ES_COMP	96.836	0,053	0,225	0,000	1,000
ENTROU_PRE_ESC	96.131	0,900	0,300	0,000	1,000
REPR	96.505	0,268	0,443	0,000	1,000
GOSTA_EST_MT	96.750	0,581	0,493	0,000	1,000
DEVER_MT	96.750	0,928	0,258	0,000	1,000
GOSTA_EST_LP	95.291	0,838	0,368	0,000	1,000
DEVER_LP	96.736	0,953	0,212	0,000	1,000
TRABALHA	96.884	0,094	0,292	0,000	1,000
TEMPO_DIR	108.465	0,358	0,480	0,000	1,000
TEMPO_DIR_ESC	108.381	0,147	0,355	0,000	1,000
NU_FUNCIONARIOS	108.399	54,382	28,864	5,000	330,000
NU_EQUIP_AL	108.399	0,250	0,233	0,000	2,250
MUNICIPAL	111.674	0,882	0,323	0,000	1,000
MANHA	111.674	0,421	0,494	0,000	1,000

URBANA	111.674	0,766	0,424	0,000	1,000
INSTALACOES	111.717	0,428	0,495	0,000	1,000

Fonte: Elaborada pelo autor

Podemos perceber que 8,8% dos alunos se declararam negros. Um pouco mais de 21,2% dos alunos apresentam idade acima da esperada para a série, ou seja, estão defasados em relação a idade correspondente à série. O percentual de alunos que pertencem às classes A, B ou C de acordo com a classificação da Associação Brasileira de Empresas de Pesquisa (ABEP) foi de 54%. Além disso 31,4% dos alunos possuem pelo menos um computador em casa.

Foi observado que a grande maioria dos alunos mora com a mãe, cerca de 92%. Em compensação a maioria das mães possui baixo grau de escolaridade. Apenas 19% completaram o ensino médio e 5% o ensino superior.

A maioria dos alunos começou a estudar na creche ou pré-escola, já que 90% dos alunos da amostra estraram na escola antes do ensino fundamental. Já foram reprovados pelo menos uma vez 27% dos alunos. A preferência da maioria é estudar língua portuguesa a matemática. E os alunos que trabalham fora de casa além de estudar somam 9%.

Com relação às variáveis dos diretores da escola, verifica-se que 35,8% deles exercem o cargo de diretor há mais de 5 anos. E 14,7% exercem esse cargo há mais de 5 anos na mesma escola. Já as variáveis no nível da escola nos mostram que em média as escolas têm 54 funcionários, sendo que a maior escola possui 330. A maioria delas são escolas municipais situadas na zona urbana. E 42,8% das escolas possuem boas instalações.

4 METODOLOGIA

4.1 Modelo multinível

Os modelos multiníveis são ferramentas estatísticas que são capazes de tratar como os fatores em um dado nível hierárquico afetam os resultados em níveis a eles subordinados. Os sistemas escolares são um exemplo de estrutura hierárquica já que os alunos estão agrupados em turmas, as turmas agrupadas em escolas, as escolas em municípios e assim por diante.

O modelo multinível leva em consideração a estrutura de agrupamento de dados. Diferentemente do modelo de regressão clássico, no modelo multinível os interceptos e os coeficientes de inclinação variam (GOLDSTEIN, 1995).

De acordo com Rasbash *et al.* (2009, p.3):

Ao focar a atenção nos níveis hierárquicos da população, a modelagem multinível permite ao pesquisador entender onde e como os efeitos estão ocorrendo. Ele fornece melhores estimativas em resposta às simples perguntas para as quais as análises de nível único já foram usadas e, além disso, permite que questões mais complexas sejam abordadas.

Devido a fatores como as diferenças socioeconômicas, geográficas e culturais, a distribuição dos alunos entre as escolas não é aleatória. No processo de entrada na escola essa seletividade contribui para que alunos de uma determinada escola sejam mais parecidos entre si do que se comparados a alunos de outras escolas, assim como alunos da mesma turma, devido a convivência diária e por estarem submetidos aos mesmos processos de ensino e professores, possuem desempenho mais parecido do que alunos de turmas diferentes. Isto quebra o pressuposto do modelo de regressão linear de que os erros são independentes e identicamente distribuídos.

Conforme Gelman e Hill (2007), os modelos de regressão multinível podem ser usados para uma variedade de objetivos, incluindo inferência causal, predição e modelagem descritiva. Entre os objetivos, os autores destacam: i) estudar como os efeitos variam entre indivíduos de grupos diferentes; ii) analisar dados estruturados; iii) obter inferências mais eficientes para os parâmetros da regressão. Já segundo Soares, César e Mambrini (2001) a utilização de um modelo que incorpora a estrutura hierárquica dos dados apresenta vantagens como: i) obtenção de melhores estimativas para os parâmetros relativos a unidades específicas, pois por meio do modelo hierárquico é possível obter uma equação para cada escola, fazendo assim uso de toda a informação presente na amostra de forma eficiente; ii) possibilidade de formular e testar hipóteses relativas a efeitos entre níveis; iii) permite a partição da variabilidade da variável resposta nos diversos níveis.

Todavia a utilização desse tipo de modelagem deve se dar apenas quando a estrutura de dados apresentar uma clara hierarquia, como é caso dos sistemas escolares. Nestes casos existe um certo grau de agrupamento. Quando não existe grau de agrupamento os modelos de regressão clássica podem ser usados com sucesso.

Os modelos multinível de componentes de variância (Longford, 1987) e o de coeficientes aleatórios (Leeuw e Kreft, 1986; Longford, 1993) são os mais utilizados. No primeiro, somente o intercepto é aleatório enquanto a variância da variável dependente é decomposta para cada nível hierárquico (efeitos fixos). Já no segundo testa o efeito randômico dos coeficientes de inclinação (efeitos aleatórios)

4.2 Modelo econométrico

O modelo de regressão multinível, especificado em quatro níveis, utiliza o aluno como a unidade do nível 1, identificado por i , a turma como unidade do nível 2, identificada por j , a escola como a unidade do nível 3, identificada por k e o município como unidade do nível 4, identificado por l .

A variável dependente de interesse neste estudo é a proficiência dos alunos do 9º ano do ensino fundamental nos municípios cearenses nas disciplinas de matemática e língua portuguesa.

Segundo Byrk e Raudenbush (1992), o modelo sem nenhuma variável explicativa é dado por:

$$\begin{aligned} y_{ijkl} &= \beta_{0jkl} + e_{ijkl}, e_{ijkl} \sim N(0, \sigma_e^2) \\ \beta_{0jkl} &= \gamma_{00kl} + u_{0jkl}, u_{0jkl} \sim N(0, \sigma_{u0}^2) \\ \gamma_{00kl} &= \pi_{000l} + r_{00kl}, r_{00kl} \sim N(0, \sigma_{r00}^2) \\ \text{e } \pi_{000l} &= \theta_{0000} + v_{000l}, v_{000l} \sim N(0, \sigma_{v000}^2) \end{aligned}$$

θ_{0000} : representa a média global da proficiência.

v_{000l} : representa o afastamento da proficiência média do município l à média global.

r_{00kl} : representa o afastamento da proficiência média da escola k à média do município l .

u_{0jkl} : representa o afastamento da proficiência média da turma j à média da escola k .

e_{ijkl} : representa o afastamento da proficiência do aluno i em relação à média da turma j .

A primeira equação representa o nível 1, a segunda o nível 2, e assim por diante. Cada equação tem associada uma componente aleatória. Assumimos que os efeitos seguem distribuição normal com média 0 e variância dos parâmetros aleatórios do modelo σ_e^2 , σ_{u0}^2 , σ_{r00}^2 , σ_{v000}^2 , e são independentes.

Conforme Goldstein (1995), um indicador para o grau de agrupamento da população em estudo é o coeficiente de intra-correlação. O coeficiente mede a proporção da variância de um determinado nível em relação a variância total. Devido à hipótese de não correlação entre os erros nos diferentes níveis, a proporção da variância de cada nível é dada por:

$$\frac{\sigma_e^2}{\sigma_e^2 + \sigma_{u0}^2 + \sigma_{r00}^2 + \sigma_{v000}^2}, \text{ para o nível 1}$$

$$\frac{\sigma_{u0}^2}{\sigma_e^2 + \sigma_{u0}^2 + \sigma_{r00}^2 + \sigma_{v000}^2}, \text{ para o nível 2}$$

$$\frac{\sigma_{r00}^2}{\sigma_e^2 + \sigma_{u0}^2 + \sigma_{r00}^2 + \sigma_{v000}^2}, \text{ para o nível 3}$$

$$\frac{\sigma_{v000}^2}{\sigma_e^2 + \sigma_{u0}^2 + \sigma_{r00}^2 + \sigma_{v000}^2}, \text{ para o nível 4}$$

Cada coeficiente varia entre 0 e 1. Quando o valor do coeficiente é nulo, significa que as turmas, escolas ou municípios são homogêneos entre si, dependendo de qual coeficiente estamos observando. Quanto mais próximo de 1 é o coeficiente, maior o impacto da escola, turma ou município no resultado acadêmico no aluno.

Substituindo a equação 4 na equação 3, depois na equação 2 e depois na equação 1, obtemos:

$$y_{ijkl} = \theta_{0000} + v_{000l} + r_{00kl} + u_{ojkl} + e_{ijkl}$$

O modelo hierárquico para os quatro níveis pode ser generalizado a partir das equações 1 e 2 (BRYK; RAUDENBUSH, 1992).

$$y_{ijkl} = \beta_{0jkl} + \sum_{f=1}^F \beta_f X_{fijkl} + \sum_{s=1}^S \gamma_s W_{sjkl} + \sum_{t=1}^T \pi_t Z_{tkl} + \sum_{p=1}^P \theta_p Q_{pl} + e_{ijkl} \quad (1)$$

$$\beta_{0jkl} = \theta_{0000} + v_{000l} + r_{00kl} + u_{0jkl} \quad (2)$$

Onde:

$f = 1, \dots, F$ variáveis do nível 1;

$s = 1, \dots, S$ variáveis do nível 2;

$t = 1, \dots, T$ variáveis do nível 3;

$p = 1, \dots, P$ variáveis do nível 4;

$i = 1, \dots, I_{jkl}$ alunos na j -ésima turma da k -ésima escola no l -ésimo município;

y_{ijkl} : proficiência do aluno i que estuda na turma j da escola k no município l ;

θ_{0000} : média global da proficiência;

β_{0jkl} : intercepto da turma j da escola k no município l ;

X_{fijkl} : variáveis associadas às características dos alunos;

W_{sjkl} : variáveis associadas às características das turmas;

Z_{tkl} : variáveis associadas às características das escolas;

Q_{pl} : variáveis associadas às características dos municípios;

$\beta_f, \gamma_s, \pi_t, \theta_p$: efeitos das variáveis dos alunos, turmas, escolas e municípios, respectivamente;

$e_{ijkl}, u_{0jkl}, r_{00kl}, v_{000l}$: efeitos aleatórios dos níveis 1, 2, 3 e 4, respectivamente;

$e_{ijkl} \sim N(0, \sigma_e^2), u_{0jkl} \sim N(0, \sigma_{u0}^2), r_{00kl} \sim N(0, \sigma_{r00}^2), v_{000l} \sim N(0, \sigma_{v000}^2)$.

Para o modelo multinível pressupõe-se que: i) O termo de erro do nível 1, e_{ijkl} , é independente das variáveis explicativas do modelo, ou seja, $\text{cov}(X_{fijkl}, e_{ijkl}) = 0$, $\text{cov}(W_{sjkl}, e_{ijkl}) = 0$, $\text{cov}(Z_{tkl}, e_{ijkl}) = 0$ e $\text{cov}(Q_{pl}, e_{ijkl}) = 0$; ii) O termo de erro dos demais níveis também são não correlacionados com as variáveis independentes; iii) Os termos de erros referentes aos níveis são independentes entre si.

Esta estrutura corresponde a um modelo multinível com intercepto aleatório, ou seja, o intercepto se divide em dois elementos: um fixo e outro aleatório. O elemento fixo, γ_{00} , representa o intercepto global, que é constante para todas as escolas e corresponde à média dos interceptos β_{0j} . O termo V_j , a parte aleatória, representa o quanto a escola j está distante do intercepto global. A componente aleatória é interpretada como efeito único da escola j em relação à média global.

Para Soares e Candian (2007) a denominação mais adequada para este indicador seria “índice de dissimilaridade entre escolas”, já que valores próximos de 1 indicam que as escolas

são muito diferentes em termos do desempenho de seus alunos, logo o efeito da escola é elevado. Valores próximos de zero indicam que as escolas são similares entre si e nesse caso, o uso de uma regressão usual seria justificável.

5 RESULTADOS

Para elaboração do modelo foi utilizado o processo apresentado por Hox (2002). O processo parte do modelo nulo, ou seja, sem nenhuma variável explicativa, e depois são adicionadas as variáveis do primeiro nível capazes de afetar o desempenho. O próximo passo é adicionar também as variáveis do segundo nível, e assim por diante até chegarmos ao modelo com todas as variáveis com efeitos fixos. Dessa forma podemos analisar a evolução da explicação alcançada após a introdução de cada variável. No próximo passo incluímos as variáveis com efeitos aleatórios ao modelo. Por último recomenda-se verificar as interações entre as variáveis.

Em cada passo são analisadas as estimativas e as variâncias para ver quais parâmetros são significantes e quanto sobra de erro residual em cada nível. Como o tamanho da amostra é maior no menor nível, faz sentido começar a adicionar variáveis a partir dele. E como os parâmetros fixos são estimados com mais precisão, começamos com a estimação dos coeficientes fixos e adicionamos os componentes de variância nos estágios mais avançados (HOX, 2002).

Para comparar os modelos utilizamos a estatística *deviance*, que é uma medida relativa de grau de ajustamento dos dados aos modelos. Quanto menor a *deviance* de um modelo em relação a outro, melhor é o modelo. Para testar se a diferença entre dois modelos é significativa, utilizamos o teste de significância χ^2 com o número de graus de liberdade igual à diferença de parâmetros entre os dois modelos.

5.1 Modelo nulo

O percentual de explicação da variação dos escores de matemática e língua portuguesa relativo ao modelo nulo está descrito conforme a Tabela 6. Em um modelo com apenas dois níveis, Barbosa e Fernandes (2000) encontraram um Índice de Correlação Intraclasse (ICC) entre as escolas para o Brasil de 37%, a grande participação percentual das escolas nos resultados é devido a amostra utilizar dados de todos os estados brasileiros e também incluir escolas particulares, o que aumenta a heterogeneidade entre as escolas.

No caso, como utilizamos um modelo de 4 níveis, em que os alunos estão inseridos em turmas, as turmas em escolas e escolas em municípios, parte da variabilidade que seria imputada apenas às escolas em um modelo de dois níveis é reduzida pela variabilidade das notas médias entre os diferentes municípios, indicando que as escolas do mesmo município possuem maior

homogeneidade. O fato de a amostrar conter apenas escolas públicas do estado do Ceará também contribui para diminuir a variabilidade entre escolas. Mesmo assim o efeito-escola, que é o percentual do resultado do aluno atribuído às diferenças entre as escolas, foi significativo, 7,69% em matemática e 4,62% em língua portuguesa.

O efeito-turma, que mostra que parte das variabilidades dos resultados entre escolas é de fato causado pela diferença entre turmas, principalmente devido a importância do professor, se mostrou mais representativo para a disciplina de língua portuguesa, indicado pelo maior percentual correspondente ao resultado devido às diferenças entre turmas a esta disciplina. Conforme Soares (2005), alunos em uma mesma turma tendem a ter resultados mais semelhantes se comparados a alunos de turmas diferentes. Em língua portuguesa a variação entre turmas foi de 6,07% enquanto em matemática foi de 5,26%.

Já o efeito do município em que a escola está inserida é maior para a disciplina matemática. Nos chama a atenção a variabilidade da nota média entre os municípios em matemática, evidenciando uma relevante heterogeneidade entre a qualidade do ensino na disciplina entre os diferentes municípios do estado do Ceará. Nota-se que o impacto da heterogeneidade entre escolas e municípios é mais expressivo para a disciplina de matemática. Quanto mais homogêneo os municípios de um estado são em relação as suas políticas e investimentos educacionais, menor será esse percentual.

A maior parte da variância dos escores ocorre devido as diferenças entre os alunos nas turmas. No modelo, essa variabilidade corresponde a variância residual estimada. Para as notas de matemática, foi de 78,33%, e 85,26% em língua portuguesa.

Tabela 7 – Índice de correlação interclasse para o modelo nulo

Fonte da variação	Matemática %	Língua Portuguesa %
Município	8,72%	4,05
Escola	7,69%	4,62
Turma	5,26%	6,07
Aluno	78,33%	85,26

Fonte: Elaborada pelo autor

Na Tabela 8 encontram-se os resultados obtidos através do modelo nulo para a proficiência em matemática. O valor do intercepto estimado foi de 257,69, indicando a média global da proficiência para esse modelo, média essa que varia entre municípios, escolas e turmas. O teste de Wald revelou que as variâncias são significativamente diferentes de zero,

uma vez que os valores da razão entre a variância e o erro-padrão (razão t) são maiores que o valor crítico de 1,96.

Tabela 8 – Proficiência em matemática no modelo nulo

Efeitos fixos	Coefficiente	Erro-padrão	Z	P-valor
Desempenho escolar	257,6867	1,2082	213,2900	0,0000
Efeitos aleatórios	Estimativa	Erro-padrão	Razão t	
Variância município	217,8634	27,1373	8,03	
Variância escola	192,1653	11,3475	16,93	
Variância turma	131,3565	6,8272	19,24	
Variância aluno	1956,3770	8,9233	219,24	

Teste LR vs modelo linear: $\chi^2(3) = 19.095,37$ Prob > $\chi^2 = 0,0000$

Fonte: Elaborada pelo autor

Em língua portuguesa, a Tabela 9 mostra que para o modelo nulo a estimativa média global da proficiência ficou em 258,88 pontos. Assim como em matemática, o teste de Wald revela que as variâncias são significativamente diferentes de zero.

O fato dos ICC's encontrados serem significativos indica que existem diferenças em relação ao desempenho médio de turmas, escolas e municípios em relação à média estadual, justificando que se proceda à análise dos dados utilizando os modelos de regressão multinível em detrimento do modelo tradicional, visto que existe correlação nas notas dos alunos em cada nível do modelo (GOLDSTEIN, 1995).

Tabela 9 – Proficiência em língua portuguesa no modelo nulo

Efeitos fixos	Coefficiente	Erro-padrão	Z	P-valor
Desempenho escolar	258,8849	0,8173	316,7600	0,0000
Efeitos aleatórios	Estimativa	Erro-padrão	Razão t	
Variância município	88,4119	12,2427	7,22	
Variância escola	101,0101	7,6963	13,12	
Variância turma	132,5964	6,4252	20,64	
Variância aluno	1862,5130	8,4926	219,31	

Teste LR vs modelo linear: $\chi^2(3) = 10.802,09$ Prob > $\chi^2 = ,0000$

Fonte: Elaborada pelo autor

5.2 Modelos com variáveis no nível do aluno (efeitos fixos)

O Modelo 2, controlado pelas variáveis que indicam o nível socioeconômico e cultural dos alunos, foi construído com a inclusão de variáveis referentes ao nível do aluno no modelo nulo. Foi assumido que o efeito das características individuais dos alunos sobre a proficiência é o mesmo entre as escolas, ou seja, os coeficientes das variáveis independentes são fixos.

A partição das variâncias, ou ICC, em matemática e língua portuguesa se dá de acordo com a Tabela 10. Após a inclusão das variáveis independentes no nível do aluno, a importância da turma nos escores diminuiu se comparada ao modelo nulo.

Tabela 10 – Partição da variância do modelo com variáveis explicativas no nível do aluno

Fonte da variação	Matemática	Língua Portuguesa
	%	%
Município	8,93%	4,24%
Escola	7,50%	4,19%
Turma	3,82%	3,84%
Aluno	79,75%	87,72%

Fonte: Elaborada pelo autor

As segunda e quinta colunas da Tabela 11 trazem os resultados do Modelo 2 para as disciplinas de matemática e língua portuguesa, respectivamente. Todas as variáveis do Modelo 2 para a disciplina de matemática foram estatisticamente significantes a 1%. Já para a disciplina língua portuguesa, a variável classe social foi a única que não se mostrou estatisticamente significativa.

Conforme podemos perceber após comparar os dois modelos, controlados pelas variáveis socioeconômicas e culturais, houve uma diminuição das variâncias entre escolas em torno de 17% para a disciplina de matemática e 23% em língua portuguesa. Esta constatação nos informa que a distribuição dos alunos entre as escolas não é aleatória.

Tabela 11 – Modelo 2 com variáveis independentes no nível do aluno

Variáveis Independentes	MODELO 2 - MATEMÁTICA			MODELO 2 - LÍNGUA PORTUGUESA		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
INTERCEPTO	252,45	1,55	0,000	247,31	1,39	0,000
FEMININO	-14,81	0,33	0,000	5,35	0,33	0,000
NEGRO	-4,84	0,58	0,000	-4,82	0,57	0,000
ATRASO_IDADE	-11,88	0,53	0,000	-9,10	0,53	0,000
CLASSE_ABC	1,55	0,37	0,000	-0,13	0,37	0,718
COMPUTADOR	6,19	0,38	0,000	8,00	0,37	0,000

DENS	-1,70	0,29	0,000	-2,53	0,29	0,000
MORA_MAE_R	2,28	0,60	0,000	2,51	0,60	0,000
MAE_EM_COMP	6,70	0,44	0,000	7,98	0,44	0,000
MAE_ES_COMP	9,03	0,76	0,000	7,26	0,75	0,000
ENTROU_PRE_ESC	7,33	0,55	0,000	7,64	0,55	0,000
REPR	-15,91	0,49	0,000	-18,46	0,48	0,000
GOSTA_EST_MT / GOSTA_EST_LP	13,32	0,35	0,000	4,11	0,45	0,000
DEVER_MT/DEVER_LP	4,29	0,68	0,000	7,12	0,80	0,000
TRABALHA	-4,71	0,58	0,000	-9,98	0,57	0,000
Efeitos aleatórios	Estimativa	Erro- padrão		Estimativa	Erro- padrão	
Variância município	190,08	23,61		78,20	10,63	
Variância escola	159,66	9,97		77,43	6,46	
Variância turma	81,27	5,95		70,95	5,35	
Variância aluno	1697,16	9,43		1619,19	9,10	
	LR test vs. linear model: $\chi^2(3) = 11053,27$ Prob > $\chi^2 = 0,00002$			LR test vs. linear model: $\chi^2(3) = 5096,07$ Prob > $\chi^2 = 0,0000$		
<i>Deviance</i>	712.176			693.142		
Número de parâmetros estimados	20			20		

Fonte: Elaborada pelo autor.

A variável sexo mostra que em média as notas das meninas são 14,81 pontos inferiores à dos meninos em matemática. Enquanto em língua portuguesa o cenário se inverte. As notas das meninas são 5,35 pontos maiores que a dos meninos. Fatos recorrentes na literatura devido às diferenças existentes entre o desenvolvimento cognitivo dos meninos e das meninas. No que se refere à raça, os alunos que se declararam negros obtiveram um desempenho inferior às outras raças, indo ao encontro do que aponta Soares (2005). Fato provavelmente relacionado à condição econômica desses alunos.

Conforme o esperado a defasagem escolar tem impacto negativo nos resultados e é uma das variáveis que apresenta maior diferença entre os alunos. Nas disciplinas de matemática e língua portuguesa os alunos defasados obtiveram, respectivamente, um escore de 11,88 e 9,10 pontos em média abaixo dos que possuem a idade correspondente para a série em análise.

O fato de o aluno pertencer às classes A, B ou C, de acordo com o Critério Brasil da ABEP, representou um ganho de 1,55 ponto em média em matemática em relação aos alunos pertencentes às classes D e E pela mesma classificação. Já o fator classe social não apresentou significância estatística para a disciplina língua portuguesa. Possuir pelo menos um computador em casa apresenta ganhos expressivos em termos de pontuação nas duas disciplinas. O fato de possuir um computador possibilita ao aluno obter informações que o auxiliam nos estudos e também é um indicativo do nível socioeconômico do aluno.

Foi utilizada a variável DENS para indicar quantas pessoas por quarto moram na casa do aluno. O objetivo é descobrir como a densidade demográfica na casa do aluno se reflete no aprendizado. Pelo Modelo 2 podemos perceber que quanto maior essa razão, pior o rendimento do aluno. É importante para o aluno ter uma casa mais tranquila e com espaço. A cada 1 unidade de acréscimo nessa razão, em média o rendimento cai 1,70 em matemática e 2,53 em português.

Morar com a mãe gera resultados positivos nas notas dos alunos em ambas as disciplinas, evidenciando o papel fundamental da mãe para um melhor rendimento do aluno. Destaca-se também o acréscimo no resultado devido à escolaridade da mãe, chegando a 9,03 pontos em média a mais em matemática para alunos que possuem mãe com ensino superior completo. Observamos uma maior importância da escolaridade da mãe no que se refere a disciplina de matemática.

Alunos que entraram na escola antes do ensino fundamental obtiveram melhores resultados comparados aos que entraram depois. O incremento nas notas das duas disciplinas foi semelhante, em torno de 7 pontos em média a mais. Já o fato de gostar da disciplina tem um impacto maior em matemática. Um aluno que gosta de matemática tem um acréscimo médio de 13,32 pontos em relação a um que não gosta. Em português o impacto é de 4,11 pontos.

Estudantes que são mais comprometidos e disciplinados, isto é, fazem o dever de casa, atingem melhores resultados em média. Esse maior comprometimento é mais importante para o aprendizado da disciplina de português do que da de matemática, em média 7,12 contra 4,29 pontos no resultado.

A variável que apresentou maior magnitude em relação aos resultados foi a que indica se o aluno reprovou alguma vez ou não. Alunos que já reprovaram obtiveram resultados piores em ambas as disciplinas. Em matemática, o resultado foi em média 15,91 pontos pior que os alunos que nunca reprovaram. Em língua portuguesa a diferença é ainda maior, 18,46 pontos em média.

Outro resultado também esperado, foi o fato de alunos que trabalham fora de casa apresentarem escores 4,71 pontos a menos em média que os alunos que não trabalham em matemática e 9,98 abaixo em língua portuguesa. Essa variável está relacionada ao nível socioeconômico do aluno.

5.3 Modelos com variáveis no nível do aluno, da turma e da escola (efeitos fixos)

O próximo modelo é obtido incluindo as variáveis do nível da escola no modelo anterior, com finalidade de explicar a variabilidade das escolas e turmas, assumindo que os efeitos das

variáveis sobre a nota é o mesmo entre as escolas (efeitos fixos). As variáveis incluídas no modelo foram: TEMPO_DIR, TEMPO_DIR_ESC, NU_FUNCIONARIOS_DIV, NU_EQUIP_AL, MUNICIPAL, MANHA, URBANA e INSTALACOES.

Após as inclusões dessas variáveis, os ICC's pouco se alteraram, conforme mostra a Tabela 12.

Tabela 12 - Partição da variância do Modelo 3

Fonte da variação	Matemática	Língua Portuguesa
	%	%
Município	8,73%	4,22%
Escola	7,48%	3,92%
Turma	3,77%	3,68%
Aluno	80,02%	88,18%

Fonte: Elaborada pelo autor.

No Modelo 3 as variáveis no nível do aluno continuaram significativas, excetuando a variável CLASSE_ABC que permanece não significava para a disciplina língua portuguesa. Já a única variável no nível da turma, indicando se a turma pertence ao turno da manhã ou não, apresentou um impacto positivo na nota média dos alunos nas duas disciplinas. Turmas no turno matutino apresentam melhores resultados.

Escolas com diretores com mais de 5 anos de experiência obtiveram resultados piores tanto em matemática como em língua portuguesa, estatisticamente significantes a 5% e 10% respectivamente, indicando que possivelmente diretores com métodos mais modernos conseguem melhorar o desempenho da escola. Já quando o diretor está a pelo menos 5 anos no cargo na mesma escola, os resultados passam a serem positivos, mostrando a importância da continuidade do trabalho do diretor.

A quantidade de funcionários apresenta impacto positivo em média nos escores de matemática e língua portuguesa, ou seja, escolas maiores atingem melhores resultados. Já a quantidade de equipamentos por aluno pertencentes à escola é significativa nas duas disciplinas, indicando que escolas que possuem mais equipamentos por aluno, conseguem em média atingir melhores resultados.

A variável do nível da escola que exerce maior impacto nos escores de língua portuguesa é a MUNICIPAL. Escolas com dependência municipal em média obtém 5,00 pontos a mais na

prova de matemática que escolas com dependência estadual e federal, e 6,36 pontos a mais em língua portuguesa.

A variável que indica a qualidade das instalações da escola e a que indica se a escola está localizada na área urbana ou rural, não foram significativas. Esperava-se que escolas com instalações em melhores condições e as localizadas em área urbana alcançassem melhores resultados.

Tabela 13 – Modelo 3 com variáveis independentes no nível do aluno, turma e escola

Variáveis Independentes	MODELO 3 – MATEMÁTICA			MODELO 3- LÍNGUA PORTUGUESA		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
INTERCEPTO	241,49	2,65	0,000	234,98	2,18	0,000
FEMININO	-14,70	0,33	0,000	5,32	0,33	0,000
NEGRO	-4,72	0,59	0,000	-4,77	0,58	0,000
ATRASO_IDADE	-11,91	0,54	0,000	-9,10	0,54	0,000
CLASSE_ABC	1,48	0,38	0,000	-0,29	0,37	0,445
COMPUTADOR	6,10	0,39	0,000	7,89	0,38	0,000
DENS	-1,73	0,30	0,000	-2,51	0,29	0,000
MORA_MAE_R	2,07	0,62	0,001	2,31	0,61	0,000
MAE_EM_COMP	6,76	0,45	0,000	8,04	0,45	0,000
MAE_ES_COMP	9,05	0,77	0,000	7,27	0,76	0,000
ENTROU_PRE_ESC	7,29	0,57	0,000	7,56	0,56	0,000
REPR	-15,80	0,50	0,000	-18,41	0,49	0,000
GOSTA_EST_MT / GOSTA_EST_LP	13,37	0,36	0,000	4,02	0,46	0,000
DEVER_MT/DEVER_LP	4,57	0,70	0,000	7,18	0,82	0,000
TRABALHA	-4,62	0,59	0,000	-9,72	0,58	0,000
TEMPO_DIR	-2,36	1,03	0,021	-1,52	0,81	0,061
TEMPO_DIR_ESC	3,90	1,39	0,005	2,93	1,10	0,008
NU_FUNCIONARIOS	0,07	0,02	0,000	0,07	0,01	0,000
NU_EQUIP_AL	7,10	1,65	0,000	4,94	1,36	0,000
MUNICIPAL	5,00	1,82	0,006	6,36	1,40	0,000
MANHA	3,19	0,58	0,000	3,56	0,53	0,000
URBANA	-0,40	1,01	0,694	0,68	0,82	0,403
INSTALACOES	-0,10	0,81	0,905	0,51	0,64	0,424
Efeitos aleatórios	Estimativa	Erro-padrão		Estimativa	Erro-padrão	
Variância município	185,53	23,43		77,41	10,77	
Variância escola	158,94	10,33		71,96	6,53	
Variância turma	80,05	6,08		67,63	5,40	
Variância aluno	1700,70	9,62		1618,68	9,26	
	LR test vs. linear model: chi2(3) = 10002,22 Prob > chi2 = 0,0000			LR test vs. linear model: chi2(3) = 4417,17 Prob > chi2 = 0,0000		
<i>Deviance</i>	685.968			667.125		

Número de parâmetros estimados	27	27
Diferença Devianciace: M2 - M3	26.208	26.017
Diferença de parâmetros: M2 - M3	7	7
Teste χ^2	3.744	3.717

Fonte: Elaborada pelo autor.

Analisando a diferença entre o *deviance* dos dois modelos, podemos perceber que o Modelo 3 se ajusta melhor aos dados do que o Modelo 2 em ambas as disciplinas. Após o aplicar o teste de significância χ^2 , observamos que os modelos 2 e 3 são significativamente diferentes. Se diferença dos *deviances* for significativa, o modelo com o menor *deviance* é considerado melhor (KREFT; DE LEEUW, 1998).

5.4 Modelos com variáveis no nível do aluno, da turma e da escola (efeitos fixos e aleatórios)

Seguindo com os modelos, tentamos descobrir se algumas variáveis apresentavam inclinações aleatórias. Devido a importância de algumas variáveis que indicam características socioeconômicas e culturais dos alunos, podemos supor que os efeitos de algumas delas não são os mesmos entre as escolas. As variáveis foram testadas e selecionadas as que apresentaram significância estatística, são elas: MAE_ES_COMP, ATRASO_IDADE e REPR.

Os resultados seguem conforme Tabela 14.

Tabela 14 – Modelo 4 com variáveis independentes no nível do aluno, turma e escola, e efeitos aleatórios.

Variáveis Independentes	MODELO 4 – MATEMÁTICA			MODELO 4 - LÍNGUA PORTUGUESA		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
INTERCEPTO	241,49	2,65	0,000	235,08	2,18	0,000
FEMININO	-14,73	0,33	0,000	5,31	0,33	0,000
NEGRO	-4,73	0,59	0,000	-4,77	0,58	0,000
ATRASO_IDADE	-12,19	0,57	0,000	-9,35	0,57	0,000
CLASSE_ABC	1,49	0,38	0,000	-0,27	0,37	0,463
COMPUTADOR	6,10	0,39	0,000	7,90	0,38	0,000
DENS	-1,72	0,30	0,000	-2,49	0,29	0,000
MORA_MAE_R	2,02	0,62	0,001	2,28	0,61	0,000
MAE_EM_COMP	6,76	0,45	0,000	8,05	0,45	0,000
MAE_ES_COMP	8,77	0,87	0,000	7,23	0,80	0,000

ENTROU_PRE_ESC	7,28	0,57	0,000	7,60	0,56	0,000
REPR	-16,02	0,52	0,000	-18,55	0,51	0,000
GOSTA_EST_MT / GOSTA_EST_LP	13,39	0,36	0,000	4,01	0,46	0,000
DEVER_MT/DEVER_LP	4,53	0,70	0,000	7,14	0,82	0,000
TRABALHA	-4,59	0,59	0,000	-9,69	0,58	0,000
TEMPO_DIR	-2,32	1,02	0,023	-1,52	0,81	0,060
TEMPO_DIR_ESC	3,97	1,39	0,004	2,97	1,10	0,007
NU_FUNCIONARIOS	0,07	0,02	0,001	0,07	0,01	0,000
NU_EQUIP_AL	7,12	1,65	0,000	4,93	1,35	0,000
MUNICIPAL	5,32	1,83	0,004	6,42	1,40	0,000
MANHA	3,29	0,57	0,000	3,63	0,52	0,000
URBANA	-0,40	1,01	0,690	0,62	0,82	0,450
INSTALACOES	0,00	0,81	0,998	0,54	0,64	0,404
Efeitos aleatórios	Estimativa	Erro- padrão		Estimativa	Erro- padrão	
Variância município	186,19	23,48		77,72	10,80	
Variância escola	155,37	10,29		69,52	6,51	
var(MAE_ES~P)	124,20	30,09		37,01	24,18	
var(ATRASO~E)	34,52	11,30		36,36	11,06	
var(REPR)	26,32	9,71		17,94	8,67	
Variância turma	78,56	6,05		66,33	5,37	
Variância aluno	1685,61	9,74		1608,95	9,38	
	LR test vs. linear model: $\chi^2(6) = 10064,51$ Prob > $\chi^2 = 0,0000$			LR test vs. linear model: $\chi^2(6) = 4452,82$ Prob > $\chi^2 = 0,0000$		
<i>Deviance</i>	685.905			667.090		
Número de parâmetros estimados	30			30		
Diferença Deviance: M3 - M4	62			36		
Diferença de parâmetros: M3 - M4	3			3		
Teste χ^2	21			12		

Fonte: Elaborada pelo autor.

Podemos notar que a educação materna tem uma alta variabilidade entre as escolas. Indicando, conforme testado por Rios-Neto, César e Riani (2002), que essa variável apresenta efeitos aleatórios, ajudando a explicar a razão da variabilidade do efeito intra-escola.

O Modelo 4 apresenta menor *deviance* em relação ao Modelo 3, e com isso é o modelo que melhor se ajusta aos dados entre os analisados. Após o Modelo 4 tentamos estimar um último modelo no qual as haveria interação entre as variáveis de efeitos aleatórios e as variáveis no nível da escola, porém as interações não apresentaram significância estatística.

6 CONCLUSÃO

Este trabalho procurou contribuir para um melhor entendimento dos determinantes do sistema educacional cearense, ratificando muitas conclusões de estudos anteriores, que utilizaram amostras diferentes. Por utilizar um modelo hierárquico mais complexo, subdividido em 4 níveis, pudemos analisar o real impacto das diferenças entre turmas, escolas e municípios nos resultados dos alunos. A maioria dos estudos, por não utilizarem esse nível de agregação, acabam incorporando a um determinado nível hierárquico as variações pertencentes a um nível não discriminado.

Foram utilizadas também variáveis explicativas ainda não estudadas. A quantidade de pessoas que moram na casa do aluno por quarto, por exemplo, utilizada como um indicador de densidade, diminui o resultado médio do aluno. Surpreende o fato da não significância estatística dos impactos relativos a escolas com melhores instalações. Também não era esperado o fato de a escola estar localizada na zona urbana ou rural não ter demonstrado significância.

Ao utilizar as duas disciplinas no estudo, foi possível compará-las e ver como as diversas variáveis impactam diferentemente devido a peculiaridades de cada disciplina. Ao analisar a partição de variâncias no modelo nulo, ficou claro a estrutura hierárquica dos dados, que demonstraram a importância das turmas, escolas e municípios nas notas dos alunos. Os resultados encontrados sugerem que o impacto da turma e o da escola são pequenos se comparados às variabilidades ocasionadas pelas diferenças socioeconômicas e culturais dos alunos.

Após serem incluídas as variáveis explicativas do nível do aluno, ou seja, após controlar por fatores sócio econômicos e culturais dos alunos, as variâncias de turmas, escolas e municípios diminuíram, mas continuaram significativas, evidenciando a importância desses fatores no resultado escolar do aluno e que a distribuição dos alunos nas escolas não é aleatória.

Os resultados levantam diversos questionamentos que podem ser utilizados em estudos futuros, como por exemplo: se o fato de a heterogeneidade entre as escolas ser menor no Ceará se comparado com a maioria dos outros estados, se levarmos em conta diversos estudos focados em outros estados que utilizaram modelo multinível, tem relação com a qualidade do sistema escolar. Como os Coeficientes de Correlação Intraclasse relativo aos municípios evoluíram no estado do Ceará, ou seja, se as políticas estaduais que estão sendo desenvolvidas estão diminuindo a heterogeneidade entre os sistemas escolares nos diversos municípios? Qual é o tamanho da importância do professor para o resultado da turma?

REFERÊNCIAS

- ALMEIDA, A. T. C.; Ramalho ; ARAÚJO JÚNIOR, Ignácio Tavares de. **Esforço da gestão escolar na rede pública de ensino fundamental e o desempenho dos estudantes no Brasil.** Planejamento e Políticas Publicas, v. 47, p. 135-166, 2016.
- BARBOSA, M. E.; FERNANDES, C, (2000). **Modelo multinível:** uma aplicação a dados de avaliação educacional. Estudos em Avaliação Educacional, n. 22, p. 135-153.
- BRESSOUX, P. **As pesquisas sobre o efeito-escola e o efeito-professor.** Educação em Revista, Belo Horizonte, FAE/UFMG, n. 38, p. 17-88, 2003.
- BRYK, S. A.; RAUDENBUSH, W. **Hierarchical Linear Models: Applications and Data Analysis Methods.** Newbury Park (CA): Sage, 1992.
- COLEMAN, J. S. et al. **Equality of education opportunity.** Washington: Office of Education and Welfare, 1966.
- DIAS, V. M., NAVIO, V.M. e FERRÃO, M. E. (2006). **Modelo Multinível do Desempenho Escolar de Alunos Socialmente Desfavorecidos em Escolas Públicas/Privadas - Aplicação aos Dados Portugueses do PISA 2000.** Psicologia e Educação. 5 (2), 65-79.
- GELMAN, A.; HILL, J. **Data analysis using regression and multilevel/ hierarchical models.** New York: Cambridge University Press, 2007.
- GOLDSTEIN, H. **Multilevel statistical models.** 2. ed. New York: John Wiley, 1995.
- HANUSHEK, Eric. **Teacher Quality.** In: IZUMI, L.T; EVERS W. M. (Ed.). Teacher Quality. Stanford: Hoover Institution Press, P. 1-12, 2002.
- HANUSHEK, Eric A. **The Long Run Importance of School Quality.** NBER Working Paper 9071. Cambridge, MA: National Bureau of Economic Research. 2002.

HANUSHEK, E. A.; KIMKO, D. D. **Schooling, labor-force quality and the growth of nations**. American Economic Review, v. 90, n. 5, p. 1184-1208, 2000.

Hox, J. (2002). **Multilevel analysis: techniques and applications**. London: IEA.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. Disponível em: <http://portal.inep.gov.br/censo-escolar>. Acesso em: 15/01/2020.

KREFT, I. & DE LEEUW, J. **Introducing multilevel modeling**. London: Sage Publications, 1998.

LEEUEW, J. e KREFT, I. **Random coefficient models for multilevel analysis**. Journal of Educational Statistics, 11(1), p. 57–85, 1986.

LONGFORD, N.T. **A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects**. Biometrika, 74, p. 817-827, 1987.

LONFORD, N.T. **Random Coefficient Models**. Oxford University Press, Oxford, UK, 1993.

MOREIRA, Kayline da S. G. M.; JACINTO, Paulo de A.; BEGOLIN, Izete P. **Determinantes da proficiência em matemática no Rio Grande do Sul: uma análise a partir de modelos hierárquicos**. Ensaios FEE, Porto Alegre, v. 38, p. 7-34, 2017.

MORTIMORE, P. **The Nature and Findings of Research on School Effectiveness in the Primary Sector**. School Effectiveness Research: Its Messages for School Improvements Edinburgh, 1991.

RASBASH, J. et al. **A user's guide to MLwiN**. Bristol: University of Bristol, 2009.

RIANI, J. L. R.; RIOS-NETO, E. L. G. **Background Familiar Versus Perfil Escolar do Município: Qual Possui Maior Impacto no Resultado Educacional dos Alunos Brasileiros?** Revista Brasileira de Estudos de População, v. 25, n. 2, p. 251-269, jul./dez. 2008.

RIOS-NETO, E. L. G.; CÉSAR, C. C.; RIANI, J. L. R. **Estratificação Educacional e Progressão Escolar por Série no Brasil**. Pesquisa e Planejamento Econômico, Rio de Janeiro, v.32, n.3, p.395-415, dez. 2002.

SILVA, N. V.; HASENBALG, C. **Recursos Familiares e Transições Educacionais**. Versão preliminar apresentada no Workshop de Demografia da Educação, da Associação Brasileira de Estudos Populacionais – Abep. Salvador, junho 2001.

SOARES, José Francisco. **Escola Eficaz: um estudo de caso em três escolas da rede pública do estado de Minas Gérias**. Belo Horizonte: Segrac, 2002.

SOARES, Tufi Machado. **Influência do Professor e do Ambiente em Sala de Aula sobre a Proficiência Alcançada pelos Alunos Avaliados no SIMAVE-2002**. Estudos em Avaliação Educacional, São Paulo, n. 28, p. 103-124, 2003.

SOARES, Tufi Machado. **Modelo de Três Níveis Hierárquicos para a Proficiência dos Alunos de 4ª Série Avaliados no Teste de Língua Portuguesa do SIMAVE/PROEB-2002**. *Rev. Bras. Educ.* [online]. 2005, n.29, pp.73-87.