

Modelos multiníveis aplicados ao desempenho escolar: uma proposta de análise com dados do SARESP

Camila Fernanda Bassetto Sampaio

Universidade Estadual Paulista

São Carlos, SP — Brasil

✉ camila.bassetto@unesp.br

 0000-0002-2513-7455

Driely Turi Ursini

Universidade Estadual Paulista

Franca, SP — Brasil

✉ driely.urini@unesp.br

 0009-0000-4455-496X

Marco Aurélio Kistemann Junior

Universidade Federal de Juiz de Fora

Juiz de Fora, MG — Brasil

✉ marco.kistemann@ufjf.br

 0000-0002-8970-3954



2238-0345 

10.37001/ripem.v15i3.4531 

Recebido • 26/03/2025

Aprovado • 07/06/2025

Publicado • 01/09/2025

Editor • Gilberto Januario 

Resumo: O presente estudo tem como objetivo investigar os fatores que impactam o desempenho escolar do aluno da terceira série do Ensino Médio, na rede pública estadual de ensino, matriculados nas escolas pertencentes à região Capital, definida de acordo com a Diretoria de Ensino. Foram coletados dados relativos às notas de Matemática, bem como as respostas dadas ao questionário aplicado aos pais dos alunos participantes da edição de 2013 do SARESP. O estudo emprega modelos multiníveis com dois níveis para a realização das análises empíricas. As variáveis selecionadas foram categorizadas e diferentes modelos foram propostos assumindo os efeitos específicos para as escolas como fixos e aleatórios. Estimativas para os parâmetros associados às variáveis foram obtidas via modelagem no software RStudio. Os resultados revelam que o nível de escolaridade dos pais, a renda familiar e o cumprimento das tarefas de casa pelo aluno tiveram um impacto positivo nas notas de Matemática.

Palavras-chave: Proficiência em Matemática. Background Familiar. Modelos Multiníveis. Ensino Médio.

Multilevel models applied to school performance: a proposal for analysis using SARESP data

Abstract: The present study aims to investigate the factors that impact the third-year high school students' academic performance in the state public education system, enrolled in schools within the Capital region, as defined by the Board of Education. Data were collected on Mathematics grades, as well as responses to a questionnaire administered to the parents of students participating in the 2013 edition of SARESP. The study employs two-level multilevel models for empirical analyses. The selected variables were categorized, and different models were proposed, assuming school-specific effects as fixed and random. Parameter estimates associated with the variables were obtained through modeling in RStudio software. The results reveal that parents' education level, family income, and students' completion of homework had a positive impact on Mathematics grades.

Keywords: Math Proficiency. Family Background. Multilevel Models. High School.

Modelos multinivel aplicados al desempeño escolar: una propuesta de análisis con datos del SARESP

Resumen: El presente estudio tiene como objetivo investigar los factores que impactan el desempeño escolar del alumno de tercer grado de la Educación Secundaria, en la red pública estatal de enseñanza, matriculados en las escuelas pertenecientes a la región Capital, definida de acuerdo con la Dirección de Enseñanza. Se recopilaron datos relativos a las calificaciones en Matemáticas, así como las respuestas dadas al cuestionario aplicado a los padres de los alumnos participantes en la edición 2013 del SARESP. El estudio emplea modelos multinivel con dos niveles para la realización de los análisis empíricos. Las variables seleccionadas fueron categorizadas y se propusieron diferentes modelos asumiendo los efectos específicos para las escuelas como fijos y aleatorios. Las estimaciones para los parámetros asociados a las variables fueron obtenidas mediante modelado en el software RStudio. Los resultados revelan que el nivel de escolaridad de los padres, el ingreso familiar y el cumplimiento de las tareas escolares por parte del alumno tuvieron un impacto positivo en las calificaciones de Matemáticas.

Palabras clave: Proficiencia en Matemáticas. Antecedentes Familiares. Modelos Multinivel. Educación Secundaria.

1 Introdução

Desde os anos 1990, o interesse pela investigação dos fatores que interferem no desempenho escolar tem crescido no âmbito nacional e internacional. No cenário internacional, os estudos sobre tais fatores tiveram, como marco inicial, o chamado Relatório Coleman, publicado em 1966, o qual analisou as causas para as diferenças de desempenho entre as escolas norte-americanas e impactou fortemente as políticas públicas nos Estados Unidos e na Sociologia da Educação (Coleman et al., 1966). O objetivo foi identificar, entre os sistemas escolar e familiar, qual explicaria melhor as desigualdades no desempenho escolar dos alunos. Para isso, os autores investigaram a ausência de igualdade de oportunidades educacionais para os estudantes em razão do sexo, da raça, da religião ou da região de origem, nas instituições públicas de ensino de todos os níveis do país. Os resultados levaram à conclusão de que a maior parte das desigualdades entre os alunos são provenientes das famílias e do contexto social das escolas, e que as diferenças de infraestrutura e equipamentos entre as escolas, a qualidade do seu corpo docente ou de seus currículos, sua localização e nível socioeconômico não justificavam a variação no desempenho escolar de alunos matriculados em diferentes escolas (Barbosa & Fernandes. 2001; Ferrão et al. 2001; Alves & Soares, 2008; Palermo et al., 2014).

Com os resultados e conclusões amplamente divulgadas, o Relatório Coleman foi alvo de críticas devido à falta de variáveis que caracterizassem os processos internos da escola, tais como a dinâmica e atuação do professor em sala de aula, o tempo de aula destinado às atividades acadêmicas, organização, gestão e clima escolar, entre outras. Este fato gerou a necessidade de estudos que buscassem identificar e entender os processos escolares associados ao desempenho escolar e, também, identificar escolas que conseguem levar o aluno, de origem social e cultural desfavorecida, à obtenção de resultados escolares que contrariam as expectativas, utilizando-se somente de suas práticas pedagógicas.

Estudos para a caracterização do efeito das escolas, dos fatores e das estruturas promotoras de desempenho surgiram como contraponto aos resultados pessimistas obtidos pelo Relatório Coleman e, atualmente, sabe-se que a escola pode fazer diferença no aprendizado do aluno. Brooke e Soares (2008) argumentam que

Embora parte importante da explicação dos baixos níveis de desempenho dos alunos

esteja em fatores extraescolares, há uma enorme variação entre resultados de escolas de um mesmo sistema que atendem alunos muito similares em termos socioeconômicos. Ou seja, a unidade escolar frequentada pelo aluno pode fazer diferença significativa na sua vida escolar. (Brooke & Soares, 2008, p. 9).

Para Riani e Rios-Neto (2008), “o debate sobre o impacto da família e da escola no resultado educacional já está bem adiantado nos países desenvolvidos, mas longe de se chegar a um consenso” (Riani & Rios-Neto, 2008, p. 251).

No Brasil, os primeiros resultados das pesquisas sobre o efeito e qualidade da escola foram publicados a partir da disponibilização de dados educacionais pelo Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP), do Ministério da Educação (MEC), ocorrida na década de 1990. Os estudos intensificaram-se a partir da consolidação do Sistema de Avaliação da Educação Básica (SAEB), cujos resultados despertaram atenção por evidenciarem um quadro de desigualdades educacionais no país, impulsionando a realização dos primeiros estudos brasileiros que analisaram o desempenho escolar, entre os quais estão, Fletcher (1998), Barbosa e Fernandes (2001), Ferrão et al. (2001), Soares et al. (2001) e Rodrigues e Monteiro (2024).

Na avaliação educacional, o desempenho do aluno é um importante indicador da eficácia do sistema de ensino. Esse parâmetro, no entanto, não deve ser observado isoladamente. Ferrão et al. (2001), Alves e Soares (2008), Laros e Marciano (2008), entre outros autores, realçam que análises de desempenho escolar devem cotejar informações em conjunto, ou seja, do contexto escolar e social do indivíduo. Isso porque há evidências empíricas de que o desempenho escolar dos alunos é influenciado não apenas por suas habilidades inerentes, mas também por fatores que englobam elementos sociais, econômicos e culturais de suas famílias. Aspectos mais abrangentes, portanto, são necessários para realizar avaliações educacionais – visto que eventuais desigualdades sociais comprovadamente têm influência nos resultados dos sistemas de ensino, como observam os estudos de Jesus e Laros (2004) e Laros et al. (2010).

Dessa forma, os resultados obtidos por meio dos sistemas educacionais de avaliação permitem validar com maior segurança os resultados verificados na área educacional brasileira. Isso porque houve avanços em recursos metodológicos, principalmente nos atuais critérios de avaliação do Sistema Nacional de Avaliação da Educação Básica (SAEB), do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP) e do Exame Nacional do Ensino Médio (ENEM). Estes sistemas têm, portanto, proporcionado segura análise crítica do desempenho e rendimento escolar da educação formal no Brasil.

O que se verifica, como predominante nas avaliações desses sistemas, é a presença de fatores diversificados, sejam eles de natureza humana, social ou cultural que possam influenciar o desempenho escolar dos alunos (Ferrão et al., 2001; Soares e Alves, 2003; Andrade e Laros, 2007; Bassetto, 2019). Desse modo, para bem identificar as variáveis associadas ao desempenho escolar, é necessário considerar as estruturas educacionais nas quais os indivíduos estão inseridos. Essas devem ser observadas abarcando aspectos do ambiente familiar, da disponibilidade e da qualidade da infraestrutura escolar e do corpo docente. Neste sentido, buscando contribuir com a literatura sobre a investigação dos motivos que impactam o desempenho escolar do indivíduo, o objetivo proposto na presente pesquisa consiste na utilização de modelos multiníveis com dados resultantes do SARESP, correspondentes às notas e classificações obtidas em Matemática e, também, às respostas dadas ao questionário socioeconômico por parte dos pais dos alunos.

O texto apresenta-se estruturado em seções, incluída esta introdução. A seção 2 traz uma

revisão de literatura sobre estudos que investigaram os fatores atuantes sobre o desempenho escolar do aluno utilizando-se da modelagem multinível. Na seção 3 descreve-se os dados e a seleção de variáveis para composição dos modelos. A seção 4 apresenta o modelo multinível selecionado e os resultados obtidos. Por fim, na seção 5, faz-se uma conclusão a partir da pesquisa realizada.

2 Modelos multiníveis com dados educacionais

A técnica de modelos multiníveis tem sido utilizada no Brasil desde o final dos anos 1990, para interpretar e analisar os dados das avaliações educacionais em larga escala, como mostram os estudos de Fletcher (1998), Ferrão et al. (2001), Alves e Soares (2008), Soares (2005), Andrade e Laros (2007), Laros e Marciano (2008), entre outros.

Considerado um dos pioneiros na utilização de modelos lineares hierárquicos no Brasil, a pesquisa educacional realizada por Fletcher (1998) verificou que a proporção de alunos brancos e do sexo masculino, na escola, causam efeitos positivos sobre o desempenho em matemática. Nos estudos de Soares e Mendonça (2003), Soares (2003), Soares e Alves (2003), Riani e Rios-Neto (2008) e de Brooke et al. (2014), os resultados indicaram que os alunos masculinos e negros têm desempenho inferior aos demais. Barbosa e Fernandes (2001) constataram que, enquanto a escolaridade do pai impacta positivamente a proficiência do aluno, o efeito da variável defasagem escolar é negativo. Soares et al. (2001), utilizando dados do Sistema de Avaliação da Educação Básica (SAEB) de 1997, também encontraram que os alunos com trajetória escolar regular, ou seja, sem reprovações e sem abandono de escola, juntamente com aqueles do sexo masculino e que estudam em escolas particulares, alcançaram melhores resultados e mostraram proficiência superior comparados aos demais alunos.

Enquanto Barbosa e Fernandes (2001), Ferrão et al. (2001) e Soares et al. (2001), utilizaram dados do SAEB, Natis (2001) considerou dados do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP) de 1997 para estudar a habilidade em Língua Portuguesa dos alunos da 4ª série de escolas públicas do Estado de São Paulo, considerando características dos alunos e das escolas. Assim como nos estudos relatados anteriormente, a autora utilizou a abordagem multinível com dois níveis, isto é, aluno e escola. Para o nível dos alunos, as variáveis consideradas foram a habilidade na prova de Língua Portuguesa, o sexo e a idade, considerada para identificar o desvio entre a idade efetiva do aluno e a idade adequada para os alunos da 4ª série do ano letivo. Para o nível da escola, as variáveis analisadas foram o tipo (1ª a 4ª série e 1ª a 8ª série) e a atuação do diretor (se mais ou menos atuante). Os resultados alcançados permitiram à autora concluir que, à medida que o aluno se afasta da idade adequada, o desempenho em Língua Portuguesa piora e, também, que os alunos do sexo feminino têm um rendimento melhor na avaliação. Concomitantemente, os resultados mostraram que um diretor mais atuante contribui para aumentar o desempenho médio dos alunos.

A partir de dados obtidos do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE), Soares e Mendonça (2003), Soares (2003), Soares (2005) e Soares e Andrade (2006) utilizaram a modelagem multinível para investigar os fatores associados ao desempenho escolar dos alunos matriculados nas escolas estaduais do estado de Minas Gerais.

Embora inseridos na literatura sobre a identificação de fatores determinantes para o desempenho escolar do aluno utilizando a abordagem multinível, os estudos de Soares e Andrade (2006) e de Laros e Marciano (2008) diferem-se dos demais descritos anteriormente. Soares e Andrade (2006) propuseram-se a medir o nível socioeconômico das escolas utilizando informações socioeconômicas dos alunos coletadas a partir de questionários existentes no

SIMAVE e, também, dos vestibulares da Universidade Federal de Minas Gerais (UFMG), nos anos de 2002, 2003 e 2004. Laros e Marciano (2008) consideraram dados do *National Educational Longitudinal Study* – NELS:88, provenientes de uma pesquisa conduzida pelo Ministério da Educação dos Estados Unidos, para ilustrar a análise da modelagem multinível como alternativa à regressão tradicional, preocupando-se em descrever a construção dos modelos multiníveis, passo a passo. Um resultado comum nestes estudos refere-se ao efeito positivo da condição socioeconômica sobre a proficiência do aluno.

As estimações realizadas por Brooke et al. (2014) corroboram as conclusões de Ferrão et al. (2001), Soares (2003) e Soares (2005) ao afirmar que o nível socioeconômico proporciona melhores desempenhos em Matemática e Língua Portuguesa, alunos do gênero masculino estão em desvantagem ao gênero feminino no que se refere ao aprendizado em Leitura, e que alunos pretos e alunos com defasagem escolar têm proficiência inferior em ambas as disciplinas.

Preocupados com as desigualdades raciais no sistema brasileiro de educação básica, Soares e Alves (2003) utilizaram a abordagem multinível em dois níveis – alunos e escolas – para investigar as diferenças nos desempenhos de alunos discriminados por raça. Analisando dados do SAEB de 2001, os autores observaram que a diferença entre brancos e negros cresce com o aumento, tanto no nível socioeconômico do aluno como no da escola, e é maior na rede privada, e decresce à medida em que aumenta o atraso escolar. Entre fatores escolares, observaram que escolas mais equipadas e com diretores mais envolvidos, estão associadas a um aumento da média do desempenho escolar em Matemática, mas aumentam também as diferenças raciais. Os autores concluíram o estudo ressaltando a

importância de formulação e implementação de políticas públicas e escolares, não só para a melhoria do desempenho escolar de uma forma geral, mas também para diminuir o impacto da origem socioeconômica e da raça do aluno no desempenho escolar. (Soares e Alves, 2003, p.161).

O estudo de Palermo et al. (2014) está inserido no âmbito dos estudos que buscam identificar variáveis associadas às características pessoais e sociais dos alunos que impactam o desempenho escolar. Os autores consideraram a proficiência em Matemática como variável dependente e utilizaram dados da Prova Brasil de 2007 referentes aos alunos do 5o. ano do Ensino Fundamental no Rio de Janeiro e concluíram que

os fatores que mais incidem no desempenho escolar são provenientes do background do aluno, sejam relativos a características intrínsecas ou provindos de condições familiares. Porém, é possível afirmar que a escola, mesmo que possua um poder explicativo limitado em relação ao aluno, apresenta processos característicos que influenciam o desempenho e os demais resultados escolares, direta ou indiretamente. (Palermo et al., 2014, p. 387).

Os estudos realizados no Brasil sobre os determinantes educacionais enfatizam o papel da família na educação dos filhos, mostrando que os fatores familiares, principalmente o nível de instrução dos pais, possuem grande influência sobre o desempenho educacional. Pesquisas que incluem, além dos fatores familiares, aqueles relacionados à escola e à sociedade em geral mostram que, apesar do importante impacto das características, os demais fatores citados também provocam efeito significativo na proficiência do aluno. Concomitantemente, a partir de estudos citados anteriormente, é evidente o impacto do nível socioeconômico no desempenho escolar dos alunos e, por isso, no estudo de qualquer realidade educacional deve-se necessariamente considerar a condição socioeconômica do aluno. Características

sociodemográficas tais como sexo, cor/raça e a defasagem existente entre a idade do aluno e a idade na série em que se encontra, sintetizam experiências vividas que impactam o seu aprendizado.

3 Dados e seleção de variáveis

O presente estudo teve como objetivo analisar os resultados do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), com foco na identificação dos fatores que impactam os resultados educacionais, tomando como base os dados da edição de 2013¹.

O SARESP foi instituído pelo governo do estado de São Paulo, por meio da Resolução SE Nº 27, de 29 de março de 1996. De acordo com o documento entendeu-se necessária a criação de uma política de avaliação do ensino a nível estadual, para ser possível articular-se com o Sistema Nacional de Avaliação da Educação Básica-SAEB/MEC. Entre outras funções, a Resolução ainda aborda ser imprescindível a recuperação da qualidade de ensino do Estado de São Paulo, assim como, a necessidade de possuir uma ferramenta capaz de fornecer dados capazes de subsidiar tomadas de decisões das Delegacias de Ensino e/ou Unidades Escolares.

Atualmente o SARESP é obrigatório para estudantes do 2º, 5º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio. As disciplinas contempladas em todas as provas, desde a primeira edição, são as de Língua Portuguesa e Matemática; porém, em alguns anos são trazidas outras disciplinas para completar a avaliação. Além dos dados que trazem os resultados das provas, esta avaliação também gera informações importantes que estruturam o Índice de Desenvolvimento da Educação de São Paulo (IDESP), como, por exemplo, o fluxo escolar, por meio de taxas de aprovação e reprovação dos alunos, e abandono escolar (São Paulo, 2019).

Para conduzir esta pesquisa, foram consideradas as Diretorias de Ensino (DE) às quais as escolas dos alunos pertencem. Neste estudo, o enfoque se concentrou na análise dos dados associados às DE situadas na região conhecida como Capital do estado de São Paulo, isto é, Caieiras, Carapicuíba, Centro (SP), Centro Oeste (SP), Centro Sul (SP), Diadema, Guarulhos Norte, Guarulhos Sul, Itapeverica da Serra, Itapeví, Itaquaquecetuba, Leste 1 (SP), Leste 2 (SP), Leste 3 (SP), Leste 4 (SP), Leste 5 (SP), Mauá, Mogi das Cruzes, Norte 1 (SP), Norte 2 (SP), Osasco, Santo André, São Bernardo do Campo, Sul 1 (SP), Sul 2 (SP), Sul 3 (SP), Suzano, Taboão da Serra. Do total de 253.845 alunos inscritos na edição do SARESP de 2013, matriculados nas escolas do Estado de São Paulo, foram excluídos da base de dados utilizada para esta pesquisa os alunos que não responderam ao questionário aplicado pela avaliação, assim como aqueles cujos pais também não responderam ao questionário socioeconômico.

Da mesma forma, foram excluídas as escolas em que os coordenadores e diretores não responderam às questões relevantes para este estudo. Com estas exclusões, a base de dados final foi formada por informações sobre 106.120 alunos matriculados na 3ª série do Ensino Médio da rede pública estadual de ensino. Dentre esse total, 43.564 alunos frequentam escolas cujas DEs estão vinculadas à região da Capital, representando aproximadamente 41% do total de alunos, para os quais a Tabela 1 apresenta informações descritivas sobre as proficiências alcançadas na disciplina de Matemática.

Tabela 1: Desempenho em Matemática na região Capital no SARESP de 2013.

Nível	Nota Mínima	Mediana	Média	Nota Máxima	Desvio Padrão	Percentual de Alunos
-------	-------------	---------	-------	-------------	---------------	----------------------

¹ O uso dos dados de 2013 deve-se ao fato de ter sido o último ano em que os questionários aplicados pelo SARESP foram respondidos por agentes escolares. A partir deste ano, somente alunos e pais de alunos responderam.

AB	147,1	240,5	235,3	274,9	28,3	56,54%
BA	275,0	304,0	306,5	349,9	20,5	38,47%
AD	350,0	364,4	367,1	399,9	12,8	4,71%
AV	400,0	408,4	409,4	427,0	6,9	0,28%

Fonte: Elaboração dos autores.

É possível observar que, na categoria Abaixo do Básico (AB), a menor nota foi de 147,1 pontos e, a maior foi de 274,9 pontos. Há, neste nível de proficiência, 56,5% dos alunos considerados na presente pesquisa, ou seja, mais da metade dos estudantes encontram-se na menor categoria de classificação. O desvio padrão está em 28,3 pontos, indicando que, quando comparados com a média do conjunto – 235,3 pontos – os estudantes podem ter feito 28,3 pontos a mais ou a menos da média. Esses dados podem apontar um elevado índice de desigualdade na pontuação do mesmo conjunto de estudantes. A amplitude (diferença entre a maior e a menor nota) desse grupo confirma o desequilíbrio, sendo de 127,8 pontos.

Na categoria Básico (BA) estão 38,47% dos estudantes da amostra pesquisada. Entre os alunos classificados neste espaço, os estudantes com notas mínima e máxima, respectivamente, atingiram 275,0 e 349,9 pontos. A média deste conjunto é de 306,5 pontos. O desvio padrão ficou em 20,5 pontos, esse dado indica que os estudantes puderam ter notas com mais ou menos de 20,5 pontos em relação à nota média do grupo. A amplitude do conjunto ficou em 74,9 pontos, indicando desigualdade considerável, porém, já menor do que o grupo AB.

No grupo classificado como adequado (AD), se encontram os estudantes que atingiram notas de acordo com o que se espera no ano em que está matriculado. Nesta categoria, o grupo analisado apresentou nota mínima e máxima, respectivamente, em 350,0 e 399,9. A quantidade de estudantes da amostra que está neste conjunto é de 4,71%. O desvio padrão de 12,8 pontos nesta categoria é menor do que nas anteriores, sendo considerado positivo, pois aponta variação das notas entre 12,8 pontos a mais ou a menos da média, de 367,1. A diferença entre a maior e a menor nota (amplitude) ficou em 49,9 pontos.

A última categoria trata dos alunos que se encontram no nível avançado (AV), os quais apresentam domínio acima dos conteúdos, habilidades e competências do ano/série em que estão inseridos. Representam 0,28% dos alunos da amostra analisada. Dentro dela, as notas mínima e máxima são, respectivamente, 400,0 e 427,0 pontos. A média é de 408,4 pontos. E o desvio padrão está em seu menor número dentre todas as categorias, pois é de 6,9 pontos, o que representa baixa variação entre as notas diante da média do grupo. Ainda na Tabela 1, observa-se que a maior parte dos estudantes está classificada na categoria AB, sugerindo que aproximadamente 57% e, portanto, mais da metade dos alunos, tem domínio insuficiente dos conteúdos escolares correspondentes ao ano/série em que se encontram.

Considerando a abrangência dos questionários aplicados pelo SARESP, que englobam significativo número de variáveis nos questionários destinados aos alunos, pais diretores e coordenadores, tornou-se necessário selecionar variáveis mais relevantes. Para identificar quais variáveis seriam relevantes, utilizou-se estudos com objetivos semelhantes aos desta pesquisa. Foram, então, identificadas as variáveis que demonstraram maior relevância relativas ao desempenho em Matemática.

O Quadro 1 identifica as variáveis e os autores associados.

Quadro 1: Autores e variáveis consideradas nos modelos multiníveis

Autores	Variáveis
Fletcher (1998)	Características do ambiente escolar e familiar, tais como nível socioeconômico, etnia e gênero.

Barbosa e Fernandes (2000)	Escolaridade do pai, gênero, tipo de rede.
Barbosa e Fernandes (2001)	Defasagem escolar, escolaridade do pai, motivação (gosta de Matemática), relação casa- escola, didática e escolaridade do professor, infraestrutura e equipamentos escolares (conservação do prédio, laboratórios, mobiliário e equipamento, rede de ensino).
Ferrão et al. (2001)	Condição socioeconômica, atraso escolar, etnia, infraestrutura, segurança e limpeza da escola, turno de funcionamento da turma.
Soares, Cesar e Mambrini (2001)	Condição socioeconômica, trajetória escolar (reprovação e abandono de escola), gênero, escolar pública e privada.
Natis (2001)	Gênero, defasagem escolar, série (1 ^a a 4 ^a e 1 ^a a 8 ^a), atuação do diretor.
Soares e Mendonça (2003)	Nível socioeconômico, escolaridade da mãe, defasagem escolar, abandono de escola, etnia, frequência a cultos religiosos.
Soares (2003, 2005)	Condição socioeconômica, gênero, anos de defasagem escolar, número aproximado de faltas do professor durante o ano, se o professor passa ou não de casa, turno de funcionamento da turma.
Jesus e Laros (2004)	Condição socioeconômica, escolaridade do pai, etnia, atraso escolar, dever de casa, recursos pedagógicos.
Andrade e Laros (2007)	Condição socioeconômica, etnia, gênero, dever de casa, defasagem escolar.
Machado et al. (2008)	Reprovação, escolaridade da mãe, dever de casa, possuir mais de 20 livros, custo do aluno na escola, tamanho de turmas, capacitação de professores, presença de laboratórios, infraestrutura da escola.
Riani e Rios-Neto (2008)	Defasagem escolar, gênero, etnia, zona rural e urbana, escolaridade da mãe, ocupação do chefe.

Fonte: Elaboração dos autores.

Com base nos estudos indicados no Quadro 1, as questões – que constam no Quadro 2 e são apresentadas na execução do SARESP – foram selecionadas. Os modelos multiníveis de dois níveis obtidos dessa forma contêm variáveis associadas aos alunos no nível 1 e às escolas no nível 2.

Quadro 2: Questões selecionadas a partir dos questionários do SARESP.

Questionário	Questões	Alternativas	
Pais	Q1: Qual é a cor ou raça do pai?	(a) Branco. (b) Negro. (c) Pardo/mulato.	(d) Amarela/oriental. (e) Indígena. (f) Não sabe.
	Q2: Qual é a cor ou raça da mãe?	Mesmas da Q1.	
	Q3: Marque a situação de trabalho do pai.	(a) Empregado. (b) Autônomo. (c) Dono de negócio próprio.	(d) Temporário. (e) Aposentado. (f) Desempregado. (g) Outra situação.
	Q4: Qual é o grau de escolaridade do pai?	(a) Não estudou. (b) EF I incompleto. (c) EF I completo. (d) EF II incompleto. (e) EF II completo.	(f) EM incompleto. (g) EM completo. (h) ES incompleto. (i) ES completo. (j) Não sabe.
	Q5: Marque a situação de trabalho da mãe.	Mesmas da Q3.	
	Q6: Qual é o grau de	Mesmas da Q4.	

	escolaridade da mãe?		
	Q7: Qual é a renda familiar de seu domicílio, em R\$?	(a) até 850. (b) de 851 a 1.275. (c) de 1.276 a 2.125. (d) de 2.126 a 4.250.	(e) acima de 4.250. (f) não sabe/não quer responder.
Aluno	Q8: Há quantos anos você está cursando o Ensino Médio?	(a) Menos de 8 anos. (b) 8 anos. (c) 9 anos. (d) 10 anos.	(e) 11 anos. (f) Mais de 11 anos. (g) Não sei.
	Q9: Após as avaliações, frequentou aulas de recuperação para melhorar os resultados?	(a) Sim, sempre. (b) Sim, às vezes. (c) Não, pois não precisei.	(d) Não, precisei, mas minha escola não ofereceu.
	Q10: Quanto tempo passa por dia fazendo as lições de casa?	(a) Menos de 20 min. (b) Entre 20 e 40 min. (c) Entre 40 min. e 1h.	(d) Mais de 1 hora. (e) Não faço lição.
	Q11: Trabalhou fora durante o Ensino Médio?	(a) Sim, todo o tempo. (b) Sim, menos de 1 ano. (c) Sim, de 1 a 2 anos.	(d) Sim, 2 a 3 anos. (e) Não trabalhou.
	Q12: Você se considera:	Mesmas da Q1.	
Coordenador	Q13: Há quantos anos exerce esse cargo/função?	(a) Menos de 1 ano. (b) De 1 a 2 anos. (c) De 3 a 5 anos. (d) De 6 a 8 anos.	(e) De 9 a 10 anos. (f) De 11 a 15 anos. (g) De 16 a 20 anos. (h) Mais de 20 anos
	Q14: Além dessa função, você exerce outra atividade?	(a) Sim, atividade docente na Rede Municipal de Ensino da capital. (b) Sim, atividade docente na Rede Municipal de Ensino.	(c) Sim, em outra Rede Municipal do Estado de São Paulo. (d) Sim, outra atividade não vinculada ao ensino. (e) Não.
	Q15: Há algum projeto de redução das taxas de abandono na sua escola?	(a) Sim, e o projeto está sendo aplicado. (b) Sim, mas ainda não foi implementado. (c) Não criamos ainda o projeto, embora exista o problema.	(d) Não, porque na minha escola não há esse tipo de problema.
	Q16: Há algum projeto de redução das taxas de reprovação na sua escola?	Mesmas da Q15.	
	Q17: Neste ano faltou professor?	(a) Sim, muitas vezes. (b) Sim, poucas vezes.	(c) Não ocorreu. (d) Não sei.
Diretor	Q18: Há quantos anos exerce esse cargo/função?	Mesmas da Q13.	
	Q19: Além dessa função, exerce outra?	Mesmas da Q14.	
	Q20: Qual a sua atividade predominante?	(a) Burocrática. (b) Administrativa de gestão escolar (faltas de professor, disciplinas).	(c) Pedagógica (d) Atendimento à comunidade escolar.

Fonte: Elaborado pelos autores com base nos dados do SARESP (2023).

As questões associadas aos alunos, pais dos alunos, coordenadores e diretores são todas

qualitativas e, portanto, fez-se necessário categorizá-las para a realização da modelagem. O Quadro 3 contém a categorização utilizada e a codificação (nomenclatura atribuída nos modelos propostos) de cada uma das variáveis selecionadas. Após a categorização das variáveis selecionadas e, com o objetivo de avaliar sua influência significativa no desempenho em Matemática, foram realizadas análises de regressão linear múltipla para cada uma das regiões e para todos os anos considerados. Simultaneamente, adotou-se o método *Stepwise* como critério para a seleção do modelo que melhor se adequa à explicação dos dados. Neste estágio, não se pretende realizar uma interpretação exaustiva das estimativas dos parâmetros associados a cada variável, uma vez que o foco principal é avaliar a relevância de cada variável no modelo.

Quadro 3: Categorização e codificação das variáveis.

Questionário	Questão	Categorização*	Codificação
Pais	Q1	0 – outra (parda/mulata, amarela, indígena); 1 – branca; 2 – negra.	corp
	Q2	Mesma da Q1.	corm
	Q3	0 – não trabalha (aposentado, desempregado, outra situação); 1 – trabalha (empregado, autônomo, outro).	trabp
	Q4	0 – não estudou; 1 – EF incompleto ou completo; 2 – EM incompleto ou completo; 3 – ES incompleto ou completo.	escp
	Q5	Mesma da Q3.	trabm
	Q6	Mesma da Q4.	escm
	Q7	0 – até R\$ 2.126,00; 1 – mais de R\$ 2.126,00.	rendaf
Aluno	Q8	0 – sem defasagem (3 anos ou menos); 1 – com defasagem (4 anos ou mais).	def_EM
	Q9	0 – sim (sempre, às vezes); 1 – não (não precisei, escola não ofereceu).	rec_mat
	Q10	0 – não faz lição de casa; 1 – faz lição de casa (menos de 20 min., entre 20 e 40 min., entre 40 min. e 1h, mais de 1h).	lic_casa
	Q11	0 – não trabalhou; 1 – trabalhou (o tempo todo, menos de 1 ano, de 1 a 2 anos, de 2 a 3 anos).	trab
	Q12	Mesma da Q1.	cora
Coordenador	Q13	0 – 10 anos ou menos; 1 – mais de 10 anos.	tcargoc
	Q14	1 – sim; 0 – não.	outra_ativc
	Q15	1 – sim; 0 – não.	tx_aband
	Q16	Mesma da Q15.	tx_reprov
	Q17	1 – sim; 0 – não.	faltaprof
Diretor	Q18	Mesma da Q13.	t_cargod
	Q19	1 – sim; 0 – não.	outrativd
	Q20	0 – burocráticas ou administrativas relacionadas à gestão escolar; 1 – pedagógicas ou atendimento à comunidade	princ_ativ

	escolar.	
--	----------	--

Fonte: Elaboração dos autores.

Após a categorização das variáveis selecionadas e, com o objetivo de avaliar sua influência significativa no desempenho em Matemática, foram realizadas análises de regressão linear múltipla para cada uma das regiões e para todos os anos considerados. Simultaneamente, adotou-se o método *Stepwise* como critério para a seleção do modelo que melhor se adequa à explicação dos dados. Neste estágio da pesquisa, não se pretende realizar interpretação exaustiva das estimativas dos parâmetros associados a cada variável, uma vez que o foco principal é avaliar a relevância de cada variável no modelo (Soares e Mendonça, 2003; Andrade e Laros, 2007). No entanto, dada a complexidade desta técnica de análise fatorial e considerando que não se alinha com o objetivo desta pesquisa, optou-se por utilizar a regressão linear múltipla, tendo a proficiência em Matemática como variável dependente. A modelagem e todas as estimativas apresentadas, neste estudo, foram conduzidas utilizando o *software* RStudio².

Considerando, assim, todas as variáveis relacionadas às questões descritas no Quadro 3, um modelo completo, englobando todas as variáveis, foi inicialmente considerado. O modelo completo é expresso da seguinte forma, onde i denota o aluno e j a escola, sendo utilizado para identificar as variáveis que apresentam significância em todas as regiões.

$$\begin{aligned}
 profic_{ij} = & \alpha + \beta_{1i}corp_i + \beta_{2i}corm_i + \beta_{3i}trabp_i + \beta_{4i}escp_i + \beta_{5i}trabm_i + \beta_{6i}escm_i + \\
 & \beta_{7i}rendaf_i + \beta_{8i}def_EM_i + \beta_{9i}rec_mat_i + \beta_{10i}lic_casa_i + \beta_{11i}trab_i + \\
 & \beta_{12i}cora_i + \beta_{13i}t_cargoc_j + \beta_{14i}outra_atvc_j + \beta_{15i}tx_aband_j + \\
 & \beta_{16i}tx_reprov_j + \beta_{17i}falta_prof_j + \beta_{18i}t_cargod_j + \beta_{19i}outra_atvd_j + \\
 & \beta_{20i}princ_ativ_j + e_{ij}
 \end{aligned} \quad (1)$$

Na Tabela 2 são apresentadas as estimativas para os parâmetros do modelo selecionado pelo método *Stepwise*, a partir daquele dado na expressão (1)³. De acordo com as estimativas mostradas na Tabela 2, as variáveis significativas para explicar a proficiência em Matemática dos alunos da rede pública de ensino matriculados na 3ª série do Ensino Médio, na Capital, são: cor do pai (corp0 e corp2), cor da mãe (corm0 e corm2), se a mãe trabalha ou não (trabm0), escolaridade do pai (escp1, escp2 e escp3), escolaridade da mãe (escm1, escm2 e escm3), renda familiar (rendaf1), se o aluno faz, ou não, o dever de Matemática (lic_casa1), o tempo que o coordenador está no cargo (t_cargoc1) e se exerce, ou não, outra atividade (outra_atvc0), se a escola possui algum projeto voltado à redução da taxa de abandono (tx_aband1) e de reprovação (tx_reprov1), se falta, ou não, professor para alguma disciplina na escola (falta_prof0) e o tempo que o diretor está no cargo (t_cargod1).

Tabela 2: Variáveis significativas.

Variáveis	Estimativas	Erro Padrão	Variáveis	Estimativas	Erro Padrão
Intercepto	254,06***	2,0682	escm2	12,17***	1,4306
corp0	-3,95***	0,4951	escm3	15,89***	1,5584
corp2	-8,02***	0,6595	rendaf1	8,00***	0,5154
corm0	-3,60***	0,4837	lic_casa1	6,78***	0,7076

² RStudio é uma plataforma de desenvolvimento integrado de código aberto para a linguagem de programação R, amplamente utilizada em análises estatísticas e geração de gráficos. Mais informações podem ser encontradas no link <https://posit.co/products/open-source/rstudio/>.

³ Nas tabelas que contém as estimativas, os símbolos *, ** e *** representam os níveis de significância 10%, 5% e 1%, respectivamente.

corm2	-8,24***	0,7865	t_cargoc1	1,96*	0,7974
trabm0	0,95*	0,4727	outra_ativc0	-1,63***	0,4726
escp1	4,51***	1,2407	tx_aband1	-2,80***	0,7354
escp2	8,69***	1,2817	tx_reprov1	-2,57*	1,0097
escp3	11,99***	1,4505	falta_prof0	-1,59*	0,7392
escm1	5,79***	1,4015	t_cargod1	2,81***	0,4379

Fonte: Elaboração dos autores.

Com exceção das variáveis *trabm*, *t_cargoc*, *tx_reprov* e *falta_prof*, as quais foram significativas a 5% para explicar o desempenho em Matemática do aluno, todas as demais apresentaram nível de significância de 1%. Todas as variáveis mostradas na Tabela 2 são consideradas na modelagem multinível proposta com dados da Capital.

4 Modelos multiníveis: resultados e discussões

Na pesquisa educacional, a população de interesse compreende tanto as escolas quanto os alunos que frequentam essas instituições. A amostragem ocorre em dois estágios distintos: inicialmente, é selecionada uma amostra de escolas, seguida pela seleção de uma amostra de alunos dentro de cada escola. No contexto deste estudo, as escolas foram agrupadas com base nas regiões geográficas do estado de São Paulo, conforme definido pelas Diretorias de Ensino às quais estão vinculadas estas escolas.

A estrutura hierárquica dos dados é composta por dois níveis, nos quais os alunos (considerados a unidade micro ou nível 1) estão agrupados em escolas (consideradas a unidade macro ou nível 2). O principal objetivo desta pesquisa é examinar o impacto das variáveis relacionadas à escola na proficiência em Matemática dos alunos. Seguindo os passos de modelagem multinível propostos por Hox (2002), inicialmente, foi estimado o modelo nulo, assim denominado por não ter variáveis explicativas, sendo formado apenas pelo intercepto e por um termo de erro. Tal modelo exerce papel fundamental no estudo da distribuição da variância total da variável resposta pelos níveis da hierarquia. É por meio do modelo nulo que o efeito da escola sobre a proficiência em Matemática do aluno é estimado. Em outras palavras, o modelo nulo serve como base para a estimação da variância explicada *versus* variância não explicada em comparação aos modelos condicionais estimados *a posteriori*.

Supondo que o subscrito *i* corresponda ao aluno e o subscrito *j* represente a escola e, considerando a existência de *J* escolas, $j = 1, \dots, J$, cada uma delas com n_j alunos, $i = 1, \dots, n_j$, o modelo nulo pode ser especificado como na expressão (2).

$$\begin{aligned} y_{ij} &= \beta_{0j} + e_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \end{aligned} \quad (2)$$

em que y_{ij} é a variável resposta (proficiência em Matemática) do *i*-ésimo aluno na *j*-ésima escola. Assume-se que $e_{ij} \sim N(0, \sigma^2)$ e $u_{0j} \sim N(0, \tau_{00})$, e independência entre o efeito fixo e os efeitos aleatórios, bem como entre os efeitos aleatórios.

De acordo com o modelo nulo, a proficiência do aluno *i* que estuda na escola *j* é definida pelo intercepto β_{0j} e pelo termo do erro e_{ij} . O intercepto β_{0j} , que representa um efeito específico da escola *j* sobre a proficiência em Matemática do aluno, é dado pela soma de γ_{00} e u_{0j} , os quais indicam, respectivamente, a média global da proficiência em Matemática dos alunos da 3ª série do Ensino Médio – portanto, um efeito fixo – e o afastamento da proficiência

média em Matemática da escola j à média global γ_{00} .

Ainda no modelo (2), a primeira linha caracteriza o nível 1, a qual tem associado um termo aleatório representado por e_{ij} cuja variância indica a variabilidade dentro das escolas, a segunda linha é o nível 2, para a qual u_{0j} representa o efeito aleatório e, cuja variância indica a variabilidade entre as escolas. Então, a variância da proficiência em Matemática está decomposta na variância entre as escolas (τ_{00}) e dentro (ou intra) escolas (σ^2) e, por isso, o modelo com apenas o intercepto aleatório é chamado modelo de componentes da variância. Ao inserir uma variável no nível 1 do modelo, o coeficiente associado pode ser tratado como fixo ou aleatório. No caso aleatório, um termo \mathbf{u}_j é inserido no modelo e então tem-se que

$$\mathbf{u}_j = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim NMV(\mathbf{0}, \mathbf{\Omega}_u), \text{ com média } \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ e matriz de variância-covariância dada por}$$

$$\mathbf{\Omega}_u = \begin{pmatrix} \sigma_{u_0}^2 & \\ \sigma_{u_{01}}^2 & \sigma_{u_{01}}^2 \end{pmatrix}.$$

Com as estimações das variâncias dentro das escolas (σ^2) e entre as escolas (τ_{00}), pode-se calcular o coeficiente de correlação intra-escolar (ICC), o qual mede a proporção da variância entre as escolas face à variância total da variável resposta, possibilitando medir a magnitude do efeito-escola, isto é, a heterogeneidade dos resultados escolares entre as escolas. O ICC varia de zero a um e é calculado por meio da razão (3) (Kreft e De Leeuw, 1998; Raudenbush e Bryk, 2002; Goldstein, 1995; Gelman, 2006).

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}, \quad 0 \leq \rho \leq 1. \quad (3)$$

Quando $\rho = 0$, as escolas são homogêneas entre si e, conseqüentemente, o desempenho do aluno independe da escola em que está matriculado. Isso significa que toda a variabilidade na proficiência em Matemática deve-se apenas às diferenças entre os alunos, mas não entre as escolas. Por outro lado, quando $\rho = 1$, a variabilidade na proficiência em Matemática depende somente da escola em que o aluno está matriculado, ou seja, as características individuais dos alunos em nada afetam o seu desempenho escolar.

Bliese (2016) argumenta que, além de calcular o ICC, é importante determinar a significância da variância τ_{00} e, para isso, sugere comparar o resultado da Deviance, calculada por $-2 \log \text{verossimilhança}$, para dois modelos, um contendo o intercepto aleatório e o outro não (Raudenbush e Bryk, 2002). Deve-se comparar a magnitude das diferenças na Deviance entre os modelos. Uma diferença significativa entre esses resultados sugere a existência de uma variação relevante no intercepto em termos de proficiência em Matemática entre as escolas e, portanto, há evidência para considerar o modelo que contém o intercepto aleatório. Tal critério sugere como mais adequado o modelo com o menor valor para a Deviance (Kreft e Leeuw, 1998). Após o cálculo do ICC e da verificação da significância de tal valor, variáveis devem ser introduzidas no modelo.

Para a região Capital do estado de São Paulo, o valor do ICC foi

$$\rho = \frac{154,28}{154,28 + 1.987,74} = 0,072, \text{ sugerindo que o percentual de variabilidade contida na}$$

proficiência em Matemática do aluno da 3ª série do Ensino Médio matriculado na rede pública de ensino e pertencente à região Capital do Estado corresponde a 7,2%.

A Tabela 2 mostrou que as variáveis significativas associadas ao nível 1 (aluno) para explicar o desempenho em Matemática alcançado no SARESP de 2013 foram a cor e a escolaridade da mãe e do pai, a situação de trabalho da mãe, a renda familiar e se faz o dever de casa. Considerando tais variáveis, diferentes modelos hierárquicos foram propostos e, por meio do critério AIC (Bozdogan, 1987), selecionou-se o mais adequado para explicar o desempenho do aluno pertencente à região Capital do Estado, no SARESP de 2013.

O modelo selecionado foi aquele que considerou, além do intercepto, o coeficiente de inclinação associado à etnia do pai, como aleatório, sugerindo que, diante de uma variância significativa para tal variável, o fato de o pai ser branco, negro ou de outra etnia poderia impactar o desempenho do filho de maneiras distintas entre as escolas. No nível 2, do modelo hierárquico, foram inseridas as variáveis que indicam o tempo que o coordenador e o diretor atuam no cargo, se o coordenador desempenha outra atividade remunerada, se a escola possui algum projeto voltado à redução na taxa de abandono e de reprovação, e se falta professor para alguma disciplina.

Entre os modelos de efeito fixo e efeito aleatório propostos, aquele correspondente ao menor valor do AIC foi selecionado e é mostrado na expressão (4).

nível 1:

$$y_{ij} = \beta_{0j} + \beta_{1j}corp_i + \beta_{2j}corm_i + \beta_{3j}trabm_i + \beta_{4j}escp_i + \beta_{5j}escm_i + \beta_{6j}rendaf_i + \beta_{7j}lic_casa_i + e_{ij}$$

nível 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01j}t_cargoc_j + \gamma_{02j}outra_ativc_j + \gamma_{03j}tx_aband_j + \gamma_{04j}tx_reprov_j + \gamma_{05j}falta_prof_j + \gamma_{06j}t_cargod_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{kj} = \gamma_{k0}, k = 2, \dots, 7$$

$$\gamma_{0lj} = \zeta_{l0}, l = 1, \dots, 6.$$

Para o modelo selecionado, as estimativas são apresentadas na Tabela 3.

Tabela 3: Estimativas para o modelo selecionado

Efeito fixo					
Parâmetros	Estimativas	P-valor	Parâmetros	Estimativas	P-valor
Intercepto	256,038 2,582	0,0000	escm2	10,649 1,408	0,0000
corp0	-3,023 0,496	0,0000	escm3	13,36 1,536	0,0000
corp2	-6,836 0,663	0,0000	rendaf1	6,748 0,509	0,0000
corm0	-3,013 0,476	0,0000	lic casa1	5,369 0,700	0,0000
corm2	-7,358 0,775	0,0000	t cargoc1	2,151 1,420	0,1300
trabm0	0,881 0,465	0,0583	outra_ativc0	1,880 0,830	0,0238
escp1	3,898 1,220	0,0014	tx_aband1	-2,207 1,279	0,0847
escp2	7,496 1,263	0,0000	tx_reprov1	-2,961 1,714	0,0844
escp3	9,731 1,429	0,0000	falta_prof1	-1,143 1,284	0,3733
escm1	5,110 1,379	0,0002	t cargod1	2,097 0,762	0,0061
Efeitos aleatórios					
Nível 1	Variância	Desvio Padrão			

Intercepto	133,22	11,54	
corp0	7,21	4,24	
corp2	17,96	2,69	
Residual	1.939,60	44,04	

Fonte: Elaboração dos autores.

Considerando tais estimativas, é válido afirmar empiricamente também que, alunos filhos de pais e mães que se declararam brancos, com níveis de escolaridade mais altos, cuja renda familiar é superior a R\$ 2.126 e que fazem o dever de casa, tiveram melhores notas no SARESP de 2013, comparados aos alunos que não pertencem às categorias citadas.

A Tabela 3 mostra ainda que, conforme estimativa do intercepto, a média dos alunos de todas as escolas e que não pertencem ao grupo daqueles com os melhores resultados, foi próxima de 256 pontos. As variáveis associadas ao nível 2 significativas para explicar o desempenho do aluno na edição do SARESP de 2013 permitem afirmar, empiricamente, que escolas que possuem projetos voltados à redução da taxa de abandono tiveram efeito negativo sobre a nota alcançada pelo aluno, pois a estimativa foi negativa. O mesmo pode-se concluir sobre as escolas com projetos destinados à redução da taxa de reprovação, variável para a qual a estimativa também foi negativa. Por outro lado, escolas cujo diretor atua no cargo há mais de 10 anos contribuíram positivamente para o desempenho em Matemática do aluno, como mostra a estimativa associada a tal variável.

Tratada como aleatória, uma variância significativa para a variável associada à etnia do pai do aluno foi estimada, evidenciando diferenças nos resultados dos alunos conforme a escola, em relação à cor do pai do indivíduo participante da avaliação. A variância significativa associada à variável que indica a cor do pai impactou, de maneiras distintas, o desempenho em Matemática no SARESP de 2013, dos alunos em diferentes escolas, na região Capital do Estado.

De acordo com o modelo selecionado, tem-se que o efeito da variável que indica a etnia do pai, sobre o desempenho do aluno, é dado por $-3,02 + \hat{u}_{1j}$ para os pais que se declararam pardo/mulato, amarelo/oriental ou indígena, e $-6,84 + \hat{u}_{1j}$ para aqueles que se declararam negros. Para cada etnia declarada pelo pai, pode-se calcular um intervalo de cobertura para as estimativas do coeficiente de inclinação associado a tal variável. Considerando as variâncias estimadas para corp0 e ccorp2, os intervalos correspondentes aos pais pardos/mulatos, amarelos/orientais ou indígenas, e negros, são dados, respectivamente, por $-3,02 \pm 1,96\sqrt{7,21}$ e $-6,84 \pm 1,96\sqrt{17,96}$, resultando nos intervalos $(-8,28; -2,24)$ e $(-15,14; 1,47)$.

Logo, pode-se afirmar, empiricamente que, para a média das escolas, espera-se, na nota do aluno, uma redução de 3 pontos se o pai for pardo/mulato, amarelo/oriental ou indígena, e de quase 7 pontos se o pai for negro. Porém, com os intervalos de cobertura para as inclinações associadas à etnia do pai, estima-se que, em 95% das escolas, o decréscimo na nota do aluno filho de pai pardo/mulato, amarelo/oriental ou indígena esteja, aproximadamente, entre 8,2 e 2,2 pontos, e para aqueles filhos de pais negros pode haver reduções de no máximo 15 pontos ou até mesmo acréscimos, sendo o máximo próximo de 1,5 ponto, ocasionando diferentes médias em Matemática entre as escolas. Todas as comparações foram baseadas na nota dos alunos com pais declarados brancos, para os quais a média em Matemática no SARESP foi próxima de 256 pontos.

5 Considerações finais

O presente estudo buscou analisar os fatores intra e extraescolares que podem

influenciar o desempenho educacional do aluno e, para tanto, utilizou dados coletados no SARESP de 2013. A modelagem hierárquica proposta permite identificar a variabilidade do desempenho em Matemática do aluno na rede pública estadual de ensino advinda de fatores escolares e do contexto familiar. Considerando as observações agrupadas por regiões do Estado, definidas a partir da Diretoria de Ensino a qual pertence a escola, estimativas foram obtidas utilizando o software RStudio. Para a região denominada Capital do Estado, pôde-se observar que a etnia, a escolaridade do pai e da mãe e a renda familiar foram significativas para explicar o desempenho do aluno na edição do SARESP de 2013, sugerindo que alunos filhos de pai e mãe brancos, com mais anos de estudo e com renda familiar acima de R\$ 2.126 alcançaram médias, em Matemática, superiores aos demais alunos.

Segundo Bassetto (2021),

Uma justificativa para o desempenho mais baixo dos alunos filhos de pais e mães que se declararam negros é pautada, segundo o INEP (2003), na entrada mais tardia dos negros na economia, devido aos pais com menores níveis de escolaridade. Outra explicação para a diferença no desempenho ao comparar alunos brancos e negros, pardos ou outras etnias, de acordo com o INEP (2003), está associada ao racismo, o qual se faz ainda presente na sociedade brasileira. Atitudes e comportamentos racistas afetam a autoestima dos alunos e causa prejuízos no desempenho educacional, o que traz, como consequência, um aumento da desigualdade no sistema de ensino. (Bassetto, 2021, p. 11-12).

O efeito de mães que não trabalham (aposentada, desempregada ou em outra situação) sobre o desempenho do filho foi positivo. Associadas às escolas, aquelas que possuem projetos voltados à redução da taxa de abandono teve impacto negativo sobre o desempenho do aluno e aquelas em que o diretor atua no cargo há mais de 10 anos contribuíram positivamente para o resultado do aluno pertencente às escolas da região Capital. As escolas em que ocorre o problema associado à falta de professor para alguma disciplina tiveram impacto negativo sobre o desempenho dos alunos. Estudos como este buscam obter conhecimento para subsidiar soluções para problemas que são constantes preocupações de todos aqueles que estão envolvidos com o sistema educacional brasileiro. Como apontam Soares e Alves (2003), a desigualdade observada, no contexto escolar, entre brancos e negros, pardos ou alunos de outras etnias, somente se reduzirá com a implementação de políticas públicas educacionais que busquem, concomitantemente, um equilíbrio entre a eficácia e a equidade na educação e a redução do impacto das questões socioeconômicas do aluno sobre o seu desempenho escolar.

Os resultados obtidos nesta pesquisa corroboraram muitas conclusões de estudos realizados na literatura, citados previamente. Assim, a partir dos resultados alcançados, espera-se ter contribuído com a literatura que aborda a identificação de fatores que impactam o desempenho educacional do aluno e ressaltar a importância de formulação e implementação de políticas públicas voltadas à educação que visem, simultaneamente, melhorar a qualidade do ensino ofertado à sociedade e diminuir o impacto das características socioeconômicas adversas sobre o desempenho do aluno.

Referências

- Alves, M. T. G. & Soares, J. F. (2008). O efeito das escolas no aprendizado dos alunos: Um estudo com dados longitudinais no Ensino Fundamental. *Educação e Pesquisa*, 34(3), 527-544. <https://doi.org/10.1590/S1517-97022008000300007>
- Andrade, J. M. & Laros, J. A. (2007). Fatores associados ao desempenho escolar: Estudo

- multinível com dados do SAEB/2001. *Psicologia: Teoria e Pesquisa*, 23(1), 33-42. <https://doi.org/10.1590/S0102-37722007000100005>
- 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. In F. Crespo (Org.), *Avaliação, Ciclos e Promoção na Educação* (pp. 155-178). Artmed Editora.
- Bassetto, C. F. (2019). Background familiar e desempenho escolar: Uma abordagem com variáveis binárias a partir dos resultados do Saresp. *Revista Brasileira de Estudos de População*, 36, e0077. <https://doi.org/10.20947/S0102-3098a0077>
- Bassetto, C. F. (2021). Análise regional dos resultados do Saresp: Uma abordagem com modelos hierárquicos. *Ciência & Educação (Bauru)*, 27, e21063.
- Bliese, P. (2016). Multilevel modeling in R (2.6): A brief introduction to R, the multilevel package and the nlme package. In R Development Core Team, *An introduction to R*. <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370. <https://doi.org/10.1007/BF02294361>
- Brooke, N., Fernandes, N. da S., Miranda, I. P. H. de & Soares, T. M. (2014). Modelagem do crescimento da aprendizagem nos anos iniciais com dados longitudinais da pesquisa GERES. *Educação e Pesquisa*, 40(1), 77-94. <https://doi.org/10.1590/S1517-97022014000100006>
- Brooke, N. & Soares, J. F. (2008). *Pesquisa em eficácia escolar: Origem e trajetórias*. Editora UFMG.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. 1. U.S. Department of Health, Education, and Welfare, Office of Education.
- Ferrão, M. E., Beltrão, K. I., Fernandes, C., Santos, D., Suárez, M. & Andrade, A. do C. (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. <https://doi.org/10.20947/S0102-3098a0008>
- Fletcher, P. (1998). *À procura do ensino eficaz*. Ministério da Educação e Cultura, Departamento da Avaliação da Educação Básica.
- Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Goldstein, H. (1995). *Multilevel statistical models* (2 ed.). John Wiley.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Lawrence Erlbaum Associates.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2003, July 14). *Piora na qualidade do ensino afeta mais estudantes negros*.
- Jesus, G. R. de & Laros, J. A. (2004). Eficácia escolar: Regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, 3(2), 93-106.
- Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. Sage Publications.

- Laros, J. A. & Marciano, J. L. P. (2008). Análise multinível aplicada aos dados do NELS:88. *Estudos em Avaliação Educacional*, 19(40), 263-278. <https://doi.org/10.18222/ae194020082061>
- Laros, J. A., Marciano, J. L. P. & Andrade, J. M. de. (2010). Fatores que afetam o desempenho na prova de Matemática do SAEB: Um estudo multinível. *Avaliação Psicológica*, 9(2), 173-186. <https://doi.org/10.15689/ap.2010.0902.04>
- Natis, L. (2001). Modelos lineares hierárquicos. *Estudos em Avaliação Educacional*, 3(5), 5-24.
- Palermo, G. A., Silva, D. B. do N. & Novellino, M. S. F. (2014). Fatores associados ao desempenho escolar: Uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. *Revista Brasileira de Estudos de População*, 31(2), 367-394. <https://doi.org/10.1590/S0102-30982014000200007>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2 ed.). Sage Publications.
- Riani, J. de L. R. & Rios-Neto, E. L. G. (2008). Background familiar versus perfil escolar do município: Qual possui maior influência no resultado educacional dos alunos brasileiros? *Revista Brasileira de Estudos Populacionais*, 25(2), 251-269. <https://doi.org/10.1590/S0102-30982008000200003>
- Rodrigues, F. C. & Monteiro, M. A. A. (2024). Um modelo para avaliação do argumento de prova em contextos de ensino baseado em argumentação coletiva. *Revista Internacional de Pesquisa em Educação Matemática*, 14(1), 1-20. <https://doi.org/10.37001/ripem.v14i1.3214>
- RStudio Team. (2023). *RStudio: Integrated development environment for R*. <http://www.rstudio.com/>
- São Paulo (Estado). Secretaria da Educação. (2019). *Relatório pedagógico: SARESP 2019 - Matemática*.
- Soares, J. F. & Alves, M. T. G. (2003). Desigualdades raciais no sistema brasileiro de educação básica. *Educação e Pesquisa*, 29(1), 147-165. <https://doi.org/10.1590/S1517-97022003000100010>
- Soares, J. F. & Andrade, R. J. D. (2006). Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Ensaio: Avaliação e Políticas Públicas em Educação*, 14(50), 107-125. <https://doi.org/10.1590/S0104-40362006000100007>
- Soares, J. F., Cesar, C. C. & Mambrini, J. (2001). Determinantes de desempenho dos alunos do ensino básico brasileiro. In F. Crespo (Org.), *Avaliação, Ciclos e Promoção na Educação* (pp. 121-154). Artmed Editora.
- Soares, T. M. (2003). Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos em Avaliação Educacional*, 28, 103-124. <https://doi.org/10.18222/ae02820032180>
- Soares, T. M. (2005). Modelo de três níveis hierárquicos para a proficiência dos alunos de 4ª série de língua portuguesa do SIMAVE/PROEB-2002. *Revista Brasileira de Educação*, 29, 73-87. <https://doi.org/10.1590/S1413-24782005000100006>
- Soares, T. M. & Mendonça, M. C. (2003). Construção de um modelo de regressão hierárquico para os dados do Simave-2000. *Pesquisa Operacional*, 23(3), 421-441.