

Análise do impacto de variáveis contextuais e socioeconômicas sobre a proficiência em avaliação de desempenho de estudantes do Ensino Fundamental de São Paulo

Gisela Sprovieri¹, Seiji Isotani², Bruno Elias Penteadó³

Abstract

This study sought to ascertain the impact of contextual and socioeconomic variables on proficiency in the Portuguese Language and Mathematics subjects of students from the 7th grade of elementary school, participating in Saresp in 2018. Through the application of linear regression (MDE), we seek to quantify the impact of the most significant variables in the composition of the students' performance in the evaluation, and in this way, relate the socioeconomic and contextual variables to proficiency in Portuguese Language and Mathematics. Opening a discussion about the use of public data generated by schools is extremely important since it can point out possible problems related to the teaching-learning process that may be responsible for mitigating their performance. Among the positive results generated by this model, we verify: the father's education level; the presence of technology (cell phone) at home and family income. Among the negatives we have: the parents' lack of knowledge about their children's school activities (homework); failure; the presence of television at home, in addition to the student's negative view of his future in studies when he believes he must stop studying to work.

Resumo

Este estudo buscou averiguar o impacto de variáveis contextuais e socioeconômicas sobre a proficiência em Língua Portuguesa e Matemática de estudantes dos 7º anos do Ensino Fundamental, participantes do Saresp em 2018. Por meio da aplicação da regressão linear (MDE), buscamos quantificar o impacto das variáveis mais significativas na composição do desempenho dos estudantes na avaliação, e dessa forma, relacionar as variáveis socioeconômicas e contextuais à proficiência em língua portuguesa e matemática. Abrir uma discussão acerca do aproveitamento de dados públicos gerados pelas escolas mostra-se de extrema importância já que pode apontar possíveis problemas relacionados ao processo ensino-aprendizagem que possam ser responsáveis por mitigar seu desempenho. Dentre os resultados positivos gerados por este modelo, verificamos: o nível de escolaridade do pai; a presença em casa de tecnologia (celular) e a renda familiar. Dentre os negativos, estão: a falta de ciência dos pais sobre as atividades escolares de seus filhos (tarefa de casa); a reprovação; a presença em

1 Pós-Graduando(a) em Computação Aplicada à Educação, USP, gispro@usp.br.

2 Prof. Dr. Seiji Isotani, ICMC-USP, sisotani@icmc.usp.br.

3 Prof. Dr. Bruno Elias Penteadó, ICMC-USP, brunopenteadó@usp.br.

domicílio da televisão, além da visão negativa do estudante em relação ao seu futuro nos estudos quando acredita ter que parar os estudos para trabalhar.

1. Introdução

São inúmeros os trabalhos na literatura acadêmica que procuram dimensionar o impacto do engajamento (e/ou interação, colaboração) sobre o desempenho dos estudantes em cursos online (EAD). Assim como, vários autores já buscaram dimensionar o impacto de inúmeras outras variáveis sobre o desempenho dos estudantes no Ensino Fundamental. Assim o fizeram Teixeira e Kassouf (2015) com a violência, Bassetto (2019) com a renda familiar, Soares et al (2010) com a expectativa do professor, Brooke (2013) com a equidade e outros impactos dos incentivos monetários para professores, Arruda (2014) com fatores familiares e socioeconômicos, tendo por base o Saresp 2012, entre outros. Este último, buscando por meio da associação (MDE) a relação dessas variáveis com o desempenho de estudantes do Ensino Fundamental I, mais especificamente, de 3ºs e 5ºs anos, de escolas públicas de São Paulo.

O Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) é uma avaliação anual aplicada pela Secretaria da Educação do Estado de São Paulo com a finalidade de diagnosticar e acompanhar a evolução da educação básica paulista. No Saresp, os alunos do 3º, 5º, 7º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio têm seus conhecimentos avaliados por meio de provas com questões de Língua Portuguesa e Matemática [SARESP, 2018].

Como muito bem destaca ARCAS (2009), tratar hoje da avaliação não é um trabalho simples. Nos anos finais do século XX e anos iniciais do século XXI, segundo ele, a avaliação ganhou destaque nas políticas educacionais empreendidas no Brasil. Ao contrário simplesmente de falarmos nas avaliações tradicionais, hoje falamos de inúmeras concepções e formas de avaliação, como: da aprendizagem, de sistemas, de programas, em larga escala, institucional, entre outros focos [ARCAS, 2009].

O Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp) foi adotado em 1996 introduzindo uma prática de avaliação institucional, externa, aplicada a toda a rede estadual de ensino de São Paulo que engloba 5640 escolas e 3.356.104 alunos, conforme publicado no site Dados Abertos da Educação. Arcas (2009) destaca o dilema gerado por sua associação, em 1998, ao regime de progressão continuada que propunha uma concepção formativa à avaliação. A influência dessas duas medidas sobre as práticas escolares e as concepções de avaliação adotadas a partir delas pelas escolas podem ser sentidas a desde então. Santos e Sabia (2015) destacam algumas das implicações e consequências do Saresp sobre o trabalho docente, tais como: utilização do Saresp como critério de aprovação (ou não) em final de ciclo escolar; utilização como modelo para elaboração de novas provas; realização do treinamento dos alunos por parte dos professores para a avaliação; utilização da avaliação externa como avaliação da aprendizagem; direcionamento e controle do trabalho docente; comercialização de notas (atribuição de pontos extras caso os alunos realizem a avaliação no esforço de professores para que os alunos participem das avaliações institucionais externas); foco estritamente nos componentes curriculares; currículo prescrito; regulação do trabalho docente; e limitação da autonomia do professor em sala de aula (SANTOS e SABIA, 2015).

Posto isso, é bom entendermos esse contexto em que o Saresp foi criado. Podemos verificar, segundo TORRES (2001) que, na década de 90, o sistema de avaliação da educação

básica passa a inserir-se em um conjunto mais complexo de inter-relações, e em cujo interior opera o aprofundamento das políticas de descentralização administrativa, financeira e pedagógica da educação, além de um novo aparato legal e uma série de reformas curriculares [TORRES, 2001].

Ora, se por um lado, a adoção da progressão continuada estimulou a concepção formativa da avaliação do rendimento dos estudantes, e por outro, o Saresp estimulou sua concepção tradicional, procurando quantificar, ou simplesmente, classificar os estudantes dentro de um sistema de referências do que se pretende desejável e esperado sobre o rendimento dos estudantes, qual seria a melhor forma de tratamento desses dados coletados advindos de sua aplicação? Qual seria uma boa opção para sua exploração de maneira a acabar ou minimizar o dilema? É neste contexto que propomos o tratamento desses dados pela MDE, de modo a criar um modelo de regressão linear múltipla em que várias variantes independentes, explanatórias - as condições socioeconômicas - possam ser relacionadas à proficiência dos estudantes em português e matemática (variante dependente), quando tomada como referência.

Dessa forma, esperamos que esta análise possa contribuir para ampliar a discussão em torno do aproveitamento e uso das informações obtidas a partir dos dados gerados no Saresp e da aplicação dos questionários socioeconômicos, e que desta feita, eles possam contribuir concretamente para a reflexão docente acerca do papel da avaliação no processo de ensino e aprendizagem, orientando novas medidas no que tange ao processo de ensino e à avaliação escolar.

2. Revisão da Literatura

Foi realizada no Google Scholar, utilizando progressivamente termos que restringiram os trabalhos, como: "desempenho escolar" ou proficiência em português e matemática, ensino fundamental, Saresp, variáveis socioeconômicas, regressão linear, 6º ao 9º ano, São Paulo, na busca de trabalhos que se assemelhassem a nossa proposta.

Esta breve revisão na literatura, restrita aos últimos dez anos, aponta para 4580 estudos sobre o desempenho ou proficiência em língua portuguesa e matemática dos estudantes do Ensino Fundamental. Destes, 645 relacionados ao Saresp (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo), 591 relacionados à MDE (Mineração de Dados Educacionais) e 217 relacionam o desempenho dos estudantes no Ensino Fundamental às variáveis socioeconômicas. Quando especificado o algoritmo "regressão linear", chegamos a 51 estudos, um número bastante reduzido se considerarmos o período a que o levantamento se refere, o que sugere ainda uma recente aplicação da técnica de regressão linear a esses dados públicos.

2.1. Trabalhos relacionados

Destes 51 trabalhos, dois foram os que mais se aproximaram de nossa proposta: o de Osio (2013), pois desenvolve e aplica modelos de regressão multinível simétricos, a fim de avaliar se o número de reprovações no histórico escolar do aluno e a infraestrutura da escola são variáveis relevantes que afetam o baixo desempenho dos alunos do ensino básico na disciplina de Matemática (OSIO, 2013) e o de Comim (2016), que busca relacionar violência na escola e proficiência dos alunos no Brasil nos anos de 2003, 2005, 2007, 2009, 2011 e 2013 com aplicação da regressão linear, concluindo que a ocorrência de um tipo de evento violento adicional está associada a uma redução na proficiência. No entanto, destes 51 trabalhos a que

chegamos, apenas três se referem a dados públicos de São Paulo relacionados ao Saresp, ou seja, dados da Educação de São Paulo, relacionando-os a variáveis contextuais ou socioeconômicas com a aplicação da regressão linear, o que aponta para um campo ainda a ser explorado.

3. Metodologia

Rodrigues et al (2013) destacam a técnica de análise de correlação como sendo muito útil em trabalhos exploratórios, quando o interesse está no grau ou força do relacionamento [RODRIGUES et al, 2013]. Este trabalho busca estabelecer a correlação linear de modo que consigamos dimensionar, ou ainda, quantificar o impacto de cada variável independente, do total de variáveis socioeconômicas e contextuais em questão, sobre o desempenho (variável dependente) dos estudantes do Ciclo II, do Ensino Fundamental, que participaram do Saresp 2018.

Este trabalho apresenta uma análise da aplicação do algoritmo de regressão linear, utilizando o RStudio a dados públicos relativos ao Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp/2018), de 4803 alunos matriculados em classes regulares do 7º ano, que tenham participado das avaliações, de Português e Matemática, e tenham concomitantemente, preenchido os [questionários socioeconômicos, de Aluno](#) e [Pais](#). Essa exigência foi necessária para que coletássemos o máximo de informações sobre os estudantes e seu desempenho e, por meio da análise dos resultados, buscássemos o impacto direto da relação entre condições socioeconômicas sobre os índices de proficiência em português e matemática.

O RStudio, é um software livre de ambiente de desenvolvimento integrado para R, com linguagem de programação para gráficos e cálculos estatísticos. Os dados utilizados nesta pesquisa são públicos, como já mencionado, obtidos em dois bancos de dados da Secretaria Estadual da Educação de São Paulo. O primeiro deles contém os resultados do Saresp 2018, dos 7ºs anos, do Ensino Fundamental II e o segundo apresenta os questionários socioeconômicos de alunos e pais que participaram desta avaliação no referido ano (2018).

Para identificação das questões que compõem cada questionário socioeconômico de pais e aluno foram mantidas as nomenclaturas, ou ainda, os códigos utilizados pelo Estado de São Paulo em seus bancos de dados públicos. Para eventual consulta, disponibilizamos o [dicionário de dados do Saresp](#).

3.1. Objetivos

Buscamos neste trabalho a natureza da relação entre “nota” e as condições socioeconômicas dos estudantes. Conhecer, dimensionar e atribuir significado ao impacto dessas condições sobre o desempenho dos estudantes talvez contribua para que as escolas públicas reflitam sobre o atendimento às demandas de sua comunidade escolar de forma mais assertiva e eficiente.

Esperamos, por meio da aplicação da regressão linear aos dados públicos em questão, compreender melhor como os resultados obtidos no Saresp 2018 são impactados pelas variáveis socioeconômicas e quanto cada uma representa, em pontos, ganho ou perda no nível de proficiência dos estudantes em língua portuguesa e matemática.

3.2. Materiais e método

O questionário do aluno, como podemos observar no **Quadro 1**, é dividido em cinco blocos de perguntas, cada bloco direcionado a um aspecto da vida do estudante, conforme os itens a seguir: Bloco 1 – Opinião sobre a escola; Bloco 2 -Vida Escolar; Bloco 3 - Participação dos pais (ou responsáveis); Bloco 4 - Hábitos de estudo e leitura; Bloco 5: Clima escolar.

O questionário dos pais também é dividido também em cinco blocos: Bloco 1: Opinião sobre a escola; Bloco 2: Hábitos de estudo, leitura e vida escolar; Bloco 3: Participação dos pais (ou responsáveis); Bloco 4: Informações básicas e, finalmente, Bloco 5: Renda e acesso a bens/serviços domésticos. Questões relacionadas à renda e acesso a bens e serviços domésticos são apresentadas apenas aos pais, no Bloco 5, conforme demonstrado no **Quadro 2**.

As variáveis presentes na correlação entre índice de proficiência, de português e matemática e dados socioeconômicos são demonstradas no **Quadro 1**, com base nos questionários de pais e alunos.

Quadro 1. Dos questionários socioeconômicos aplicados aos alunos

QUESTIONÁRIO	BLOCO	Questões	VARIÁVEIS (gerais)
ALUNOS	1 – OPINIÃO SOBRE A ESCOLA	ALUNO Q1 a Q8	<ul style="list-style-type: none"> ● satisfação com a escola ● condições ambientais da escola
	2 – VIDA ESCOLAR	ALUNO Q9 e Q10	<ul style="list-style-type: none"> ● início dos estudos ● reprovação
	3 - PARTICIPAÇÃO DOS PAIS (OU RESPONSÁVEIS)	ALUNO Q17 a Q33	<ul style="list-style-type: none"> ● incentivo dos pais a ler e estudar ● acompanhamento de tarefas ● manifestação de interesse sobre o que acontece na escola ● hábitos de leitura dos pais
	4 – HÁBITOS DE ESTUDO E LEITURA	ALUNO Q19 a Q22	<ul style="list-style-type: none"> ● formação do pai ● formação da mãe ● situação de trabalho do pai ● situação de trabalho da mãe
	5 – CLIMA ESCOLAR	ALUNO Q34 a Q49	<ul style="list-style-type: none"> ● Tipos de violência sofrida ● previsão do futuro ● da relação entre e com professores ● da relação entre e com alunos ● da relação entre e com a direção ● da relação entre e com funcionários

Quadro 2. Dos questionários socioeconômicos aplicados aos pais

QUESTIONÁRIO	BLOCO	Questões	VARIÁVEIS (gerais)
--------------	-------	----------	-----------------------

PAIS	1 – OPINIÃO SOBRE A ESCOLA	PAIS Q1 a Q8	<ul style="list-style-type: none"> • satisfação com a escola • condições ambientais da escola
	2 - HÁBITOS DE ESTUDO, LEITURA E VIDA ESCOLAR	PAIS Q9 a Q11	<ul style="list-style-type: none"> • tarefa de casa • uso de tecnologia
	3 - PARTICIPAÇÃO DOS PAIS (OU RESPONSÁVEIS)	PAIS Q12 a Q18	<ul style="list-style-type: none"> • interesse dos pais • acompanhamento de tarefas • diálogo com professores • participação em eventos da escola
	4 - INFORMAÇÕES BÁSICAS	PAIS Q19 a Q22	<ul style="list-style-type: none"> • formação do pai • formação da mãe • situação de trabalho do pai • situação de trabalho da mãe
	5 - RENDA E ACESSO A BENS/SERVIÇOS DOMÉSTICOS	PAIS Q23 a Q50	<ul style="list-style-type: none"> • renda familiar • posse de materiais de leitura • acesso à internet e TV a cabo • posse de bens de consumo

O **Quadro 1** apresenta resumidamente as variáveis depreendidas a partir das questões que compõem o questionário socioeconômico aplicado aos alunos, além de sua organização em blocos, em que cada um aborda um setor específico da vida do estudante relacionado aos estudos.

O **Quadro 2** apresenta resumidamente as variáveis depreendidas a partir das questões que compõem o questionário socioeconômico aplicado aos pais. Ele demonstra a organização do questionário em blocos, assim como ocorre com o do aluno. Os temas abordados relacionam-se com o desempenho dos pais em sua função de proteção, cuidados e orientação de seus filhos que diretamente ou indiretamente podem influenciar no desempenho e desenvolvimento dos estudantes nos estudos.

3.2.1. Sobre a fase de pré-processamento dos dados

Inicialmente, a planilha de dados contava com 137 atributos, sendo que 38 se relacionavam à identificação e condições da prova (do aluno, escola, diretoria de ensino e caderno de prova). Destes 38 atributos foram preservados apenas 06, além 99 atributos relacionados aos questionários socioeconômicos, totalizando assim 105 atributos: região metropolitana (região metropolitana); período (manhã ou tarde) sexo (tp_sexo); presença ou não de deficiência

(tem_nec) e proficiência de Português (profic_lp) e proficiência de Matemática (profic_mat) e período, totalizando ao final, 105 atributos selecionados para esta análise.

Ainda na etapa de pré-processamento dos dados, 32 do total de 38 atributos relacionados a identificação do aluno (nomedep, dt_nascimento), turma (turma, tipoclasse, serie, classe, periodo), diretoria de ensino (cdrede, de_codesc,) nome da rede (nomedepbol), série (todos do 7º ano) e cidade (codmun, mun) foram excluídos. Além destes, foram excluídos os atributos relacionados aos tipos de deficiência presentes entre os alunos do 7º ano que participaram das avaliações (nec_esp_1, nec_esp_2, nec_esp_3, nec_esp_4, nec_esp_5), os relacionados ao tipo de prova (tip_prova), ao número do caderno de prova (cad_prova_mat e cad_prova_lp), participação ou não na prova de português (particip_lp), participação ou não na prova de matemática (particip_mat). Essa escolha se deve à seleção prévia dos alunos que tenham participado das duas avaliações (português e matemática).

Informações relacionadas à “data de nascimento”, “período”, “código de deficiência” (tipos); “código de diretoria de ensino”, “nome da DE”, “código de município”, “nome do município” foram desprezadas por não serem relevantes ao estudo em foco. Nossa intenção foi evitar qualquer informação redundante ou que possibilitasse a identificação das escolas ou estudantes. O atributo “tem nec” (indicativo se tem ou não deficiência) foi preservado. No entanto, os atributos relativos aos tipos de deficiência (de nec1 a nec 5) foram ignorados. Isso se deve ao fato de que o total de alunos portadores de qualquer um dos tipos de deficiência, entre os alunos do 7º ano em 2018, e que participaram deste processo avaliativo somaram apenas 14 indivíduos (0,29%); um número pouco expressivo em meio ao total de estudantes em análise, o que impossibilitaria estabelecer qualquer correlação entre sua aprendizagem e a influência de sua deficiência no processo avaliativo. Sendo assim, os 14 estudantes fizeram parte do estudo sem levarmos sua condição em consideração. Outros atributos foram excluídos, como: nome e código da escola e do aluno, porque não há interesse na identificação das unidades escolares para que não haja qualquer possibilidade de discriminação por parte do estudo, e “etapa de ensino” e “série” porque todos os dados selecionados se referem aos 7ºs anos do Ensino Fundamental, do estado de São Paulo, e que participaram das avaliações de matemática e língua portuguesa do Saresp em 2018, como já foi dito anteriormente.

Das variáveis relativas aos resultados, foram excluídas aquelas relacionadas à porcentagem de acertos em língua portuguesa e matemática (porc_acert_lp; porc_acert_mat); nível, classificação e número de pontos em português e matemática (nivel_profic_lp, nivel_profic_mat, classific_lp, classific_mat, total_ponto_lp e total_ponto_mat). Como já foi mencionado, para este estudo foi selecionado os atributos relacionados à proficiência (profic_lp e profic_mat) para referência ao desempenho dos alunos nas duas avaliações, por se tratar de uma variável que possibilita a quantificação das variáveis socioeconômicas quando associadas a ele. Neste caso, a proficiência obedece a critérios de classificação do nível de proficiência do estudante de acordo com a pontuação obtida nas avaliações, neste caso, de língua portuguesa e matemática, como indicado no **Quadro 3**.

3.3. Do resultado do Saresp 2018

De acordo com a escala de valores pré-definida pelo governo do estado para a classificação dos estudantes em níveis de proficiência, como mostra o **Quadro 3**, prevê-se que o estudante que obtenha menos até 175 pontos em LP domine um número mínimo de habilidades previstas para o componente curricular. Na medida que sua pontuação se desloque para faixas mais elevadas

de pontuação, prevê-se que o domínio de habilidades diversas e mais complexas aumenta substancialmente, indo do nível insuficiente ou abaixo do básico (até 175 pontos) até o avançado, maior ou igual a 300 pontos em língua portuguesa. Em matemática, até 200 pontos, correspondendo ao nível insuficiente ou abaixo do básico até o avançado com pontuação igual ou maior que 300 pontos.

Quadro 3. Da classificação e pontuação relativa aos 7ºanos

PROFICIÊNCIA	CLASSIFICAÇÃO	PONTUAÇÃO
NIVEL_PROFIC_LP (5º ano EF a 3ª série EM em todos os componentes)	ABAIXO DO BÁSICO	< 175
	SUFICIENTE/BÁSICO	175 a 224
	ADEQUADO	225 a 274
	AVANÇADO	>= 275
NIVEL_PROFIC_MAT (5º ano EF a 3ª série EM em todos os componentes)	ABAIXO DO BÁSICO	< 200
	SUFICIENTE/BÁSICO	200 a 249
	ADEQUADO	250 a 299
	AVANÇADO	>= 300

O **Quadro 4** apresenta a média de proficiência obtida, em pontos, pelos estudantes no Saesp 2018. Ou seja, a proficiência mínima, média e máxima dos estudantes obtida em cada componente curricular (língua portuguesa e matemática). Ou seja, a proficiência mínima dos estudantes em LP foi 105,2 pontos, a máxima 338,4 pontos, enquanto a média foi de 235,518. Em matemática, a proficiência mínima foi de 114,4 pontos, a máxima de 375,5 pontos, enquanto a média ficou em 243,006 pontos.

Quadro 4. Da proficiência em LP e Mat no Saesp 2018

Proficiência em língua portuguesa	profic_lp	Mínimo	105.2
		Máximo	338.4

		Média	235.518
		Desvio padrão	44.937
Proficiência em matemática	profic_mat	Mínimo	114.4
		Máximo	375.5
		Média	243.066
		Desvio padrão	44.134

3.4. Sobre a Regressão Linear Composta

A regressão linear é expressa como função linear em que uma ou mais variáveis, neste caso, as respostas aos questionários socioeconômicos de pais e alunos, denominadas variáveis explanatórias (tomadas aqui como variáveis independentes) são relacionadas a uma variável dependente, a proficiência em português e matemática, considerando-se um provável índice de erro (ϵ). Este se espera que seja nulo, ou pelo menos, próximo disso. Neste caso, o presente estudo considera os dados como representativos de uma população, alunos do 7º ano, que seguem os critérios pré-estabelecidos para este estudo. A regressão linear neste caso, não pode servir de amostra, ou seja, não pode ser generalizada automaticamente a todos alunos dos 7ºs anos que participaram ou participarão do Sabesp em outros anos. Para tal conclusão, este estudo deverá ser aplicado a outras turmas de alunos para verificação de uma possível recorrência [MATOS, 1995].

Como bem pode exemplificar, nosso estudo apresenta uma variável dependente, numérica (proficiência dos estudantes) relacionada à inúmeras variáveis explanatórias, qualitativas, ou ainda, categóricas ou nominais, advindas dos questionários socioeconômicos de pais e alunos. Trata-se aqui, portanto, de uma regressão estimada como composta, ou seja, por um conjunto de variáveis explanatórias que compõem um modelo de regressão linear múltipla. Ela pode ser representada pela seguinte função:

$$y_i = a + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} + \epsilon_i \quad i=1..n$$

3.5. Participantes

Participaram do Saesp, em 2018, o total de 39.214 estudantes do 7º ano, locados nas sete regiões em que o estado de São Paulo é dividido (região metropolitana de São Paulo, região de Campinas, interior, região de Sorocaba, região metropolitana de Ribeirão Preto, Baixada Santista, e finalmente, a região do Vale do Paraíba e Litoral Norte).

Desse total (39124 estudantes), foram selecionados apenas 4803 que preencheram todos os quesitos necessários ao presente estudo: ter participado das duas avaliações (língua portuguesa e matemática) e ter preenchido os questionários socioeconômicos (o de pais e o de aluno). Estas condições foram estabelecidas de modo a propiciar o levantamento do maior número possível de informações relativas às condições contextuais e socioeconômicas de cada aluno para então relacioná-las a sua proficiência (de português e matemática).

Sendo assim, foram descartados todos os estudantes que não participaram de pelo menos uma das avaliações, e/ou naturalmente das duas, e não responderam a pelo menos um dos questionários socioeconômicos (de pais ou aluno).

Como mostra a **Tabela 1**, ao final do processo de seleção dos participantes válidos para o estudo, obtivemos: 1160 estudantes da região metropolitana de São Paulo, 260 da região metropolitana de Campinas, 2630 do Interior, 268 da região metropolitana de Sorocaba, 95 da região Metropolitana de Ribeirão Preto, 204 da região metropolitana da Baixada Santista 204, e finalmente, 186 da região Metropolitana do Vale do Paraíba e Litoral Norte. Destes, 47,42% pertencem ao gênero feminino e 52,57% ao gênero masculino. 1733 estudantes (36,08%) pertencem ao turno da manhã, enquanto 3070 (63,91%), ao turno da tarde.

Tabela 1. Dos participantes

INFORMAÇÕES	ATRIBUTO	VARIÁVEIS	População	%
Região a que pertence o aluno	Região Metropolitana	1. Região Metropolitana de São Paulo	1160	24,15
		2. Região Metropolitana de Campinas	260	5,41
		3. Interior	2630	54,75
		4. Região Metropolitana de Sorocaba	268	5,57
		5.Região Metropolitana de Ribeirão Preto	95	1,9
		6. Região Metropolitana da Baixada Santista	204	4,24
		7. Região Metropolitana do Vale do Paraíba e Litoral Norte	186	3,87
Sexo do aluno	TP_SEXO	1. Feminino	2278	47,42
		2. Masculino	2525	52,57
Período em que o aluno é matriculado	PERÍODO	Manhã	1733	36,08
		Tarde	3070	63,91

Presença ou ausência de deficiência	Tem_Nec	0. Não tem	4789	99,70
		1. Tem	14	0,29

4. Resultados

4.1. Discussão dos resultados obtidos relacionados à proficiência em língua portuguesa

Das 105 variáveis submetidas à análise, apenas 80 foram consideradas significativas para a composição da proficiência dos estudantes em língua portuguesa. Neste caso, a região a que pertence o estudante não representou nenhum impacto sobre seu resultado. Pertencer ao sexo masculino gerou um impacto positivo da ordem de 5,503586 pontos, ou ainda, um pouco mais de cinco pontos e meio, sobre seu desempenho. Pertencer ao período da tarde gerou um impacto negativo em seu desempenho, da ordem de -3,122611 pontos. Possuir algum tipo de deficiência gerou um forte impacto negativo sobre o desempenho do estudante, ou seja, -24,238692 pontos. No entanto, a sua baixa ocorrência entre os estudantes, apenas 14, torna esse impacto muito pouco significativo. O mesmo ocorrendo em matemática, como veremos a seguir. Podemos sintetizar o impacto das variáveis socioeconômicas, de aluno e pais, como mostra a **Tabela 3**, sobre o impacto na proficiência em língua portuguesa (conforme APÊNDICE A).

4.1.2. Com referência à proficiência em língua portuguesa

Com base nos resultados apresentados na **Tabela 4** sobre a correlação entre as variáveis socioeconômicas dos pais e a proficiência em língua portuguesa, podemos concluir que a tarefa de casa exerce enorme impacto negativo sobre a proficiência quando os pais demonstram pouco ou nenhum interesse em acompanhá-las. Sendo assim, podemos verificar que ele foi de -6,924369 a -15,335080 pontos.

O fato de os pais não conversarem com os professores de seu filho (14,9%), não resulta em perda sobre seu desempenho em língua portuguesa, pelo contrário, ainda apresenta um ganho razoável (da ordem de 4,656689 pontos), mas não participar ativamente das reuniões de Conselho (18,55%), por exemplo, apresenta uma perda de -3,937036 pontos. Esse resultado pode ser explicado se considerarmos que são nessas reuniões que são dadas orientações aos pais de como proceder em relação ao acompanhamento da vida escolar de seus filhos durante as atividades, estudos em casa e ao que ocorre na escola, de modo geral.

Quanto à formação do pai, seu impacto sobre a proficiência em língua portuguesa mostrou-se bastante expressiva, de 5,198050 a 11,712921, quanto maior seu nível de formação.

Quanto maior a renda familiar, maior o impacto positivo sobre a proficiência, de 7,617628 a 10,933721. No entanto, o fato de desconhecer a renda da família também surtiu um efeito positivo no desempenho, mesmo que consideravelmente menor, 3,697077 pontos apenas.

O fato de no domicílio haver livros (pais_q27) não garantiu ganho na proficiência em língua portuguesa, pelo contrário, houve uma perda de -2,741712 pontos. Esse fenômeno talvez possa ser explicado pelo fato de que a presença de livros no domicílio não seja garantia de que o aluno os esteja lendo.

Sobre a posse de bens e serviços, em geral, demonstraram causar grande impacto no desempenho dos estudantes, o que se justificaria pela melhora da qualidade de vida e conforto

das famílias. As exceções ocorrem, no entanto, quando da posse das máquinas de lavar e secar roupas em que o impacto sobre o desempenho é da ordem de -7,115033 e -3,039771 pontos, respectivamente.

A posse de computador e celulares (smartphones) gera um ganho no seu desempenho (no Saresp/2018, neste caso), de 6,374566, 7,892511, 9,141549 ou 8,164309, ou seja, quanto maior a presença das tecnologias no domicílio, maior o impacto na proficiência em língua portuguesa.

Considerando a visão do aluno, sua satisfação com a qualidade de ensino oferecida pela escola equivale a um impacto na proficiência de língua portuguesa da ordem de 8,757509 pontos. No entanto, quando indicada sua insatisfação, ela ainda é responsável por 7,930042 positivos. Este índice talvez seja responsável por suplantar a insatisfação indicada pelos pais sobre a estrutura da escola, mais especificamente de suas condições ambientais, e com isso, mesmo satisfeitos ou insatisfeitos, ainda são responsáveis por 4,603546 a 4,171774 sobre o desempenho em língua portuguesa.

Ter sido reprovado até uma vez então, no 7º ano do Ensino Fundamental, Ciclo II, gera um impacto negativo da ordem de -17,026977 sobre sua proficiência final em LP. Caso tenha sido reprovado duas vezes, esse impacto negativo se amplifica para 22,697517 pontos negativos.

4.2. Discussão sobre os resultados obtidos relacionados à proficiência em matemática

Das 105 variáveis submetidas à análise, apenas 75 foram consideradas significativas para a composição da proficiência dos estudantes em matemática. Neste caso, apenas a região metropolitana de São Paulo representou impacto sobre seu desempenho em matemática, da ordem de -8,882245. Pertencer ao sexo masculino gerou impacto negativo, ao contrário do observado em português (de -6,744169 pontos), ou ainda, quase 7 pontos do total de seu desempenho em matemática. Pertencer ao período da tarde gerou um impacto negativo em seu desempenho, da ordem de -5,385508 pontos. Possuir algum tipo de deficiência também gerou forte impacto negativo sobre o desempenho desses estudantes em matemática, ou seja, -34,071260 pontos, mas com baixa ocorrência como já explicado anteriormente quando tratamos do desempenho em português de estudantes com deficiência. Podemos sintetizar o impacto das demais variáveis socioeconômicas, de aluno e pais, sobre a proficiência em matemática como pode ser verificado na **Tabela 4. Sobre o impacto na proficiência em Matemática (APÊNDICE B)**.

4.2.1. Com referência à proficiência em matemática

Obtivemos um resultado muito próximo ao de língua portuguesa, quanto à satisfação dos pais em relação à escola (pais_q1), ou seja, essa variável apesar de frequente entre os respondentes (61,77% em cada uma das disciplinas), não correspondeu a um impacto positivo sobre o desempenho de seus filhos, pelo contrário, resultou em -3,458848 em sua proficiência de matemática. Este impacto sobre o desempenho dos estudantes talvez esteja relacionado ao papel de autoridade imposto pelos pais, já que a eles, em geral, cabe a prerrogativa da escolha da escola em que o filho deverá estudar. Ao contrário, quando questionado sobre o espaço para esportes, a resposta satisfatória do pai equivale a 6,373114, um resultado razoável, mas apenas para 6,76% dos estudantes.

As respostas dos pais sobre a frequência com que os estudantes realizam suas tarefas de casa (pais_q10) demonstrou que fazer tarefa de casa esporadicamente (às vezes) representa um impacto de -7,983555 a -9,414092 sobre a proficiência dos estudantes em matemática. Mas ao afirmarem que o estudante “nunca faz a tarefa” ou “a escola não passa lição de casa”, as respostas dos pais representam um impacto tremendamente maior sobre seu desempenho em matemática (de -15,727681 e -10,052149, respectivamente), mesmo considerando sua baixa frequência entre os estudantes (2,35% e 3,41, respectivamente).

Dos pais dos 4803 estudantes, 4013 afirmam que seus filhos não fazem uso do aplicativo Minha Escola SP, ou seja, 83,55% dos estudantes em estudo. Esta alta frequência de resposta acaba por impactar positivamente seu desempenho em matemática, 6,061281.

Quanto ao nível de escolaridade do pai ou responsável (pais_q19), a resposta “com especialização, mestrado ou doutorado” dada pelos pais, apontou um acréscimo no desempenho em matemática dos estudantes da ordem de 11,577479 pontos, apesar de pouco frequente (1,79%).

Para o atributo pais_q22, a respeito da situação de trabalho da mãe, sua ocorrência apesar de baixa (2,76%), apontou para um ganho substancial na proficiência em matemática dos estudantes, de 8,607597. Enquanto, por sua vez, a renda familiar (pais_q23) demonstrou ser uma das variáveis de maior impacto sobre a proficiência em matemática dos estudantes. O resultado apontou que quanto maior a renda das famílias, melhor é seu desempenho (conforme nos mostra a Tabela 7).

Das variantes mais significativas do questionário aplicado aos alunos, a reprovação (aluno_10) mostrou-se de grande impacto sobre a proficiência dos estudantes em matemática. O resultado apontou que quanto mais reprovações tanto maior é a perda em seu desempenho (-16,143003, para uma reprovação e -20,416138, para duas ou mais). Sobre o incentivo e participação dos pais nos estudos e realização das tarefas de casa, as respostas da maioria dos alunos apontaram a tendência à autonomia dos estudantes para a realização das tarefas, pois mesmo não os ajudando a realizá-las ou conferindo as tarefas realizadas pelos estudantes, a variante apresenta um impacto positivo sobre seu desempenho em matemática de 9,470331 (aluno_13), com uma frequência de 44,72% entre os estudantes.

Na visão do aluno, o pouco incentivo dos pais para que frequentem a escola e não faltem às aulas (aluno_14) gerou um impacto negativo de -5,385843, para 9,14% dos estudantes.

Usar um pouco a internet para reforçar os estudos ou não usar (aluno_18) ainda gera impacto positivo sobre o desempenho em matemática, 4,894749 e 6,553667 pontos, respectivamente. No entanto, usá-la para assistir vídeos às vezes ou nunca (aluno_21), ao contrário, gera um impacto negativo na proficiência de matemática, de -6,258320 e -7,356857, respectivamente.

Não gostar de matemática (aluno_22) não correspondeu necessariamente em baixo desempenho na disciplina. Pelo contrário, gerou um impacto positivo de 13,240634. Fazer lição de casa de matemática às vezes ou nunca (aluno_23) corresponde a -9,934504, apesar dessa ocorrência incidir sobre apenas 7,57% dos alunos analisados. Para 2,22% deles, que afirmam não ter tarefas de casa, o impacto sobre sua proficiência de matemática é de -7,587583.

O fato de não gostar de estudar língua portuguesa (aluno_24) não incidiu negativamente sobre seu resultado em matemática. Pelo contrário, para 901 estudantes essa resposta ainda lhe conferiu 7,335842.

Sobre a forma preferida de estudar (de aluno_27 a aluno_31), a única que gerou impacto positivo sobre o desempenho do aluno em matemática, foi “prestar atenção às aulas”, ou seja, 8,072681 sobre sua proficiência. As demais formas de estudo geraram impacto negativo, inclusive quando “complementada com pesquisas na internet, por programas no computador ou aplicativos no celular”, tendo gerado -5,629265 pontos (negativos).

Preparar-se dias antes para a prova ou no dia anterior (aluno_33), uma recorrência entre 17,21% dos alunos, representa 5,164845 sobre a proficiência em matemática, enquanto não estudar representa 10,347784 pontos para 11,88% deles.

A respeito do clima escolar, o fato de “ser chamado de nomes feios por outros alunos apenas de vez em quando” (aluno_35), sem frequência gera para 45,82% dos estudantes um resultado positivo sobre sua proficiência em matemática, de 6,454041. enquanto isso, a visão dos estudantes sobre seu futuro (aluno_42) gera -9,317138 para 4,33% deles, pois acreditam que devem parar de estudar para trabalhar.

A visão que os estudantes têm acerca da qualidade da relação entre os alunos tem impacto de 6,767128 a 7,753605 pontos sobre seu desempenho; entre professores e direção, 9,445677 e 11,602337; entre alunos e direção, -6,687594 para 48,99% dos estudantes que acreditam ser uma relação boa e -6,749234 para 13,86 que afirmam não ter conhecimento sobre isso e, finalmente, para 10,86% dos estudantes que afirmam não conhecer a relação entre alunos e funcionários da escola, têm -8,213450 sobre seu desempenho em matemática comprometido.

Quando consideradas as variáveis relacionadas ao clima escolar, a baixa ocorrência de violência na escola gera resultados positivos sobre o desempenho em matemática. Este resultado foi confirmado em Comim (2016) citado na revisão de literatura.

4.3. Limitações do estudo

Neste caso, algumas inconsistências nos resultados da aplicação da regressão linear aos atributos relacionados aos itens de posse de bens de consumo (máquina de lavar) apontam para a necessidade de aplicação da Teoria de Resposta ao Item, um recurso estatístico que segundo Andrade, Tavares & Valle (2000) consiste em um instrumento poderoso nos processos quantitativos de avaliação educacional, pelo fato de permitir a construção de escalas de variáveis calibradas, ou seja, a estimação de um parâmetro relacionado aos itens de posse de bens para só então relacioná-los ao atributo de referência (proficiência em língua portuguesa ou matemática) por meio da regressão linear. Dentre os métodos de aplicação da Teoria de Resposta ao Item, o de agrupamento de habilidades talvez ofereça uma alternativa para a solução dos problemas de incongruência nos resultados obtidos (ANDRADE, TAVARES & VALLE, 2000).

5. Considerações Finais

Como vimos até aqui, a Mineração de Dados Educacionais (MDE) pode oferecer grande oportunidade de análise de dados gerados a partir de avaliações e práticas realizadas na própria escola. Desse modo, é possível identificar suas fragilidades, conhecer sua comunidade e

promover mudanças, criando oportunidades para o engajamento dos alunos nas atividades escolares, atendendo suas demandas onde possam estar seus pontos mais frágeis. Além disso, podemos fazer uma leitura mais personalizada desses dados públicos, pois eles traduzem todas as variáveis envolvidas no processo de ensino e aprendizagem, desde dados contextuais específicos de cada escola e de sua comunidade escolar, até as condições socioeconômicas das famílias em que nossos estudantes estão inseridos e de que seu desempenho é fruto.

Podemos concluir destacando que o modelo gerado corresponde a um índice de significância da relação das variantes socioeconômicas e proficiência em língua portuguesa menor que 5% (ou seja, $\leq .05$, 1:4); o mesmo ocorrendo em relação à proficiência em matemática ($\leq .05$, 1:4) o que demonstra a validade do modelo.

Referências

Arcas, P. H. (2009). *Implicações da progressão continuada e do SARESP na avaliação escolar: tensões, dilemas e tendências* (Doctoral dissertation, Universidade de São Paulo).

Arruda, Z. J. (2014). Associação entre desempenho acadêmico no SARESP e fatores familiares e socioeconômicos dos alunos do ensino fundamental I das escolas estaduais de São Paulo.

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.

Brooke, N. (2013). Sobre a equidade e outros impactos dos incentivos monetários para professores. *Estudos em avaliação educacional*, 24(55), 34-62.

de Andrade, D. F. Tavares, H. R., da Cunha Valle, R. (2000). Teoria da Resposta ao Item: conceitos e aplicações. ABE, São Paulo.

dos Santos, U. E., de Pádua Sabia, C. P. (2015). Percurso histórico do Saresp e as implicações para o trabalho pedagógico em sala de aula. *Estudos em Avaliação Educacional*, 26(62), 354-385.

Fossatti, E. C., & Berti, S. M. Regressão Linear Múltipla: Como simplificar por meio do Excel e SPSS? André da Silva Pereira e Thayane Woellner Sviercoski Manosso.

Matos, M. A. (1995). Manual operacional para a regressão linear. *Faculdade de Engenharia da Universidade do Porto*, 63.

Rodrigues, R. L. De Medeiros, F. P., & Gomes, A. S. (2013). Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 24, No. 1, p. 607).

Teixeira, E. C., & Kassouf, A. L. (2015). Impacto da violência nas escolas paulistas sobre o desempenho acadêmico dos alunos. *Economia Aplicada*, 19(2), 221-240.

Torres, R. M. (2001). Educação para todos: a tarefa por fazer. A Resposta.

Osio, M. M. G. (2013). Análise de modelos de regressão multiníveis simétricos. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos. doi:10.11606/D.55.2013.tde-05072013-161440. Recuperado em 2021-02-10, de www.teses.usp.br

Dalcin, A. K. (2016). Uma análise da relação entre violência na escola e proficiência dos alunos.

Questionários Saresp. Disponível em: <https://dados.educacao.sp.gov.br/dataset/question%C3%A1rios-saresp>. Acesso em: 06-06-2020.

Dicionário de dados do Questionário Saresp. Disponível em: <https://dados.educacao.sp.gov.br/dicionario-de-dados-do-questionario-saresp> Acesso em: 06-06-2020.

Proficiência do sistema de avaliação de rendimento escolar do estado de São Paulo (SARESP) por instância. Disponível em: <https://dados.educacao.sp.gov.br/dataset/profici%C3%A2ncia-do-sistema-de>

[avalia%C3%A7%C3%A3o-de-rendimento-escolar-do-estado-de-s%C3%A3o-paulo-saresp-6](#) Acesso em 06-06-2020.

SARESP - Sistema de avaliação. Disponível em: www.educacao.sp.gov.br/saresp . Acesso em 06/06/2020.

APÊNDICE A

Tabela 2. Impacto das variáveis sobre a proficiência em língua portuguesa

QUESTÃO	VARIÁVEL	Resposta	Número de Respondentes	%	Impacto sobre a proficiência de LP (PONTOS)
PAIS_Q1	Satisfação	2. Satisfeito com a escola	2967	61,77	-3.023.048
PAIS_Q10	Tarefa de casa	2. Sim, às vezes.	1430	29,77	-8.781.981
PAIS_Q10		3. Sim, quase nunca.	336	6,99	-6.924.369
PAIS_Q10		4. Não, nunca.	113	2,35	-10.006.792
PAIS_Q10		6. Não sei.	54	1,1	-15.335.080
PAIS_Q14	Diálogo	3. Não.	716	14,9	4.656.689
PAIS_Q16	Participação em reuniões	2. Sim, um pouco.	891	18,55	-3.937.036
PAIS_Q19	Escolaridade do pai	4. Completou o Ensino Médio, mas não completou o Ensino Superior.	1474	30,68	5.198.050
PAIS_Q19		6. Completou a Pós-Graduação	76	1,5	11.712.921
PAIS_Q23	Renda familiar	4. De R\$2.862,01 a	522	10,83	7.617.628

		R\$4.770,00.			
PAIS_Q23		5. De R\$4.770,01 a R\$7.632,00.	171	3,5	10.933.721
PAIS_Q23		8. Não sei/não quero responder.	1075	22,38	3.697.077
PAIS_Q 27	Livros em casa	1. Sim.	3372	70,2	-2.741.712
PAIS_Q34	Televisão	2. Um.	2167	45,11	-9.584.189
PAIS_Q34		3. Dois.	1760	36,64	-11.335.209
PAIS_Q34		4. Três.	623	12,97	-11.338.744
PAIS_Q34		5. Quatro ou mais.	167	3,47	-10.488.913
PAIS_Q36	Freezer	2. Um.	2770	57,67	3.359.244
PAIS_Q39	Celular/ Smartphone	2. Um.	930	19,36	6.374.566
PAIS_Q39		3. Dois.	1023	21,29	7.892.511
PAIS_Q39		4. Três.	1334	27,77	9.141.549
PAIS_Q39		5. Quatro ou mais.	1230	25,6	8.164.309
PAIS_Q40	Computador	2. Um.	2221	46,24	3.307.855
PAIS_Q40		3. Dois.	453	9,43	5.856.960
PAIS_Q42	Máquina de lavar	3. Dois.	262	5,45	-7.115.033
PAIS_Q43	Máquina de	2. Um.	1315	27,37	-3.039.771

PAIS_Q44	secar roupa	2. Um.	206	4,2	-7.222.570
PAIS_Q44		4. Três.	6	0,12	-41.469.824
ALUNO_01	Grau de satisfação	3. Insatisfeito.	293	6,1	7.930.042
ALUNO_02	Qualidade de ensino	4. Avaliação muito positiva.	147	3,06	8.757.509
ALUNO_08	Banheiro	2. Avaliação muito negativa.	875	18,21	4.603.546
ALUNO_08		3. Avaliação regular.	940	19,57	4.171.774
ALUNO_08		5. Avaliação extremamente positiva.	1220	25,4	4.569.656
ALUNO_09	Quando começou a estudar	3. Não lembro.	824	17,15	-5.275.797
ALUNO_10	Reprovação	2. Sim, 1 vez.	4336	90,27	-17.026.977
ALUNO_10		3. Sim, 2 vezes ou mais.	94	1,95	-22.697.517
ALUNO_11	Incentivo a leitura	2. Sim, um pouco.	266	5,53	2.625.225
ALUNO_12	Incentivo a estudar	2. Sim, um pouco.	608	12,65	-3.602.743
ALUNO_13	Ajudam ou conferem a lição de casa	2. Sim, um pouco.	712	14,82	5.745.829
ALUNO_13	Ajudam ou conferem a lição de casa	3. Não.	2148	44,72	7.585.742

ALUNO_14	Incentivo a ir à escola	2. Sim, um pouco.	439	9,14	-4.145.448
ALUNO_14		3. Não.	76	1,58	-10.736.731
ALUNO_15	Clima	3. Não.	1565	32,58	-4.643.735
ALUNO_16	Ler (jornal, revistas, livros)	2. Sim, um pouco.	1671	34,79	-4.504.719
ALUNO_16		3. Não.	1090	22,69	-8.988.261
ALUNO_18	Internet para estudos	2. Sim, um pouco.	2.765	57,56	6.261.153
ALUNO_18		3. Não.	1.362	28,35	11.619.051
ALUNO_21	Internet para assistir vídeos	2. Sim, às vezes.	275	5,72	-4.951.296
ALUNO_21		3. Nunca.	1050	21,86	-8.578.765
ALUNO_22	Gosta de estudar Mat	2. Não.	1335	27,79	-2.841.007
ALUNO_23	Faz a lição de casa de Mat	2. De vez em quando.	364	7,57	-7.582.633
ALUNO_23		3. Nunca.	2841	59,15	-8.807.253
ALUNO_23		4. Não tenho lição de casa.	107	2,22	-5.595.129
ALUNO_25	Faz a lição de casa de LP	2. De vez em quando.	1469	30,58	-7.744.140
ALUNO_25		4. Não tenho lição de casa.	2858	59,5	-5.499.100
ALUNO_26	Quantos livros leu	3. 2 ou 1.	574	11,95	-7.707.442
ALUNO_26		4. Nenhum.	1497	31,16	-11.365.831
ALUNO_27	Estuda: prestando	2. Gosto só um pouco.	2468	51,38	5.044.085

ALUNO_27	atenção	3. Gosto bastante.	310	6,45	13.146.070
ALUNO_28	Estudar: lendo anotações	3. Gosto bastante.	2169	45,15	-3.311.289
ALUNO_30	Estudar: pelos livros da escola	2. Gosto só um pouco.	2130	44,34	-3.985.512
ALUNO_30		3. Gosto bastante.	1073	22,34	-7.300.746
ALUNO_31	Estuda na internet, no computador ou aplicativos no celular	2. Gosto só um pouco.	2430	48,71	-5.931.328
ALUNO_33	Como se prepara para as provas	2. Eu estudo alguns dias antes ou no dia anterior às provas.	827	17,21	4.883.334
ALUNO_33		4. Eu não estudo para as provas.	1370	28,52	4.586.926
ALUNO_35	Clima: fui xingado	2. De vez em quando.	2201	45,82	3.950.132
ALUNO_36	Clima: fui ameaçado por outros alunos	2. De vez em quando.	724	15,07	6.846.666
ALUNO_36		3. Nunca.	241	5,01	9.092.101
ALUNO_39	Clima: tiraram ou destruíram coisas que me pertenciam	2. De vez em quando.	1066	22,19	5.330.602

ALUNO_42	Previsão de futuro	3. Eu vou parar de estudar para trabalhar.	208	4,33	-12.543.096
ALUNO_44	Relação entre os alunos	2. Regular.	2485	51,73	5.703.640
ALUNO_44		3. Boa.	266	5,53	5.418.515
ALUNO_45	Relação entre professores e direção	4. Não sei.	106	2,2	12.058.035
ALUNO_46	Relação entre os professores	3. Boa.	730	15,19	9.647.840
ALUNO_46		4. Não sei.	100	2,08	9.777.520
ALUNO_47	Relação entre alunos e direção	3. Boa.	304	6,32	-8.561.556
ALUNO_47		4. Não sei.	666	13,86	-8.005.218

APÊNDICE B

Tabela 3. Impacto das variáveis sobre a proficiência em matemática

QUESTÃO	VARIÁVEL	Resposta	Número de Respondentes	%	Impacto sobre a proficiência de LP (PONTOS)
PAIS_Q1	Satisfação	2. Satisfeito com a escola	2967	61,77	-3.458.848
PAIS_Q6	Espaço para esportes	2. Satisfeito com a escola	367	7,64	6.373.114
PAIS_Q10	Faz lição de casa	2. Sim, às vezes.	1430	29,77	-7.983.555
PAIS_Q10		3. Sim, quase nunca.	336	6,99	-9.414.092
PAIS_Q10		4. Não, nunca.	113	2,35	-15.727.681

PAIS_Q10		5. Não, a escola não passa lição de casa.	164	3,41	-10.052.149
PAIS_Q11	Utilizam o app Minha Escola SP no celular	2. Não.	4013	83,55	6.061.281
PAIS_Q19	Escolaridade do pai	6. Completou a Pós-Graduação (especialização, mestrado ou doutorado).	86	0,34	11.577.479
PAIS_Q22	Gosta de estudar matemática	2. Não.	1335	27,79	-13.240.634
PAIS_Q23	Renda familiar	2. De um a dois salários-mínimos (de R\$954,01 a R\$1.908,00).	1304	27,14	4.403.873
PAIS_Q23		3. De dois a três salários-mínimos (de R\$1.908,01 a R\$2.862,00).	865	18	6.016.783
PAIS_Q23		4. De três a cinco salários-mínimos (de R\$2.862,01 a R\$4.770,00).	522	10,86	8.492.797
PAIS_Q23		5. De cinco a oito salários-mínimos (de R\$4.770,01 a R\$7.632,00).	171	3,56	13.343.666
PAIS_Q 25	Revista de informação	1. Sim.	948	19,73	3.581.373

PAIS_Q26	Dicionário	1. Sim.	3435	71,51	-4.074.891
PAIS_Q36	Freezer	2. Um.	2770	57,67	3.429.412
PAIS_Q38	Telefone fixo	2. Um.	1592	33,14	2.655.185
PAIS_Q39	Telefone celular	2. Um.	930	19,36	8.001.142
PAIS_Q39		3. Dois.	1023	21,29	9.798.647
PAIS_Q39		4. Três.	1334	27,77	8.917.457
PAIS_Q39		5. Quatro ou mais.	1230	25,6	9.841.523
PAIS_Q42	Máquina de lavar	3. Dois.	262	5,45	-7.047.810
PAIS_Q44	Máquina de lavar louça	2. Um.	206	4,28	-7.233.929
PAIS_Q44		4. Três.	6	0,12	-36.161.779
PAIS_Q46	Aspirador de pó	2. Um.	2858	59,5	3.497.728
PAIS_Q48	Motocicleta	2. Um.	1030	21,44	-2.912.126
PAIS_Q50	Empregada mensalista	4. Três.	18	0,37	-21.103.313
ALUNO_01	Satisfação com a escola	3. Insatisfeito.	481	10,01	5.384.598
ALUNO_02	Qualidade do ensino	3.Avaliação regular.	101	2,1	12.351.610
ALUNO_02		4. Avaliação muito positiva.	147	3,06	12.546.124
ALUNO_02		5.Avaliação extremamente positiva.	675	14,05	9.213.046
ALUNO_08	Banheiro	2.Avaliação muito	876	18,23	3.984.390

		negativa.			
ALUNO_08		4.Avaliação muito positiva.	744	15,49	4.175.635
ALUNO_08		5. Avaliação extremamente positiva.	1220	25,4	4.992.281
ALUNO_09	Começou a estudar	3. Não lembro.	824	17,15	-4.826.004
ALUNO_10	Reprovação	2. Sim, 1 vez.	4336	90,27	-16.143.003
ALUNO_10		3. Sim, 2 vezes ou mais.	94	1,95	-20.416.138
ALUNO_13	Ajudam na lição de casa	2. Sim, um pouco.	712	14,82	6.654.886
ALUNO_13		3. Não.	2148	44,72	9.470.331
ALUNO_14	Incentivam a ir à escola	2. Sim, um pouco.	439	9,14	-5.385.843
ALUNO_16	Ler (jornal, revistas, livros)	3. Não.	1090	22,69	-4.579.711
ALUNO_18	Internet para os estudos.	2. Sim, um pouco.	2765	57,56	4.894.749
ALUNO_18		3. Não.	1362	28,35	6.553.667
ALUNO_21	Internet para assistir vídeos	2. Sim, às vezes.	275	5,72	-6.258.320
ALUNO_21		3. Nunca.	1050	21,86	-7.356.857
ALUNO_22	Gosta de estudar matemática	2. Não.	1335	27,79	-13.240.634
ALUNO_23	Faz a lição de casa de	2. De vez em quando.	364	7,57	-9.934.504
ALUNO_23	Matemática	3. Nunca.	2841	59,15	-13.293.311

ALUNO_23		4. Não tenho lição de casa.	107	2,22	-7.587.583
ALUNO_24	Gosta de estudar LP	2. Não.	901	18,75	7.335.842
ALUNO_25	Faz a lição de casa de LP	2. De vez em quando.	1469	30,58	-3.341.623
ALUNO_26	Livros que leu este ano	2. 4 ou 3.	1205	25,08	-2.914.295
ALUNO_26		3. 2 ou 1.	574	11,95	-6.628.815
ALUNO_26		4. Nenhum.	1497	31,16	-9.932.163
ALUNO_27	Estuda: prestando atenção	3. Gosto bastante.	310	6,45	8.072.681
ALUNO_28	Estuda: lendo anotações	2. Gosto só um pouco.	1208	25,15	-2.900.188
ALUNO_30	Estuda: pelos livros	2. Gosto só um pouco.	2130	44,34	-4.844.264
ALUNO_30		3. Gosto bastante.	1073	22,34	-5.824.173
ALUNO_31	Estuda: na internet, por programas no computador ou aplicativos no celular	2. Gosto só um pouco.	2430	50,59	-5.629.265
ALUNO_32	Tempo fazendo lição de casa	2. Entre 30 minutos e 1 hora.	425	8,84	2.507.524
ALUNO_33	Como se prepara para as provas	2. Eu estudo alguns dias antes ou no dia anterior às provas.	827	17,21	5.164.845
ALUNO_33		4. Eu não	571	11,88	10.347.784

		estudo para as provas.			
ALUNO_35	Sobre experiências na escola: fui chamado de	2. De vez em quando.	2201	45,82	6.454.041
ALUNO_35	nomes feios por outros alunos.	3. Nunca.	1941	40,41	5.039.951
ALUNO_42	Previsão do futuro	3. Eu vou parar de estudar para trabalhar.	208	4,33	-9.317.138
ALUNO_44	Relação entre os alunos	2. Regular.	2485	51,73	6.767.128
ALUNO_44		3. Boa.	266	5,53	7.753.605
ALUNO_45	Relação entre professores e direção	3. Boa.	1025	21,34	9.445.677
ALUNO_45		4. Não sei.	106	2,20	11.602.337
ALUNO_47	Relação entre alunos e direção	3. Boa.	2353	48,99	-6.687.594
ALUNO_47		4. Não sei.	666	13,86	-6.749.234
ALUNO_48	Relação entre alunos e funcionários.	4. Não sei.	522	10,86	-8.213.450