EFEITOS DAS AVALIAÇÕES SERIADAS SOBRE O DESEMPENHO NO SARESP: UMA ANÁLISE COM REGRESSÃO LINEAR

Driely Turi URSINI
Camila Fernanda BASSETTO
Universidade Estadual Paulista (UNESP)
Educação, desenvolvimento e aprendizagem
driely.ursini@unesp.br

Palavras-chave: Avaliação Educacional. Aprendizagem. Regressão linear.

Resumo

O presente estudo tem por objetivo estimar o impacto das avaliações seriadas, ocorridas durante o ano de 2025, sobre o desempenho na avaliação de larga escala aplicada no final do período letivo. Para tanto, foram considerados dados referentes às notas alcançadas em Matemática nas aplicações bimestrais da Prova Paulista e os resultados do Sistema de Avaliação e Rendimento Escolar do Estado de São Paulo (SARESP) por uma turma de estudantes matriculados no 8º ano do Ensino Fundamental em uma escola de ensino público. Deve-se considerar a dependência existente entre as observações, haja vista a medição sobre o mesmo indivíduo nos diferentes bimestres, e a multicolinearidade resultante entre as variáveis explicativas, devido ao fato de a mesma informação ser considerada mais de uma vez. Neste cenário, para medir o efeito da Prova Paulista sobre o SARESP, utilizou-se uma análise de regressão linear em que o desempenho em Matemática teve como covariável a média das notas obtidas nas avaliações bimestrais. Os resultados indicam crescimento do desempenho médio ao longo das avaliações bimestrais, acompanhado de redução relativa da variabilidade entre os estudantes, sugerindo maior homogeneidade da aprendizagem ao longo do período analisado. Observou-se, ainda, uma aproximação entre os resultados das avaliações bimestrais e o desempenho registrado no SARESP. O estudo evidencia o potencial das análises quantitativas no acompanhamento da aprendizagem e na interpretação de dados provenientes de avaliações educacionais em larga escala.

Palavras-chave: Avaliação Educacional. Aprendizagem. Regressão linear.

1. Introdução

O presente trabalho analisa o acompanhamento longitudinal da aprendizagem de estudantes do ensino fundamental a partir de avaliações bimestrais e de uma avaliação externa em larga escala. O estudo parte da compreensão de que avaliações aplicadas em diferentes momentos do ano letivo podem fornecer indicadores relevantes

sobre o desempenho escolar e contribuir para a interpretação dos processos de aprendizagem dos estudantes.

O objetivo da pesquisa consiste em analisar o desempenho longitudinal de estudantes do 8º ano do ensino fundamental a partir de três aplicações da Prova Paulista e sua relação com os resultados do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP). Busca-se identificar padrões de evolução do desempenho, variabilidade da aprendizagem e possíveis aproximações entre os resultados das avaliações realizadas ao longo do ano letivo.

Para isso, adotou-se uma abordagem quantitativa exploratória, fundamentada em procedimentos de estatística descritiva. Foram analisados dados de 38 estudantes da rede pública estadual paulista, considerando medidas de tendência central e dispersão, como média, mediana, desvio padrão e coeficiente de variação. A utilização desses indicadores possibilitou observar tendências gerais de desempenho e acompanhar mudanças na variabilidade dos resultados ao longo das avaliações.

A realização deste estudo justifica-se pela relevância das avaliações educacionais no acompanhamento da aprendizagem e na produção de informações para subsidiar reflexões sobre o desempenho escolar. Além disso, a análise quantitativa de dados provenientes de avaliações seriadas e externas permitem compreender os processos de ensino e aprendizagem no contexto da educação básica.

O trabalho organiza-se em seções, incluindo esta introdução. A segunda seção contém breves discussões sobre avaliações em larga escala, usos dos resultados educacionais e acompanhamento do desempenho escolar. Na seção três são descritos os dados analisados e os procedimentos estatísticos utilizados na pesquisa. A quarta seção contém interpretações pedagógicas dos indicadores obtidos, enquanto a quinta seção sintetiza os principais achados do estudo. Por fim, a seção seis apresenta as considerações finais e possíveis encaminhamentos para pesquisas futuras.

2. Fundamentação teórica

A análise quantitativa do desempenho escolar tem assumido papel central no acompanhamento da aprendizagem e na avaliação de políticas educacionais, especialmente no contexto das avaliações em larga escala. No Brasil, as avaliações externas justificaram-se, inicialmente, pela necessidade de se obter informações para subsidiar a tomada de decisões no contexto educacional, assumindo uma função diagnóstica, sem que houvesse consequências diretas para as escolas, para os seus atores e sobre as práticas pedagógicas nelas realizadas. Porém, a partir da primeira década do século XXI, com as propostas e as mudanças técnicas introduzidas nas

principais iniciativas de avaliar, observam-se modificações também nos objetivos e funções subjacentes às avaliações (Bauer, 2024).

Instrumentos avaliativos aplicados ao longo do ano letivo permitem observar o desempenho dos estudantes em diferentes momentos do tempo, enquanto exames externos buscam aferir, de forma padronizada, os resultados educacionais alcançados. Segundo Weiss (1998), o conceito de uso das avaliações refere-se à maneira como os resultados são considerados no processo de tomada de decisão. A autora destaca diferentes classificações para o uso desses resultados.

> O primeiro deles é o instrumental, que pode ser compreendido como um uso voltado para a tomada de decisão. Nesse sentido, a avaliação contribui para a produção de descobertas que influenciam a ação investigada e indicam quais decisões poderão ser tomadas a partir dessas evidências. O segundo é denominado uso conceitual, no qual os resultados oriundos da avaliação podem modificar a concepção sobre o programa, sua finalidade e sua própria existência. O terceiro corresponde ao uso como instrumento de persuasão. Nesse caso, a avaliação é utilizada para legitimar determinada posição ou para conquistar novos defensores de uma política ou programa. Por fim, o quarto tipo é o uso para o esclarecimento, no qual a avaliação influencia instituições e agentes que não estão diretamente relacionados ao programa ou à política analisada (Basso et al., 2022, p. 504).

Como aponta o estudo de Basso et al. (2022), os dados das avaliações de larga escala são utilizados direta e indiretamente na formulação de políticas educacionais no Brasil, uma vez que os resultados atuam tanto como insumos indispensáveis para o desenvolvimento de programas educacionais com objetivos específicos relacionados ao aumento da proficiência dos alunos, quanto como diagnósticos da educação em uma rede de ensino específica. Nesse contexto, avaliações educacionais em larga escala têm sido amplamente utilizadas para produzir indicadores capazes de acompanhar o desempenho dos sistemas educacionais e subsidiar a formulação de políticas públicas voltadas à melhoria da qualidade da educação (Soares; Xavier, 2013).

Na rede estadual de São Paulo, em 1996 foi implementado o SARESP, com uma proposta de produzir um diagnóstico detalhado sobre a Educação Básica ofertada no Estado, a fim de orientar os gestores educacionais no monitoramento e na implementação de políticas públicas voltadas à melhoria da qualidade educacional. Nesta avaliação de larga escala são avaliados, de forma censitária ou amostral, os conhecimentos, habilidades e competências dos alunos matriculados no 3º, 5º, 7º e 9º anos do Ensino Fundamental e daqueles que estão cursando a 3ª série do Ensino Médio, por meio de provas compostas por questões sobre Língua Portuguesa, Matemática, Ciências Humanas, Ciências da Natureza e redação. Os resultados obtidos a partir desta avaliação são utilizados para guiar estratégias e ações Secretaria Estadual

de Educação e para compor o cálculo do Índice de Desenvolvimento da Educação do Estado de São Paulo (IDESP). De acordo com a escala proposta pelo SARESP, os alunos são classificados em quatro níveis de proficiência: Abaixo do Básico (AB), Básico (BA), Adequado (AD) e Avançado (AV). Para essa categorização, são utilizados os critérios para as expectativas de aprendizagem de cada ano/série do currículo do Estado de São Paulo e, dentro das perspectivas citadas, estão os conteúdos, habilidades e competências. (SÃO PAULO, 2009).

Em 2011, foi anunciada, pela Secretaria de Educação (Seduc-SP), a Avaliação da Aprendizagem em Processo (AAP), substituída em 2023 pela Prova Paulista. A Prova Paulista é uma avaliação destinada aos estudantes do 5º ano do Ensino Fundamental até a 3ª série do Ensino Médio, pautada nos conteúdos abordados presentes nos materiais disponibilizados aos docentes. A aplicação ocorre em todas as escolas da rede estadual e tem a finalidade de acompanhar o desenvolvimento do currículo, em todas as áreas do conhecimento, além de apoiar os professores na identificação dos conteúdos que os estudantes aprenderam ao longo do bimestre (Desiderio; Sousa, 2025).

Segundo Desiderio e Sousa (2025), a Prova Paulista faz parte de um

> [...] movimento mais amplo de implementação de iniciativas de avaliação de sistemas e redes de ensino pelos entes federados, as quais têm sido utilizadas como instrumentos tanto para induzir a melhoria da qualidade da educação quanto para a gestão das políticas educacionais. (Desiderio e Sousa, 2025, p. 2).

Compreender como os resultados das Provas Paulistas se relacionam com o desempenho no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) constitui um desafio relevante na área educacional e demanda o uso de métodos quantitativos adequados ao tipo de dado disponível. No sistema educacional paulista, a Prova Paulista consiste em avaliações bimestrais aplicadas ao longo do ano letivo, enquanto o SARESP configura-se como uma avaliação externa de caráter somativo, aplicada ao final do período letivo. Essas avaliações geram conjuntos de dados longitudinais, nos quais os mesmos estudantes são observados repetidamente ao longo do tempo. A análise desses dados possibilita identificar tendências de desempenho e acompanhar mudanças na aprendizagem dos estudantes ao longo do período letivo.

A análise quantitativa de dados educacionais possibilita compreender padrões de desempenho, variabilidade e evolução da aprendizagem dos estudantes ao longo do ano letivo. A utilização de medidas estatísticas, como média, mediana e indicadores de

dispersão, permite analisar o comportamento da turma, as diferenças de desempenho entre os estudantes e as relações entre as avaliações realizadas ao longo do período.

3. Desenvolvimento

Os dados utilizados neste estudo são provenientes de avaliações educacionais aplicadas a estudantes do 8º ano do Ensino Fundamental da rede pública estadual paulista. O conjunto de dados é composto por resultados individuais de 38 estudantes, obtidos em avaliações da Prova Paulista, correspondentes ao 1º, 2º e 3º bimestres do ano letivo, bem como pela nota obtida no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), aplicado ao final do período letivo. Cada estudante possui, portanto, quatro registros de desempenho: três referentes às avaliações bimestrais e um referente à avaliação externa final. Adicionalmente, foram utilizados registros internos de forma agregada, sem identificação dos estudantes ou da unidade escolar, em conformidade com princípios éticos relacionados à pesquisa educacional.

As avaliações consideradas utilizam escala de pontuação de 0 a 10. A partir desses dados, foram calculadas medidas de tendência central e dispersão com o objetivo de sintetizar o comportamento estatístico do desempenho dos estudantes ao longo das avaliações. Foram analisadas medidas como média, mediana, desvio padrão e coeficiente de variação, possibilitando observar o desempenho médio da turma e a variabilidade entre os estudantes.

O uso de procedimentos de estatística descritiva justifica-se pelo caráter exploratório da pesquisa e pela natureza do conjunto de dados analisado. Esses indicadores permitem identificar tendências gerais de desempenho e acompanhar possíveis mudanças na aprendizagem dos estudantes ao longo do período observado.

Na Tabela 1, apresentam-se as estatísticas descritivas para as notas.

Tabela 1. Estatísticas descritivas das notas bimestrais e no SARESP.

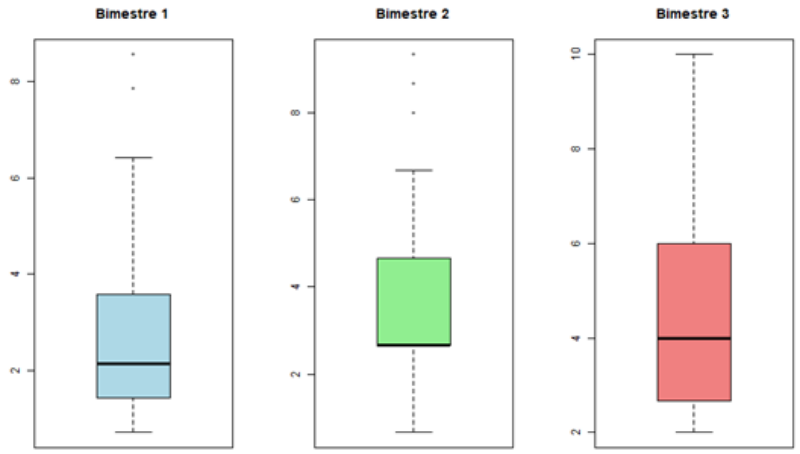
Estatística descritiva	Bimestres			SARESP
	PP1	PP2	PP3	
Média	2,89	3,65	4,72	4,11
Mediana	2,14	2,67	4,0	3,5
Nota mínima	0,71	2,67	2,67	1,5
Nota máxima	8,57	9,33	10,0	9,0
1º quartil	1,43	2,67	2,67	3,0
3º quartil	3,57	4,50	5,83	5,0
Desvio padrão	1,89	2,06	2,29	1,85

Fonte: Elaboração das autoras a partir dos dados coletados.

Os valores mostrados na Tabela 1 sugerem que os estudantes alcançaram melhores resultados na avaliação aplicada no 3º bimestre, quando comparados aos outros dois. Enquanto no primeiro bimestre, a nota mínima foi inferior a 1, no terceiro bimestre a nota mais baixa registrada foi quase quatro vezes maior, igual a 2,67. Este último bimestre também foi o único a registrar a nota máxima igual a 10. Esses indicadores contribuem para compreender não apenas a evolução das médias, mas também possíveis mudanças na heterogeneidade da aprendizagem dos estudantes ao longo do período analisado.

As estatísticas descritivas são ilustradas graficamente na Figura 1, a qual corresponde aos boxplots das notas alcançadas em cada um dos bimestres.

Figura 1. Boxplots dos resultados bimestrais.



Fonte: Elaboração das autoras a partir dos dados coletados.

No boxplot referente às notas do primeiro bimestre, vale destacar que somente dois estudantes alcançaram pontuações próximas de 8, pertencendo a maioria no intervalo de notas entre 1,43 e 3,57 as quais são consideradas ruins. No segundo bimestre, o desempenho da turma foi um tanto melhor, pois além de três alunos terem alcançado pontuações iguais ou superiores a 8, o boxplot ilustra ainda que a maioria dos alunos obteve notas entre 2,67 e 4,50. No terceiro bimestre, como ilustra o boxplot, 50% dos estudantes tiveram notas entre 4,0 e 6,0.

Para explicar o desempenho no SARESP e medir o efeitos das avaliações bimestrais, propõe-se o modelo dado na expressão (1):

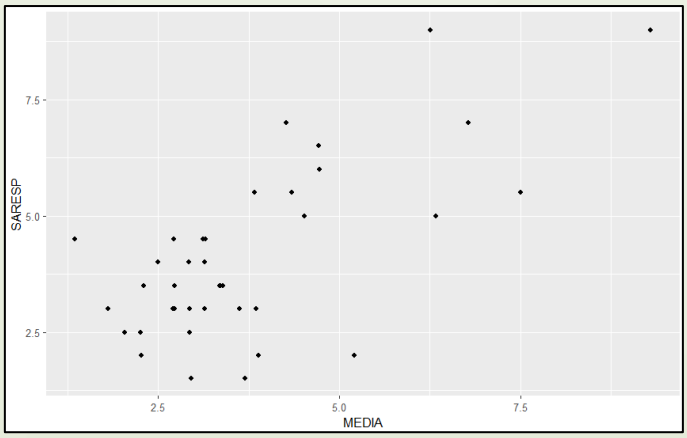
$$SARESP_i = \beta_0 + \beta_1 NB_i + \varepsilon_i \quad (1)$$

onde *i* é o estudante, *i* = 1,...,38; β_0 é o intercepto e representa o desempenho no SARESP quando a média das notas alcançadas nas avaliações bimestrais é zero, β_1 é o parâmetro correspondente à variável explicativa NB, dada pela média das notas

bimestrais do primeiro, segundo e terceiro bimestres, e ε é o termo de erro, o qual representa os demais fatores que podem impactar o desempenho no SARESP e que não estão considerados neste modelo.

Antes de estimar os parâmetros do modelo proposto, investiga-se a relação existente entre as variáveis dependente e independente. Tal investigação pode ocorrer por meio do gráfico de dispersão (Field, 2012), como mostra a Figura 2.

Figura 2. Gráfico de dispersão entre as variáveis SARESP e Média bimestral.



Fonte: Elaboração das autoras.

Por esta Figura 2, é possível inferir que existe uma relação aproximadamente linear (positiva) entre as variáveis.

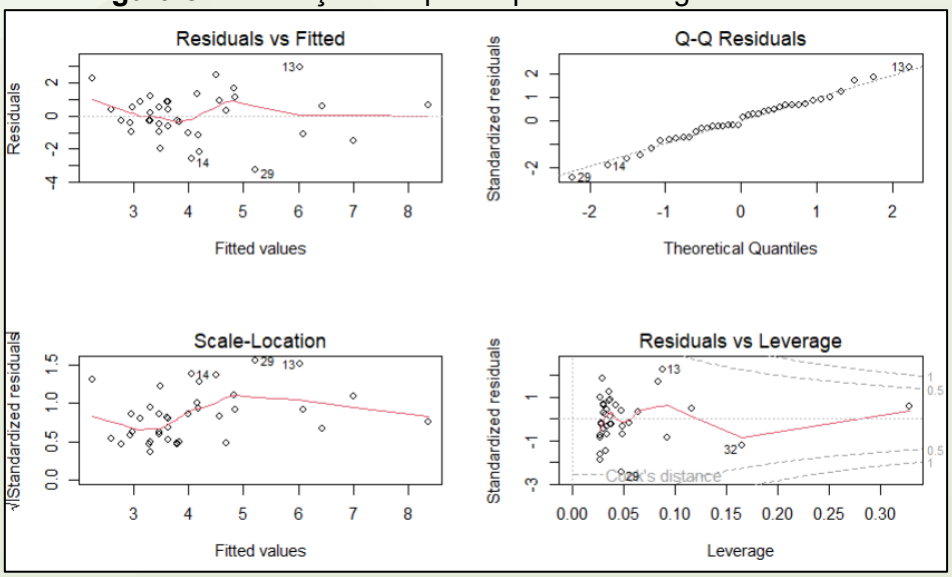
A etapa seguinte consiste em verificar os pressupostos de existência de valores discrepantes (outliers), homocedasticidade e distribuição normal.

Por valores discrepantes, caracterizam-se aqueles que se diferenciam drasticamente dos demais, isto é, são valores que fogem da normalidade e que causam anomalias nos resultados obtidos. Estes podem viesar as conclusões. A homocedasticidade faz-se presente nas situações em que o termo de erro é constante, ou seja, exibe um comportamento próximo de uma, similar em todos os valores da variável independente. Ao contrário, a heterocedasticidade apresenta-se como uma forte dispersão dos dados em torno de uma reta.

Na Figura 3, estão os gráficos associados aos pressupostos que devem ser atendidos pela regressão linear. O gráfico de Resíduos versus Valores Previstos (Residual vs Fitted) permite fazer uma análise, tanto da linearidade como da homocedasticidade. A linha vermelha indica uma relação linear quando está na posição horizontal, de tal forma que, quanto mais próxima da linha cinza, mais perfeitamente linear. Para os dados considerados neste estudo, a linha vermelha apresenta uma leve curvatura, mas no geral, não mostra um padrão parabólico ou senoidal forte. Apesar

desta leve distorção, a premissa de linearidade está atendida. No gráfico Q-Q Resíduos (Q-Q Residuals), estão apresentados no eixo y os resíduos padronizados e no eixo x os resíduos teóricos (rodapé), isto é, aqueles resíduos esperados em uma distribuição Normal. Para que os resíduos apresentem distribuição Normal, devem estar próximos da linha pontilhada.

Figura 3. Verificação dos pressupostos da regressão linear.



Fonte: Elaboração das autoras.

No gráfico de homocedasticidade (Scale Location), os pontos devem estar espalhados com uma largura constante e a linha vermelha deve ser horizontal, ou parcialmente horizontal. Alterações entre subidas e descidas desta linha ou um formato de funil, abrindo e fechando, nos pontos, são indicativos de heterocedasticidade. Para os dados em análise, observa-se que a linha vermelha sobe um pouco e, logo em seguida, decresce levemente e que a dispersão dos pontos varia um pouco, mas não há um formato de funil, indicando, portanto, a homocedasticidade. O gráfico de Resíduos versus Outliers (Residual vs Leverage) identifica a existência de resíduos que se enquadram como pontos de alavancagem, ou seja, se alguma observação está discrepante ao ponto de influenciar na estimação do modelo. Caso existam, os resíduos devem estar além da faixa -3 e 3, o que não ocorre neste caso, como mostram as estatísticas descritivas dos resíduos:

Tabela 2. Estatísticas descritivas dos resíduos.

Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
-2,4181	-0,6617	-0,0392	0,0029	0,6432	2,2825

Fonte: Elaboração das autoras.

Observando os valores Mínimo e Máximo, percebe-se que os resíduos não estão fora do intervalo -3 e 3. Sendo assim, não há outliers.

Ainda como parte dos pressupostos da regressão linear, foi realizado teste de Shapiro-Wilk, cujo objetivo é avaliar se os dados seguem uma distribuição Normal. O teste é estruturado da seguinte forma:

$$H_0: os\ dados\ seguem\ uma\ distribuição\ Normal$$
$$H_a: os\ dados\ não\ seguem\ uma\ distribuição\ Normal$$

A decisão do teste fundamenta-se no valor do p-valor. Quando $p > 0{,}05$, não rejeita-se a hipótese nula e, portanto, admite-se que os dados seguem uma distribuição Normal. Para os dados em análise no presente estudo, o resultado apontou $p = 0{,}9332$, permitindo, então, considerar a hipótese nula.

Para estudar a relação entre duas variáveis (autocorrelação), cujos parâmetros de interesse são estimados usando o método de Mínimos Quadrados Ordinários (MQO), emprega-se o teste de Durbin-Watson. As hipóteses estabelecidas são dadas por:

$$H_0: as\ variáves\ são\ independentes$$
$$H_a: as\ variáveis\ não\ são\ independentes$$

De acordo com este teste[1], para que resíduos sejam independentes, a estatística correspondente deve ser um valor entre 1 e 3. Para os dados em análise, tem que DW = 1,7888, com p-valor = 0,554.

Por fim, verifica-se a homocedasticidade dos dados, o que implica em testar se os erros tem variância comum, com o teste de Breusch-Pagan, com as hipóteses:

$$H_0: os\ dados\ são\ homocedásticos$$
$$H_a: os\ dados\ não\ \ homocedásticos$$

Como aponta o resultado obtido no RStudio, o valor da estatística BP = 2,103, com p-valor igual a 0,147 e, portanto maior que 0,05. Neste caso, a hipótese nula não é rejeitada e, então, conclui-se que existe homocedasticidade.

Na seção seguinte, são apresentadas e interpretadas as estimativas obtidas, via software RStudio, para os parâmetros do modelo dados na expressão (1).

4. Resultados e discussão

Os resultados apresentados na Tabela 1 indicam crescimento progressivo do desempenho médio dos estudantes ao longo das aplicações da Prova Paulista. A média da turma passou de 2,9 na PP1 para 3,7 na PP2, alcançando 4,7 na PP3. Esse

[1] Uma autocorrelação positiva é identificada por um agrupamento de resíduos com o mesmo sinal. Uma autocorrelação negativa é identificada por rápidas mudanças nos sinais de resíduos consecutivos.

comportamento sugere uma evolução gradual da aprendizagem ao longo do ano letivo, evidenciando melhora no desempenho dos estudantes entre as avaliações bimestrais.

As medianas também apresentaram crescimento ao longo do período analisado, passando de 2,1 na PP1 para 4,0 na PP3. Esse resultado indica que o avanço observado não ocorreu apenas em casos isolados, mas esteve presente em uma parcela significativa da turma, sugerindo melhoria mais ampla do desempenho coletivo.

Em relação à dispersão dos resultados, observa-se aumento do desvio padrão ao longo das avaliações bimestrais. Entretanto, o coeficiente de variação apresentou redução progressiva, passando de 0,65 na PP1 para 0,49 na PP3. Esse comportamento indica que, embora a dispersão absoluta tenha aumentado, a variabilidade relativa entre os estudantes diminuiu quando comparada ao crescimento das médias.

Do ponto de vista educacional, a redução do coeficiente de variação pode sugerir maior homogeneidade da aprendizagem ao longo do período analisado. Esse resultado pode estar relacionado a diferentes fatores pedagógicos, como adaptação dos estudantes às avaliações, continuidade do trabalho docente, acompanhamento da aprendizagem e consolidação gradual dos conteúdos desenvolvidos ao longo do ano letivo. Observa-se, ainda, que a média obtida na terceira aplicação da Prova Paulista (4,7) aproxima-se do desempenho registrado no SARESP (4,1), indicando possível relação entre os resultados das avaliações realizadas ao longo do ano e o desempenho apresentado na avaliação externa final. Esse resultado reforça o potencial das avaliações seriadas como instrumentos de acompanhamento da aprendizagem e monitoramento do desempenho escolar.

A Tabela 3 contém as estimativas para os parâmetros do modelo dado na expressão (1), o qual busca medir a associação entre o desempenho no SARESP e nas avaliações bimestrais realizadas via Prova Paulista.

Tabela 3. Estimativas para os parâmetros do modelo.

Coeficiente	Estimativa	Erro padrão	Valor t	p-valor
Intercepto (β_0)	1,2225	0,5543	2,205	0,0339*
Média (β_1)	0,7678	0,1353	5,673	1,9e-0,6***

Fonte: Estimativas obtidas via software RStudio, a partir dos dados coletados.

No contexto educacional, tem-se que, para os alunos que alcançados média zero nas provas bimestrais, a nota no SARESP será, aproximadamente, igual a 1,22, como indica a estimativa para o intercepto do modelo, representado por β_0. A estimativa para β_1, de quase 0,77, indica que, para cada aumento de um ponto na média das avaliações bimestrais, a nota estimada no SARESP aumenta, em média 0,77 ponto.

O erro padrão mede a variabilidade associada à estimativa do parâmetro β_1. Quanto menor, mais precisa é a inferência das médias bimestrais sobre o desempenho no SARESP. O valor t é o resultado da razão entre a estimativa do parâmetro e o erro padrão correspondente. Resultados maiores que dois indicam rejeição da H_0, isto é, de que os parâmetros sejam iguais a zero. Por fim, os valores relacionados ao p-valor, inferiores aos níveis tradicionais de significância, representados por 1%, 5% e 10%, sugerem que o impacto das médias bimestrais sobre o desempenho no SARESP é significativo do ponto de vista estatístico.

Os resultados obtidos ao longo das avaliações evidenciam crescimento gradual do desempenho médio dos estudantes nas aplicações da Prova Paulista, acompanhado de redução relativa da variabilidade entre os resultados da turma. A elevação das médias e medianas ao longo do período analisado sugere avanço coletivo da aprendizagem, enquanto a diminuição do coeficiente de variação indica maior homogeneidade do desempenho entre os estudantes. Observou-se, ainda, aproximação entre os resultados registrados na terceira aplicação da Prova Paulista e o desempenho obtido no SARESP, o que sugere possível relação entre as avaliações realizadas ao longo do ano letivo e a avaliação externa final. Esses resultados reforçam o potencial das avaliações seriadas como instrumentos de acompanhamento da aprendizagem e monitoramento do desempenho escolar. Aditivamente, a modelagem estatística, por meio da regressão linear, comprovou que a média das avaliações bimestrais atua como um preditor linear positivo e confiável para as notas do SARESP, dado o nível de significância envolta das estimativas obtidas para os parâmetros do modelo considerado.

5. Conclusões

O presente estudo buscou analisar o desempenho longitudinal de estudantes do ensino fundamental a partir de avaliações bimestrais e de uma avaliação externa em larga escala. A utilização de indicadores estatísticos e regressão linear permitiu identificar tendências de evolução do desempenho e acompanhar mudanças na variabilidade dos resultados ao longo do ano letivo.

Os resultados empíricos obtidos sugerem que as avaliações seriadas podem contribuir para o acompanhamento da aprendizagem dos estudantes, permitindo observar padrões de desempenho ao longo do tempo e oferecendo subsídios para a interpretação de resultados educacionais. Além disso, a aproximação entre os resultados das avaliações bimestrais e do SARESP evidencia o potencial dessas avaliações como instrumentos de monitoramento do desempenho escolar.

Do ponto de vista educacional, o estudo reforça a importância das análises quantitativas como ferramentas de apoio à compreensão dos processos de ensino e aprendizagem, especialmente no contexto das avaliações em larga escala. Embora o trabalho apresente caráter exploratório e esteja limitado a um conjunto específico de estudantes, os resultados indicam possibilidades de utilização de indicadores estatísticos no acompanhamento da aprendizagem e na análise de dados educacionais.

Por fim, sugere-se que pesquisas futuras ampliem o conjunto de dados analisados e incorporem outras variáveis educacionais, de modo a aprofundar a compreensão sobre os fatores associados ao desempenho escolar ao longo do tempo.

Principais referências

BASSO, F. V.; FERREIRA, R.R.; OLIVEIRA, A. S. Uso das avaliações de larga escala na formulação de políticas públicas educacionais. **Ensaio: Aval. Pol. Públ. Educ**., v. 30, n. 115, p. 501–519, abr./jun. 2022. DOI: 10.1590/S0104-40362021002902436. Disponível em: https://www.scielo.br/j/ensaio/a/h4wpxfkMkZvSZhXSB8gt7XG/?lang=pt. Acesso em: 08 mar. 2026.

BAUER, A. Large scale assessment and establishment of educational goals and indicators: uses and abuses of quantophrenia. **Educar em Revista**, v. 40, e94095, 2024. DOI: 10.1590/1984-0411.94095. Disponível em: http://educa.fcc.org.br/scielo.php?pid=S0104-40602024000102002&script=sci_arttext. Acesso em: 10 mar. 2026.

DESIDERIO, G. L.; SOUSA, S. M. Z. L. **Avaliação Externa Bimestral: Análise Da Trajetória E Propósitos Da Aap E Da Prova Paulista**. **SciELO Preprints**, 2025. DOI: 10.1590/SciELOPreprints.12539. Disponível em: https://preprints.scielo.org/index.php/scielo/preprint/view/12539. Acesso em: 9 maio. 2026.

FIELD, A.; FIELD, Z., MILES, J. Discovering statistics using R. 2012.

SÃO PAULO (Estado). Secretaria da Educação. **Matrizes de referência para a avaliação: documento básico – SARESP**. São Paulo: SEE, 2009. Disponível em: https://file.fde.sp.gov.br/saresp/saresp2013/Arquivos/Saresp2013_MatrizRefAvaliacao_DocBasico_Completo.pdf. Acesso em: 9 maio 2026.

SOARES, J. F.; XAVIER, F. P. Pressupostos educacionais e estatísticos do Ideb. **Educação & Sociedade**, v. 34, n. 124, p. 903–923, jul. 2013. DOI: 10.1590/S0101-73302013000300013. Disponível em: https://www.scielo.br/j/es/a/JLzr4qdx89rjrNXnydNcvcy/?lang=pt. Acesso em: 12 mar. 2026.

WEISS, C. Have we learned anything new about the use of evaluation**? American Journal of Evaluation**, v. 19, n. 1, p. 21–33, 1998. DOI: 10.1177/109821409801900103. Disponível em: https://journals.sagepub.com/doi/10.1177/109821409801900103. Acesso em: 12 mar. 2026.