



Efeitos rendimento escolar, infraestrutura e prática docente na qualidade do ensino médio no Brasil

Ricardo Alexandrino Garcia*
Eduardo Luiz Gonçalves Rios-Neto**
Adriana de Miranda-Ribeiro***

O artigo identifica e discute fatores escolares que mais influenciam a qualidade do ensino médio no Brasil: rendimento escolar, infraestrutura ou prática docente. Foram utilizados os dados do Censo Escolar e do Exame Nacional do Ensino Médio. A técnica de mineração de dados levantou características-chave para o bom desempenho escolar. Um modelo de regressão logística identificou os fatores escolares que mais influenciam o desempenho escolar dos alunos da rede pública de ensino médio regular no Brasil. Investigaram-se, também, os diferenciais de desempenho escolar de alunos do ensino médio segundo o tipo de escola e tipo de ensino, bem como os diferenciais interestaduais, considerando o sistema público de ensino médio regular. Os resultados apontam que alunos de escolas estaduais estão em desvantagem e que os de escolas privadas e federais possuem desempenho semelhante, quando considerado apenas o tipo de escola. Ao se levar em conta também o tipo de ensino, os alunos do ensino regular das escolas federais apresentam o melhor desempenho, cerca de 1,3 vez maior do que o dos alunos das escolas estaduais. Não há grandes disparidades interestaduais, embora os resultados apontem diferenciação regional. Dentre os fatores escolares, a qualificação docente se mostrou o fator mais impactante no desempenho escolar.

Palavras-chave: Qualidade do ensino. Desempenho escolar. Ensino médio. Mineração de dados. Enem. Brasil.

* Instituto de Geociências, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brasil (alexandrinogarcia@gmail.com; <https://orcid.org/0000-0001-7144-9866>).

** Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Faculdade de Ciências Econômicas (Face), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brasil e Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro-RJ, Brasil (riosnetoedu@gmail.com; <https://orcid.org/0000-0002-3161-5791>).

*** Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Faculdade de Ciências Econômicas (Face), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brasil (adrianamiribeiro@cedeplar.ufmg.br; <https://orcid.org/0000-0002-6845-7926>).

Introdução

O debate acerca do sistema de ensino público no país tem se direcionado, nos últimos tempos, para questões sobre a qualidade dos serviços educacionais oferecidos pelos estabelecimentos escolares, principalmente dos ensinos médio e fundamental. Em relação ao ensino médio, em virtude do elevado índice de abandono e das implicações para o contexto geral da sociedade, a qualidade dos serviços educacionais assume proporções ainda mais relevantes. Dada a complexidade inerente aos processos de ensino e de aprendizagem, avaliar a qualidade dos serviços educacionais em qualquer nível não é tarefa fácil. Há, no mínimo, três dimensões que devem ser consideradas em sua abordagem: a do aluno, a da família e a da escola (SOARES *et al.*, 2015).

No Brasil, até meados dos anos 2000, a maioria dos estudos sobre educação que empregaram abordagens quantitativas versava sobre temas como: analfabetismo, percurso escolar e fracasso escolar; fluxo escolar e análise de coortes; letramento; políticas e educação básica; financiamento da educação e municipalização; fatores sociais e educação; jovens e educação; avaliação educacional (GATTI, 2004). Dentre estes, alguns trabalhos já buscavam relacionar diretamente características dos estabelecimentos escolares com elementos da qualidade de ensino vinculados ao desempenho escolar (BARBOSA; FERNANDES, 2000; SOARES; ALVES; OLIVEIRA, 2001; CÉSAR; SOARES, 2001; RIOS-NETO; CÉSAR; RIANI, 2002; FERNANDES; NATENZON, 2003).

A relação entre qualidade de ensino e desempenho escolar não é trivial (IMBERNÓN, 2016). No entanto, desde as publicações do Relatório Coleman (COLEMAN *et al.*, 1966) e do trabalho de Eric A. Hanushek (1986), não é raro encontrar na literatura brasileira trabalhos que mensuram qualidade de ensino, eficácia escolar ou produtividade escolar, com base em algum tipo de teste ou prova de conhecimentos sobre determinado conteúdo dos ensinos fundamental ou médio (LARRUBIA *et al.*, 1976; ALBERNAZ; FERREIRA; FRANCO, 2002; JESUS; LAROS, 2004).

A disseminação dos dados das avaliações educacionais de larga escala muito tem contribuído para o avanço dos estudos sobre educação, no que concerne ao emprego de abordagens quantitativas (LAROS; MARCIANO, 2008). Destacam-se alguns estudos, como o de Cerqueira e Sawyer (2007) que, empregando o método *Grade of Membership (Fuzzy Clustering)*, propõe uma tipologia (perfis) de estabelecimentos escolares, tendo como base os dados do Censo Escolar de 2000. Os perfis extremos de escolas mostram a forte associação entre infraestrutura e eficiência escolar: escolas com melhor infraestrutura apresentam menores taxas de distorção idade-série, de reprovação e de abandono, bem como maiores taxas de aprovação, e vice-versa. Alves e Soares (2007), por meio de um levantamento longitudinal dos alunos da 5ª série do ensino fundamental de sete escolas de Belo Horizonte, mensuram os impactos do efeito-escola no desempenho escolar. Os autores apontam a importância do efeito-escola para o progresso dos alunos e ressaltam

o papel da formação das turmas, indicando que turmas homogêneas do ponto de vista das habilidades reforçam o aumento da desigualdade.

O estudo de Andrade e Soares (2008) investiga o efeito das escolas de educação básica brasileira sobre o desempenho cognitivo de alunos, por meio da modelagem hierárquica (*multilevel analysis*) de dados do Sistema Nacional de Avaliação da Educação Básica (Saeb) de 1995, 1997, 1999, 2001 e 2003. Embora o efeito das escolas, de um modo geral, tenha sido semelhante para todos os anos considerados, há um percentual não desprezível de escolas cujas práticas pedagógicas deveriam ser tomadas como modelo, com impacto positivo sobre o desempenho dos alunos. Laros e Marciano (2008) utilizaram análise multinível para quantificar o efeito-escola em 33.962 alunos, com base no teste de língua portuguesa do Saeb de 2001. Os autores apontam algumas características relacionadas à turma, dentre elas a quantidade de repetência, como fatores que influenciam o baixo rendimento. Baseado na aplicação de modelos hierárquicos logísticos (*multilevel analysis*), Riani e Rios-Neto (2008) analisam os determinantes do resultado educacional nos ensinos fundamental e médio no Brasil, considerando fatores relacionados à família e à estrutura escolar dos municípios brasileiros. O estudo aponta que a melhoria na qualidade dos recursos humanos e de infraestrutura dos serviços educacionais aumenta a probabilidade de frequentar a escola na idade correta. A qualidade dos recursos humanos é mensurada, dentre outras características, pelo percentual de docentes com curso superior. Além dos impactos diretos, essa melhoria pode ter impacto indireto, ao diminuir, eventualmente, o peso negativo de fatores relacionados ao ambiente familiar do aluno. Os autores destacam o grande efeito da proporção de professores com ensino superior sobre a qualidade da educação.

Marteleto e Andrade (2013) quantificam o efeito família no desempenho educacional dos brasileiros que participaram do Programa de Avaliação Internacional de Estudantes (Pisa) de 2006, por meio de modelos hierárquicos lineares, mostrando que as lacunas de aprendizagem, em face da forte desigualdade social, são potencializadas nas escolas brasileiras. Com base nos dados do Saeb, Rodrigues, Rios-Neto e Pinto (2013) analisam os diferenciais de rendimento escolar segundo níveis socioeconômicos dos alunos, empregando o método de decomposição contrafactual. Os resultados mostram que o aumento na cobertura escolar no país, verificado no final dos anos 1990, contribuiu para a redução da média e o crescimento da desigualdade no desempenho escolar, em função da queda do nível socioeconômico médio dos alunos. Utilizando dados da Prova Brasil de 2007 e análise multinível, Palermo, Silva e Novellino (2014) analisam os fatores associados ao desempenho escolar dos alunos do 5º ano do ensino fundamental da rede municipal de ensino do Rio de Janeiro, indicando que há um efeito relevante da turma e da escola sobre o desempenho dos alunos. Soares *et al.* (2015), combinando informações da Pesquisa Nacional por Amostra de Domicílios (PNAD) de vários anos e um *survey* com 3.418 mineiros, investigam os fatores associados ao abandono escolar no ensino médio público de Minas Gerais. O estudo revela que o aluno vulnerável ao abandono escolar é caracterizado por

baixo rendimento escolar e baixos níveis socioeconômicos. Lourenço *et al.* (2017) mostram que aspectos sociais e pedagógicos impactam na probabilidade de obtenção de melhores notas no Índice de Desenvolvimento da Educação Básica (Ideb), ao analisarem, empregando modelos regressivos logísticos multinomiais, os diferenciais municipais do Ideb de 2013.

Para lidar de forma eficiente com a multidimensionalidade envolvida na quantificação e análise da qualidade do ensino, por um lado, e com a gama cada vez maior de acesso aos microdados oriundos das avaliações em larga escala e dos censos educacionais, por outro, o campo de pesquisa que vem se destacando nos estudos educacionais é o da mineração de dados educacionais (*educational data mining* – EDM). O termo mineração de dados, também conhecido como descoberta de conhecimentos em bancos de dados (*knowledge discovery in databases* – KDD), refere-se ao processo de identificação das relações entre dados que podem produzir novos conhecimentos e gerar novas descobertas por meio da análise de grandes quantidades de dados. Por conseguinte, a EDM pode ser definida como a área de pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais (BAKER; ISOTANI; DE CARVALHO, 2011). Um dos primeiros trabalhos com esse enfoque no país é o de Kampff, Reategui e Lima (2008), que busca identificar, a partir da mineração de dados, perfis de alunos com risco de evasão ou reprovação, em ambientes virtuais de aprendizagem do ensino de graduação a distância. Os autores mencionam que a mineração de dados auxilia na descoberta de fatores potencialmente úteis e não facilmente percebidos nos dados. Posteriormente, outros autores trabalharam o tema.

Namen, Borges e Sadala (2013) analisam o potencial das técnicas de mineração de dados na modelagem de indicadores de qualidade do ensino fundamental, provenientes das bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), e discutem seus impactos na definição e avaliação de políticas públicas educacionais. Para os autores, a mineração de dados permite agilidade e facilidade na construção de bons indicadores. Costa, Cazella e Rigo (2014) empregaram técnicas de EDM para analisar o desempenho dos alunos de um curso de especialização em educação a distância da Universidade Aberta do Sistema Único de Saúde (UNA-SUS). O modelo resultante da mineração de dados classificou o potencial de evasão com precisão de 97,5%. O processo KDD foi aplicado por Fonseca e Namem (2016) aos dados da Prova Brasil de 2011, com o intuito de identificar os fatores que relacionam o perfil de professores que lecionam matemática com o grau de proficiência obtida por seus alunos. Os autores ressaltam que a técnica permite a inclusão de novas questões na discussão, para além daquelas já consolidadas na literatura. Nascimento, Cruz Junior e Fagundes (2018) utilizam as técnicas de mineração de dados para explicar os indicadores de evasão e reprovação escolares construídos com base nos dados educacionais fornecidos pelo Inep, pontuando as vantagens da técnica em selecionar variáveis mais bem relacionadas ao fenômeno estudado. O estudo de Freitas Júnior *et al.* (2019) emprega ferramentas de mineração de dados para analisar o Ideb das escolas públicas do município de Maceió, visando auxiliar

no processo decisório dos gestores educacionais pela adoção de medidas de melhoria da gestão escolar. Os autores destacam as vantagens da técnica de identificar aspectos relevantes não tradicionalmente tratados nos estudos da área.

Chama a atenção a escassez, na literatura especializada, de estudos quantitativos sobre os determinantes da qualidade de ensino ou de desempenho escolar, seja na dimensão dos alunos, das famílias ou das escolas, envolvendo o Exame Nacional do Ensino Médio (Enem). Travitzki, Calero e Boto (2014), trabalhando com os microdados de 17.359 escolas (Enem de 2009 e 2010), em modelos multiníveis, avaliam as limitações e os potenciais do exame como um indicador da eficácia da escola no Brasil. O estudo identifica várias deficiências do Enem como indicador de eficácia escolar, concluindo que o indicador revela mais sobre a condição socioeconômica das escolas do que sobre o mérito das próprias escolas, e sugere trabalhos que avaliem melhor os resultados do exame. Com base em uma perspectiva intergeracional, Travitzki, Ferrão e Couto (2016), por meio de dados do Enem de 2009 a 2012, examinam a diminuição das desigualdades educacionais do ensino médio entre as unidades da federação (UF), considerando as relações entre desempenho educacional e atributos sociodemográficos. O estudo indica a existência de substancial variabilidade de desempenho entre as escolas e a necessidade de estudos adicionais, relacionados aos determinantes do efeito escolar, levando em conta as dimensões intra e extraescolares. Outro estudo com essa mesma perspectiva é o trabalho de Dutra *et al.* (2019) que, com base nos microdados do Enem de 2011 a 2015, encontra diferenças significativas dos escores de desempenho escolar entre alunos oriundos de institutos federais no país. Os autores destacam, positivamente, que o Enem é um exame de larga escala, com periodicidade definida e com característica somativa.

À parte a controvérsia que havia a respeito da utilização do Enem como indicador de desempenho escolar (VIANNA, 2003), é inegável sua importância e abrangência, tornando-se, atualmente, uma ferramenta relevante na avaliação da qualidade do ensino médio brasileiro. O Enem é aplicado a aproximadamente 5 milhões de estudantes a cada ano e é o segundo maior exame de ensino médio no mundo, depois do *gaokao*, da China (TRAVITZKY, 2017; DUTRA *et al.*, 2019). Ainda assim, seus microdados têm sido pouco utilizados. A mitigação deste fato é uma das contribuições do presente trabalho. Outra é a utilização conjunta dos dados do Enem e do Censo Escolar na análise da qualidade de ensino. Ainda, o trabalho utiliza a técnica de mineração de dados na identificação de características-chave escolares que expliquem o desempenho de seus alunos no Enem.

São consideradas três dimensões na análise dos indicadores: condições de infraestrutura, rendimento escolar e qualificação docente. Pretende-se responder a quatro perguntas. Primeiro, se há diferenças significativas no desempenho escolar de alunos do ensino médio segundo o tipo de escolas, públicas ou privadas. Em segundo lugar, se existem diferenças significativas no desempenho escolar dos alunos segundo o tipo de ensino médio, regular ou profissionalizante. Em terceiro, com relação ao sistema público de ensino médio regular, se há diferenciais interestaduais relevantes de desempenho escolar. Por fim, busca-se

investigar se os fatores escolares que mais contribuem para o melhor nível de qualidade das escolas da rede pública de ensino médio regular estão relacionados às condições de infraestrutura, à progressão discente ou à qualificação docente. Pretende-se discorrer em larga escala sobre a qualidade do ensino médio no Brasil e buscar evidências que possam contribuir para a melhoria do sistema educacional brasileiro.

O presente artigo objetiva analisar os diferenciais de desempenho no Enem das escolas que oferecem ensino médio, segundo os tipos de ensino e de administração, abordar os diferenciais interestaduais de desempenho escolar, considerando as escolas de ensino médio regular, e identificar quais os fatores – infraestrutura, qualificação docente ou progressão discente – são preponderantes na determinação do nível de desempenho das escolas no Enem.

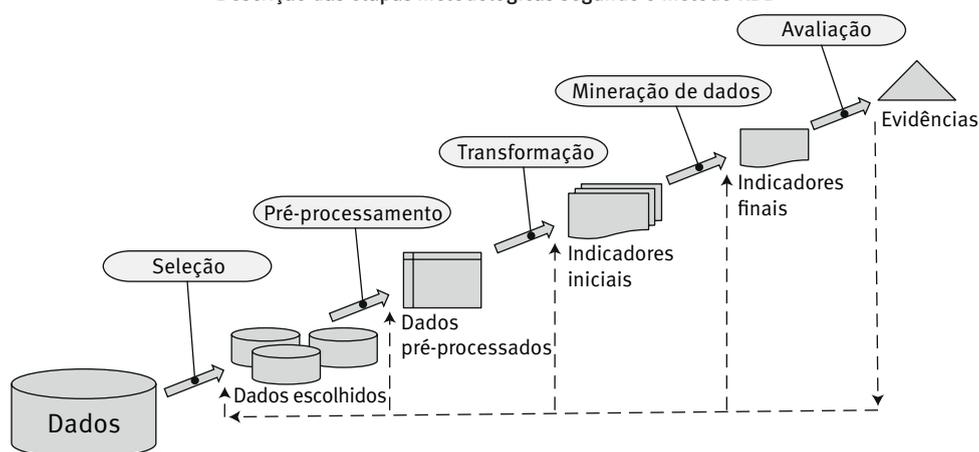
Para cumprir os objetivos, são utilizados os microdados das edições de 2016, 2017 e 2018 do Enem e os microdados do Censo Escolar 2017. Para a seleção dos indicadores, são aplicadas técnicas de mineração de dados – análise de regressão (*stepwise*) e análise de discriminante (*stepwise*) – e, para a quantificação do impacto dos indicadores, é utilizado um modelo de regressão logística múltipla (RLM). Uma das definições do conceito de efeito escola, segundo Andrade e Soares (2008), corresponderia à parcela da proficiência dos alunos que pode ser atribuída às práticas do e no estabelecimento escolar. Em síntese, o que se propõe aqui, com o modelo de RLM, é a decomposição do *efeito escola* em três componentes: *efeito infraestrutura escolar*, *efeito rendimento escolar* e *efeito prática docente* na qualidade do ensino médio.

Material e métodos

A estratégia de mineração de dados escolares encontra respaldo nos trabalhos de Baker, Isotani e De Carvalho (2011, p. 5), Nascimento, Cruz Junior e Fagundes (2018, p. 2) e Namen, Borges e Sadala (2013, p. 697). O método KDD, segundo Fonseca e Namem (2016, p. 138), é composto por cinco fases metodológicas, cujo objetivo principal é a “identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, que estão embutidos nos dados”. Essas cinco fases correspondem, respectivamente: à seleção das bases de dados; ao pré-processamento, caracterizado pela estruturação dos dados em um formato que permita sua manipulação, seja por meio de planilhas eletrônicas ou de pacotes estatísticos; à transformação, entendida como a geração de variáveis secundárias elaboradas pela combinação ou transformação de variáveis primárias, bem como junção ou separação de bases de dados pré-processadas, cuja meta é a identificação dos indicadores iniciais a serem minerados; à mineração de dados, descrita como aplicação de procedimentos computacionais e/ou métodos estatísticos que visem selecionar casos e/ou variáveis que indiquem padrões que tornem mais compreensíveis o fenômeno que se está estudando; e à avaliação, composta pela análise crítica dos resultados da etapa anterior e que pode envolver a aplicação de testes estatísticos ou computacionais, e cujas

evidências podem influenciar, recursivamente, as etapas anteriores do processo. A Figura 1 sintetiza os procedimentos descritos, que foram utilizados nesse trabalho.

FIGURA 1
Descrição das etapas metodológicas segundo o método KDD



Fonte: Adaptado de Fayyad *et al.* (1996).

A primeira fase do método KDD (seleção) iniciou-se com a escolha dos dados. Para alcançar os objetivos desse estudo, o primeiro passo foi a construção de duas bases de dados: uma com as informações sobre o desempenho dos estudantes que participaram das edições de 2016, 2017 e 2018 do Enem; e outra com informações sobre os estabelecimentos escolares constantes nos dados do Censo Escolar de 2017. O Enem é um exame nacional, que visa avaliar o nível de proficiência do egresso da educação básica, por meio da identificação de competências e habilidades adquiridas ao longo da vida escolar. O exame é composto por testes de matemática, linguagens e códigos, ciências humanas, ciências naturais e redação. Os resultados nos testes são mensurados pelo sistema de avaliação Teoria de Resposta ao Item (TRI), que avalia os acertos e atribui pontuação distinta a cada questão, de acordo o nível de dificuldade (DUTRA *et al.*, 2019). O resultado final, ou score, é calculado como a média aritmética do resultado dos quatro testes e da redação. Além dos escores parciais e final de cada candidato, a base de dados contém informações sociodemográficas, condição de conclusão do EM e alguns dados sobre a escola, como tipo, dependência administrativa, código da escola,¹ dentre outras. O Censo Escolar é uma pesquisa que contabiliza anualmente os registros educacionais, sendo a principal fonte de informações para as análises das condições de oferta e atendimento do sistema educacional brasileiro. O seu levantamento reúne informações sobre todas as etapas e modalidades de ensino, oferecendo um quadro detalhado sobre as características das escolas, dos profissionais da educação e dos alunos no país (INEP, 2017).

¹ A variável código da escola permite a junção das bases de dados do Censo Escolar e do ENEM.

Na fase de pré-processamento (segunda fase) das bases de dados do Enem, foram selecionados apenas os registros dos estudantes que assinalaram serem concluintes do ensino médio no ano em questão. Além de atender aos objetivos do estudo (inferir sobre a qualidade do ensino), garantiu-se o vínculo do aluno à escola por ele declarada na inscrição. Ao todo, nas três edições do Enem, foram computados os dados de aproximadamente 20,8 milhões de inscritos, dos quais 25,4% corresponderam aos dos estudantes de interesses dessa pesquisa. As bases dos três anos foram empilhadas. Na terceira fase (transformação), a base de dados foi agregada segundo o código da escola, sendo que cada linha passou a conter as informações sobre as escolas que tinham alunos concluintes que participaram do Enem. A nova base foi formada por seis variáveis: código da escola, UF da escola, dependência administrativa, total de alunos concluintes participantes, tipo de ensino médio e indicador de desempenho escolar. Este último corresponde à média dos escores dos alunos concluintes de cada escola, considerando os anos de 2016, 2017 e 2018. A utilização da média dos três anos visa retirar eventuais flutuações de desempenho e centralizar o desempenho no ano intermediário, 2017, ano ao qual também se referem as informações da base de dados do Censo Escolar.

$$EE_i = \frac{\sum S_{16} + \sum S_{17} + \sum S_{18}}{N} \quad (1)$$

Em que: EE_i é o indicador de desempenho da escola i ; S_{16} é o escore total de um aluno da escola i em 2016; S_{17} corresponde ao escore total de um aluno da escola i em 2017; S_{18} refere-se ao escore total de um aluno da escola i em 2018; e N é o total de alunos concluintes na escola i que participaram do Enem nos anos considerados.

Uma vez consolidado o banco de dados do Enem, foi possível realizar os primeiros cruzamentos de informações e testar estatisticamente alguns diferenciais de desempenho escolar, com o intuito de buscar evidências que ajudassem a responder às questões que norteiam este trabalho. Os métodos estatísticos utilizados na comparação entre várias médias de desempenho escolar foram a análise de variância (Anova) e a distância mínima de significância de Fisher (LSD). Em uma Anova, a hipótese nula é a de que não há diferença significativa entre as médias dos grupos; a hipótese alternativa pressupõe que exista pelo menos uma diferença significativa entre elas. Ao se calcular o valor da estatística F , com base na relação entre as variâncias (quadrado médio), dentro e entre os grupos, verifica-se probabilidade associada a ele (valor p). Em geral, se o valor p for menor que 0,05, a hipótese nula é rejeitada e conclui-se que as médias dos grupos não são iguais entre si. Para se averiguar se as médias de todos os cruzamentos são estatisticamente diferentes umas das outras, é necessário aplicar um teste de comparação de grupos. Existem diversos testes que podem efetuar múltiplas comparações por pares de categorias da variável, entendida como fator no modelo Anova. Um dos pioneiros é o teste LSD, que informa onde estão os pares de grupos cujas médias são significativamente diferentes (FISHER, 1935).

Para a base de dados do Censo Escolar, a segunda fase do método KDD (pré-processamento) consistiu na construção de diversos indicadores, a partir das variáveis originais,

e da incorporação de outros indicadores, previamente processados e disponibilizados pelo Inep.² A fase seguinte (transformação) consistiu na seleção dos indicadores iniciais. Essa escolha foi pautada pela maior correspondência às três dimensões que se relacionam ao conceito de qualidade de ensino: infraestrutura escolar, progressão discente e qualificação docente. O Quadro 1 apresenta os indicadores selecionados. A terceira fase do método KDD foi encerrada com a integração dos indicadores iniciais provenientes das duas bases de dados.

QUADRO 1
Indicadores construídos a partir dos microdados do Censo Escolar, selecionados para a etapa de mineração de dados

N	Indicador	N	Indicador
1	Localização	69	Percentual de professores do ensino médio – artes
2	Água consumida pelos alunos passa por um processo de filtragem	70	Percentual de professores do ensino médio – educação física
3	Abastecimento de água – inexistente	71	Percentual de professores do ensino médio – história
4	Abastecimento de energia elétrica – inexistente	72	Percentual de professores do ensino médio – geografia
5	Esgoto sanitário – inexistente	73	Percentual de professores do ensino médio – filosofia
6	Existência de sala de professores	74	Percentual de professores do ensino médio – ensino religioso
7	Existência de laboratório de informática	75	Percentual de professores do ensino médio – sociologia
8	Existência de laboratório de ciências	76	Percentual de professores do ensino médio – estudos sociais/sociologia
9	Existência de quadra de esportes coberta	77	Percentual de professores do ensino médio – informática e computação
10	Existência de biblioteca e/ou sala de leitura	78	Percentual de professores do ensino médio – atendimento especial
11	Existência de banheiro para alunos com deficiência/mobilidade reduzida	79	Percentual de professores do ensino médio – diversidade sociocultural
12	Média de disciplinas do ensino médio por professor	80	Percentual de professores do ensino médio – línguas
13	Média de turmas do ensino médio por professor	81	Percentual de professores do ensino médio – pedagogias
14	Percentual de docentes EM com curso superior completo	82	Percentual de professores do ensino médio – outras especialidades
15	Razão professor aluno – ensino médio	83	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – ciências naturais
16	Total de turmas no ensino médio – regular		
17	Média do número de matrículas no ensino médio		
18	Desvio padrão do número de matrículas no ensino médio	84	Percentual de docentes com formação superior de bacharelado na mesma disciplina que lecionam, mas sem licenciatura ou curso de complementação pedagógica concluído – ciências naturais
19	Taxa de aprovação no ensino médio – total		

(Continua)

² Os indicadores incluídos foram as taxas de aprovação, reprovação e abandono escolar. Essas taxas, embora refiram-se ao ano em questão e tenham como base as matrículas registradas no Censo Escolar, são calculadas a partir da combinação destas com informações coletadas ao final do ano letivo; portanto, não estão incluídas nos microdados do Censo Escolar.

(Continuação)

N	Indicador	N	Indicador
20	Taxa de aprovação no ensino médio – primeira série	85	Percentual de docentes com formação superior de licenciatura em área diferente daquela que lecionam, ou com bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente daquela que lecionam – ciências naturais
21	Taxa de aprovação no ensino médio – segunda série		
22	Taxa de aprovação no ensino médio – terceira série		
23	Taxa de reprovação no ensino médio – total	86	Percentual de docentes com outra formação superior, mas sem licenciatura ou bacharelado ou complementação pedagógica na área que lecionam – ciências naturais
24	Taxa de reprovação no ensino médio – primeira série		
25	Taxa de reprovação no ensino médio – segunda série	87	Percentual de docentes sem formação superior – ciências naturais
26	Taxa de reprovação no ensino médio – terceira série	88	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – ciências humanas
27	Taxa de abandono no ensino médio – total		
28	Taxa de abandono no ensino médio – primeira série		
29	Taxa de abandono no ensino médio – segunda série	89	Percentual de docentes com formação superior de bacharelado na mesma disciplina que lecionam, mas sem licenciatura ou curso de complementação pedagógica concluído – ciências humanas
30	Taxa de abandono no ensino médio – terceira série		
31	Razão de microcomputadores e alunos do ensino médio	90	Percentual de docentes com formação superior de licenciatura em área diferente daquela que lecionam, ou com bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente daquela que lecionam – ciências humanas
32	Razão de aparelhos de televisão e alunos do ensino médio		
33	Razão de aparelhos de DVD e alunos do ensino médio		
34	Razão de copiadoras e alunos do ensino médio	91	Percentual de docentes com outra formação superior, mas sem licenciatura ou bacharelado ou complementação pedagógica na área que lecionam – ciências humanas
35	Razão de impressoras e alunos do ensino médio		
36	Razão de projetores multimídias (Datashow) e alunos do ensino médio	92	Percentual de docentes sem formação superior – ciências humanas
37	Percentual de computadores da escola acessados pelos alunos do EM	93	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – linguagem e códigos
38	Acesso à Internet		
39	Tamanho da escola medido por número de matriculados no EM		
40	Porcentagem das matrículas do ensino médio que atendem o EM regular	94	Percentual de docentes com formação superior de bacharelado na mesma disciplina que lecionam, mas sem licenciatura ou curso de complementação pedagógica concluído – linguagem e códigos
41	Porcentagem das matrículas do ensino médio que atendem o ensino técnico e profissionalizante integrado		
42	Percentual de alunos que fizeram o Enem	95	Percentual de docentes com formação superior de licenciatura em área diferente, ou bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente da que lecionam – linguagem e códigos
43	Percentual de professores – docente		
44	Percentual de professores – auxiliar/assistente educacional	96	Percentual de docentes com outra formação superior, mas sem licenciatura ou bacharelado ou complementação pedagógica na área que lecionam – linguagem e códigos
45	Percentual de professores – profissional/monitor atividade compl.		
46	Percentual de professores – tradutor intérprete de libras	97	Percentual de docentes sem formação superior – linguagem e códigos

(Continua)

(Continuação)

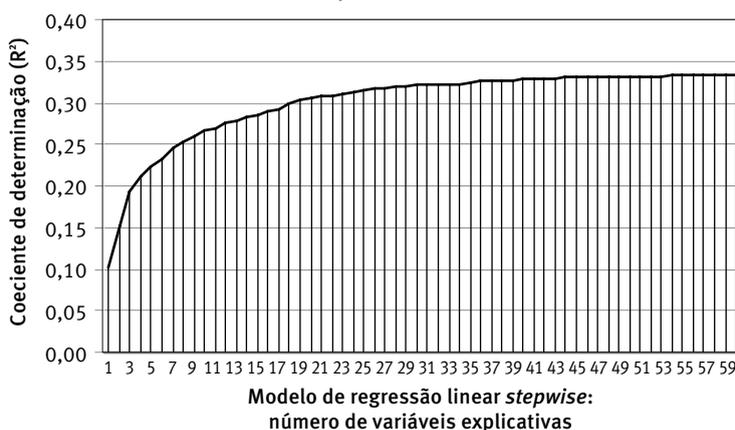
N	Indicador	N	Indicador
47	Percentual de professores – docente titular – coordenador de tutoria (de módulo ou disciplina) – EAD	98	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – matemática
48	Percentual de professores – docente tutor – auxiliar (de módulo ou disciplina) – EAD	99	Percentual de docentes com formação superior de bacharelado na mesma disciplina que lecionam, mas sem licenciatura ou curso de complementação pedagógica concluído – matemática
49	Percentual de docentes sem informação de formação adequada – geral		
50	Idade média dos professores do ensino médio		
51	Desvio padrão da idade dos professores do ensino médio	100	Percentual de docentes com formação superior de licenciatura em área diferente daquela que lecionam, ou com bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente daquela que lecionam – matemática
52	Razão professor aluno – ensino médio: química		
53	Razão professor aluno – ensino médio: física		
54	Razão professor aluno – ensino médio: matemática	101	Percentual de docentes com outra formação superior, mas sem licenciatura ou bacharelado ou complementação pedagógica na área que lecionam – matemática
55	Razão professor aluno – ensino médio: biologia		
56	Razão professor aluno – ensino médio: português	102	Percentual de docentes sem formação superior – matemática
57	Razão professor aluno – ensino médio: história	103	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – geral
58	Razão professor aluno – ensino médio: geografia		
59	Percentual de professores do ensino médio – química		
60	Percentual de professores do ensino médio – física	104	Percentual de docentes com formação superior de bacharelado na disciplina que lecionam, mas sem licenciatura ou curso de complementação pedagógica concluído – geral
61	Percentual de professores do ensino médio – matemática		
62	Percentual de professores do ensino médio – biologia		
63	Percentual de professores do ensino médio – língua portuguesa		
64	Percentual de professores do ensino médio – língua inglesa	105	Percentual de docentes com formação superior de licenciatura em área diferente da que lecionam, ou bacharelado nas disciplinas da base curricular comum e complementação pedagógica em área diferente da que lecionam – geral
65	Percentual de professores do ensino médio – língua espanhola		
66	Percentual de professores do ensino médio – língua francesa		
67	Percentual de professores do ensino médio – outras línguas portuguesas	106	Percentual de docentes com outra formação superior, sem licenciatura, bacharelado ou complementação pedagógica na área que lecionam – geral
68	Percentual de professores do ensino médio – língua indígena	107	Percentual de docentes sem formação superior – geral

Fonte: Inep. Censo Escolar 2017.

A fase de mineração de dados (quarta fase) consistiu em duas etapas. A primeira teve início com a análise de regressão, tendo o indicador de desempenho (EE) escolar como variável dependente e os 107 indicadores iniciais como covariáveis. Na etapa seguinte, os resultados dos 60 modelos sugeridos pelo método *stepwise* foram examinados, na busca por um modelo mais eficiente.

O resultado do modelo saturado – ou seja, com todas as covariáveis – indicou um grau de ajuste de 0,339, segundo seu coeficiente de determinação (R^2). O modelo *stepwise*, por sua vez, reduziu o número de variáveis para 60, com um ajuste de 0,336. A evolução do grau de ajuste do modelo em função do número de covariáveis significativas da modelagem estatística pode ser observada na Figura 2. Verifica-se que a taxa de incremento do R^2 diminui consideravelmente a partir de um modelo com oito covariáveis e passa a ser muito baixo a partir de 20 variáveis. O modelo escolhido foi o de 16 covariáveis, que apresentou um coeficiente de determinação de 0,289, com um pequeno número de variáveis.

FIGURA 2
Evolução do coeficiente de determinação (R^2) do modelo *stepwise*, em função do número de variáveis explicativas



Fonte: Inep: Censo Escolar 2017; Enem, 2016 a 2018

Os resultados de um modelo *stepwise*, como alertado por Smith (2018), devem ser interpretados com cautela. Por não indicarem o arranjo mais adequado de covariáveis, ou por apontarem um arranjo que se mostra eficiente apenas para o conjunto de casos amostrados, tais resultados podem, muitas vezes, incorrer em erros graves. No caso desse estudo, por tratar-se de praticamente o universo das escolas públicas de ensino médio, é o primeiro tipo de erro que pode ocorrer. De fato, algumas das 16 variáveis selecionadas pelo modelo não são de fácil interpretação, quando analisadas à luz do desempenho escolar, e não coadunam com as dimensões analíticas desse estudo. Procedeu-se, então, à escolha de novos indicadores, por meio de análise exploratória. Buscou-se uma correlação tanto teórico-metodológica quanto estatística, entre qualidade de ensino e desempenho escolar, de um lado, e infraestrutura escolar, progressão discente e qualificação docente, de outro, tal como sugerido por Cunha, Perez e Aidar (2001). Dessa forma, efetuou-se a segunda etapa da quarta fase do método KDD, com substituição de nove, dos 16 indicadores iniciais. O ajuste do modelo com os novos indicadores foi de 0,265. O Quadro 2 traz os indicadores selecionados pelo modelo e pela análise exploratória. Os indicadores serão inseridos na análise discriminante, etapa da metodologia.

QUADRO 2

Indicadores selecionados para inserção na análise de discriminante, segundo modelo aplicado

Modelo <i>stepwise</i> de 16 indicadores ($R^2 = 0,289$)	Modelo final de 16 indicadores ($R^2 = 0,265$)
Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – geral	Percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – geral
Existência na escola – biblioteca e/ou sala de leitura	Existência na escola – biblioteca e/ou sala de leitura
Existência na escola – laboratório de ciências	Existência na escola – laboratório de ciências
Existência na escola – quadra de esportes coberta	Existência na escola – quadra de esportes coberta
Percentual de docentes com curso superior completo do ensino médio	Percentual de docentes com curso superior completo do ensino médio
Taxa de abandono no ensino médio – total	Taxa de abandono no ensino médio – total
Localização	Localização
Desvio padrão da idade dos professores do ensino médio	Média de disciplinas do ensino médio por professor
Desvio padrão do número de matrículas no ensino médio	Média de turmas do ensino médio por professor
Percentual de docentes com formação superior de licenciatura em área diferente daquela que lecionam, ou com bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente daquela que lecionam – geral	Razão de microcomputadores por aluno
Percentual de professores – docente	Razão professor aluno – ensino médio
Percentual de professores do ensino médio – artes	Tamanho da escola (número de matriculados no EM)
Percentual de professores do ensino médio – educação física	Taxa de reprovação no ensino médio – total
Percentual de professores do ensino médio – ensino religioso	Total de turmas no ensino médio – regular
Percentual de professores do ensino médio – língua indígena	Existência na escola – laboratório de informática
Acesso à internet	Existência na escola – sala de professores

Fonte: Inep: Censo Escolar, 2017.

Até aqui, a qualidade do ensino médio pode ser entendida como resultante de uma função de produção educacional adaptada, na qual o desempenho escolar é um produto de 16 insumos. À semelhança do que foi proposto por Larrubia *et al.* (1976), nesse estudo os insumos referem-se a um único nível hierárquico. A diferença é que a função de produção educacional, no estudo citado, é relativa aos alunos e, no presente trabalho, refere-se às escolas. Imbernón (2016) pontua que o conceito de qualidade de ensino é multidimensional, sendo que essa multidimensionalidade não é capturada por modelos econométricos. Para Dourado (2009), uma escola de qualidade caracteriza-se por apresentar, entre outros aspectos, um ambiente agradável, professores satisfeitos e alunos com bom desempenho escolar.

No presente estudo, a qualidade do ensino foi entendida como um atributo da escola, mesmo que o indicador tenha sido obtido com base no desempenho individual dos alunos. Sendo um atributo complexo e multidimensional, optou-se por representá-lo em sua forma

discreta, isto é, em faixas de qualidade de ensino e, assim, buscar uma maior aderência conceitual na avaliação dos indicadores finais, na quinta etapa do método KDD (avaliação).

Como a distribuição dos valores do indicador de desempenho escolar (EE) das escolas obedecia a uma distribuição relativamente normal, essa medida foi então discretizada segundo seus desvios padrões. Foram estabelecidos três níveis de qualidade para o ensino: baixo, para valores de EE inferiores à média menos 1,6 desvio padrão; intermediário para valores entre a média e mais ou menos 1,6 desvio padrão; e alto, para valores superiores à média mais 1,6 desvio padrão. Em seguida foi empregada uma análise de discriminante (*stepwise*), com o intuito de avaliar os indicadores finais.

A análise discriminante é uma técnica poderosa de mineração de dados, cujo método *stepwise* é capaz de selecionar as variáveis com grande poder de discriminação e eliminar as de menor contribuição, com base no maior valor da estatística F e/ou menor valor da estatística (*lambda*) de Wilks (FERNANDEZ, 2002). Os resultados desse procedimento mostraram que todos os indicadores foram significativos na discriminação dos níveis de qualidade de ensino, tal como apresentado na Tabela 1.

TABELA 1
Resultados da estatística *lambda* de Wilks e da estatística F e significância estatística das variáveis selecionadas para o modelo

Número de variáveis	<i>Lambda</i> de Wilks	Estatística F	Significância estatística
1	0,965	315,4	0,000
2	0,888	535,5	0,000
3	0,865	438,5	0,000
4	0,859	346,2	0,000
5	0,858	279,4	0,000
6	0,854	239,9	0,000
7	0,85	212,5	0,000
8	0,823	223,7	0,000
9	0,819	204,1	0,000
10	0,814	190,2	0,000
11	0,808	179,2	0,000
12	0,784	189,1	0,000
13	0,783	175,0	0,000
14	0,778	167,3	0,000
15	0,775	158,8	0,000
16	0,769	153,3	0,000

Fonte: Inep. Censo Escolar 2017.

Por meio do método KDD, constatou-se que os 16 indicadores selecionados são relevantes no que se relaciona à questão da qualidade do ensino médio das escolas em escala nacional. Cabe, agora, buscar quais são os indicadores mais relevantes entre si. Para avaliar os efeitos dos indicadores escolares nos níveis de qualidade do ensino médio das escolas públicas, foi empregado o método da análise de regressão logística multinomial (HOSMER, LEMESHOW, 2000)

A regressão logística multinomial (RLM) pode ser entendida como um modelo linear generalizado. Dado que as categorias de uma variável resposta, Y , são codificadas em 0, 1, ou 2, para desenvolver o modelo, assume-se que se têm “ p ” covariáveis e um termo constante, especificado pelo vetor x , de extensão $p+1$. Com isso, verificam-se as probabilidades condicionais das categorias da variável Y . O método RLM possui caráter analítico e foi concebido como um modelo regressivo, para ocasiões em que a variável dependente é categórica ou ordinal, como os dados em questão. A variável resposta são as classes de nível da qualidade do ensino médio (baixo, intermediário e alto) e as variáveis independentes são os 16 indicadores selecionados. A classe de referência do nível da qualidade de ensino médio, para efeitos analíticos do modelo, foi a segunda (nível intermediário).

Os indicadores finais foram, por último, dispostos da seguinte forma nas três dimensionalidades que norteiam este estudo:

- infraestrutura escolar, composta por oito indicadores (localização, tamanho da escola, cinco itens existentes na escola – sala de professores, laboratório de informática, quadra de esporte coberta, laboratório de ciências e biblioteca e/ou sala de leitura – e razão microcomputadores por aluno);
- dois indicadores de rendimento escolar (taxa de reprovação e taxa de abandono);
- seis indicadores de prática docente (percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – geral, percentual de professor com ensino superior completo, total de turma no ensino médio, razão professor aluno, médias de turmas por professor e média de disciplina por professor).

Resultados

O desempenho escolar dos concluintes do ensino médio, por ocasião de suas participações no Enem, é mostrado na Tabela 2. Os resultados indicam que há diferenciais segundo tipo de escola e de ensino, respondendo às duas primeiras perguntas desse trabalho. Embora as respostas fossem esperadas, do ponto de vista do senso comum e da própria literatura, é fundamental enxergar a dimensão desses diferenciais. Para além dos diferenciais citados, os resultados mostram que, entre as escolas públicas, a dependência administrativa importa. Foram realizados testes Anova para as médias e LSD para as comparações entre os pares de médias dos vários grupos analisados. Foram encontrados resultados significativos ($p < 0,05$) em todas as análises de variância e em praticamente todos os testes LSD realizados.³

A média geral do escore desses estudantes foi de 512,44 pontos. Ao se compararem apenas os valores referentes ao tipo de dependência administrativa, a menor média foi

³ Os poucos resultados não significativos desse teste podem ser conferidos nas Tabelas 1 e 2 do Anexo.

observada entre os estudantes de escolas estaduais (492,25), seguida pela média dos estudantes das escolas municipais (528,15), das escolas federais (592,20) e das escolas privadas (593,63). A diferença de 100 pontos verificada entre as médias extremas mostra-se mais relevante quando se considera que os alunos de escolas estaduais representavam quase 80% do total de alunos pesquisados. Entre as escolas públicas, as federais têm uma média bastante superior, quase se igualando às escolas privadas; embora em número reduzido – mas não desprezível –, esse diferencial, principalmente em relação às estaduais, pode trazer algumas reflexões importantes. Estariam esses diferenciais relacionados a práticas pedagógicas, como sugere o estudo de Andrade e Soares (2008), ou à progressão dos alunos, como indicam Laros e Marciano (2008), ou mesmo à desigualdade social no país, como pontuado por Marteleto e Andrade (2013)? A Anova para essas médias revelou-se significativa ($p < 0,0001$) e o teste LSD indicou que houve diferenças significativas ($p < 0,05$) entre todas elas.

As mesmas reflexões se aplicam na análise dos resultados dos escores de desempenho acadêmico segundo o tipo de ensino médio oferecido pela escola (técnico, regular, ou profissionalizante integrado). As escolas que ofereciam apenas o ensino médio regular obtiveram o menor valor (508,7) e as que ofereciam ensino técnico, o maior (557,06). A Anova para essas médias também foi significativa ($p < 0,0001$) e o teste LSD indicou, igualmente, que houve diferenças significativas ($p < 0,05$) entre todas elas. Quando se compararam os escores de desempenho, levando-se em conta os dois aspectos conjuntamente (tipo de ensino e dependência administrativa), notou-se que o menor valor foi percebido entre os alunos das escolas públicas estaduais que ofereciam o ensino médio regular (488,64) e o maior, entre as escolas públicas federais que também ofereciam o ensino médio regular (637,71). Nesse caso, embora a Anova tenha sido significativa ($p < 0,001$), o teste LSD não se mostrou significativo ($p < 0,05$) para os seguintes pares de valores: os das escolas privadas de ensino técnico (511,89) e as escolas estaduais de ensino profissionalizante integrado (511,61); e os das escolas municipais de ensino técnico (545,47) e as municipais de ensino profissionalizante integrado (545,81).

As análises seguintes se concentraram nas escolas públicas de ensino médio regular. Como mencionado, elas abrangem a maioria dos estudantes pesquisados (3,4 milhões, ou 70,51%) e a maior parte das escolas com alunos concluintes do ensino médio que prestaram o Enem no período (17,6 mil, ou 62,35%), como pode ser observado na Tabela 3, que também apresenta o desempenho acadêmico médio dos alunos dessas escolas segundo unidades da federação (UF). Foram realizados testes Anova e LSD entre as médias, cujos resultados se encontram no Anexo. Foram encontradas diferenças regionais nos escores médios do Enem no país, respondendo à terceira pergunta deste trabalho. Apesar dos diferenciais evidentes, observa-se que a amplitude entre os valores mínimo e máximo é relativamente pequena. Os maiores escores foram observados no Distrito Federal (513,08), Rio Grande do Sul (508,78), Santa Catarina (507,42) e Minas Gerais (506,56), enquanto os menores ocorreram no Maranhão (465,37), Ceará (467,67), Amapá (468,64) e Alagoas

(468,69). A Anova revelou-se significativa ($p < 0,001$), indicando que as médias não podem ser consideradas iguais. No entanto, o teste LSD não se mostrou significativo ($p < 0,05$) entre os seguintes pares de valores: Acre (479,33) e Mato Grosso (479,30); Amapá (468,64) e Alagoas (468,69); e Amazonas (473,49) e Roraima (474,01).

TABELA 2

Estudantes concluintes do ensino médio que participaram do Enem, escolas com alunos concluintes do ensino médio que participaram do Enem e escores médios dos estudantes, segundo tipo de ensino e dependência administrativa
Brasil – 2016-2018

Tipo de ensino e dependência administrativa	Estudantes	Escolas	Escores
Ensino profissionalizante (técnico)			
Federal	110.374	516	586,41
Estadual	123.983	554	536,27
Municipal	6.410	33	545,47
Privada	13.079	431	511,89
Total	253.846	1.534	557,06
Ensino médio regular			
Federal	11.235	33	637,71
Estadual	3.406.604	17.432	488,64
Municipal	18.836	181	518,59
Privada	759.000	7.472	596,53
Total	4.195.675	25.118	508,70
Ensino médio e profissionalizante			
Federal	6.916	23	610,65
Estadual	351.811	1.121	511,61
Municipal	3.920	21	545,81
Privada	61.566	476	575,22
Total	424.213	1.641	522,77
Total			
Federal	128.525	572	592,20
Estadual	3.882.398	19.107	492,25
Municipal	29.166	235	528,15
Privada	833.645	8.379	593,63
Total	4.873.734	28.293	512,44

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

TABELA 3
Estudantes concluintes do ensino médio que participaram do Enem, escolas com alunos concluintes do ensino médio que participaram do Enem e escores médios dos estudantes, segundo unidades da federação
Brasil – 2016-2018

Unidades da federação	Estudantes	Escolas	Escores
Rondônia	38.857	180	484,35
Acre	21.788	151	479,33
Amazonas	101.754	311	473,49
Roraima	7.516	110	474,01
Pará	162.315	565	475,50
Amapá	20.630	105	468,64
Tocantins	31.205	230	472,89
Maranhão	148.036	788	465,37
Piauí	55.857	405	470,66
Ceará	266.391	537	467,67
Rio Grande do Norte	40.855	236	478,90
Paraíba	60.514	376	480,91
Pernambuco	194.190	750	481,40
Alagoas	47.301	209	468,69
Sergipe	38.256	167	487,24
Bahia	198.043	971	481,70
Minas Gerais	346.718	2.199	506,56
Espírito Santo	47.526	194	492,49
Rio de Janeiro	210.477	1.093	499,79
São Paulo	755.009	3.780	496,48
Paraná	141.001	1.260	498,82
Santa Catarina	94.716	678	507,42
Rio Grande do Sul	129.834	961	508,78
Mato Grosso do Sul	35.371	244	491,53
Mato Grosso	47.976	415	479,30
Goiás	133.252	635	491,25
Distrito Federal	61.211	92	513,08
Brasil	3.436.599	17.642	489,29

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

Os resultados da Tabela 3 indicam, por um lado, a baixa disparidade dos escores de desempenho escolar entre as unidades da federação. A diferença entre o maior valor (513,08, no DF) e o menor (465,37, no MA) foi de 47,71 pontos, ou 9,3%. Por outro lado, o nível geral do desempenho médio escolar no Brasil foi baixo (489,29).⁴ Embora reflitam, de certa forma, as desigualdades socioeconômicas do país, os resultados podem indicar que, em termos de desempenho no Enem, as desigualdades interestaduais tenham menor efeito, bem como que as desigualdades internas, mesmo em unidades da federação mais desenvolvidas, puxam a média das UF para baixo. Marteleto e Andrade (2013),

⁴ Para efeitos de comparação, o aluno que ingressou na Universidade Federal de Minas Gerais (UFMG) em 2017, pela modalidade “ampla concorrência”, com a menor nota na prova do Enem de 2016, obteve um escore geral de 520,08 pontos (UFMG, 2017).

por exemplo, chamam a atenção para a potencialização das desigualdades nas escolas brasileiras. Destarte, coube indagar quais dos aspectos escolares selecionados estariam mais associados aos níveis de qualidade do ensino médio, calculados com base nesses desempenhos escolares.

A mesclagem dos microdados do Censo Escolar com os do Enem permitiu que 17.552 escolas públicas de ensino médio fizessem parte do modelo RLM. As escolas foram classificadas em três níveis de proficiência: baixo, intermediário e alto. A escala de níveis foi construída utilizando-se, como critério, a média e o desvio padrão das proficiências de todas as escolas (considerando o período de 2016 a 2018), conforme descrito na metodologia. A proficiência das escolas foi calculada como a média da proficiência dos alunos concluintes que fizeram a prova do Enem em 2016, 2017 e 2018. Na Tabela 4, os resultados indicam que, das pouco mais de 17.500 escolas analisadas, 13,25% foram classificadas com qualidade de ensino baixa, 74,35% com qualidade de ensino intermediária e 12,38% com alta qualidade de ensino.

TABELA 4
Escolas, proficiência média, desvio padrão, valores mínimo e máximo do desempenho médio escolar, segundo nível de qualidade do ensino médio das escolas públicas
Brasil – 2016-2018

Qualidade do ensino médio das escolas públicas	Escolas	Desempenho escolar (2016-2018)			
		Média	Desvio padrão	Mínimo	Máximo
Baixa	2.327	438,92	19,82	276,67	455,47
Média	13.051	483,90	14,51	455,49	511,35
Alta	2.174	528,19	22,16	511,36	728,42
Total	17.552	483,42	27,94	276,67	728,42

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

O modelo foi alimentado por 16 indicadores, sendo seis qualitativos⁵ e dez quantitativos.⁶ A Tabela 5 traz a distribuição percentual dos indicadores qualitativos pelos três níveis de qualidade de ensino médio das escolas públicas no Brasil e fornece um interessante panorama: 87,4% dessas escolas estavam localizadas em áreas urbanas, 92,1% possuíam sala de professores, 86,7% estavam equipadas com biblioteca ou sala de leitura, 84,3% tinham laboratório de informática e 37,3% declararam possuir laboratório de ciências.

⁵ Optou-se, para fins de composição e análise das tabelas, em rotular os indicadores qualitativos dando ênfase à sua categoria de referência. Assim, o indicador localização passou a ser designado *escola em área urbana, existência na escola – sala de professores, passou para escola com sala de professores* e assim por diante.

⁶ O indicador *percentual de docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina que lecionam, com curso de complementação pedagógica concluído – geral* foi rotulado, nas tabelas a seguir, por *percentual de docentes com formação qualificada*.

TABELA 5
Proporção de escolas, por nível de qualidade do ensino médio das escolas públicas, segundo indicadores qualitativos
Brasil – 2016-2018

Indicadores qualitativos (categoria de análise)	Qualidade do ensino médio			Total
	Baixa	Média	Alta	
Escola em área urbana	61,80	90,74	94,80	87,41
Escola com sala de professores	77,31	94,06	96,41	92,13
Escola com laboratório de informática	64,03	86,79	91,63	84,37
Escola com laboratório de ciências	21,23	37,02	56,03	37,28
Escola com quadra de esportes coberta	47,66	78,34	84,82	75,07
Escola com biblioteca e/ou sala de leitura	65,15	89,16	94,89	86,69
Escolas (n. absolutos)	2.327	13.051	2.174	17.552

Em porcentagem

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

Com relação aos indicadores quantitativos, a análise da Tabela 6 revela que as 17.552 escolas analisadas continham, em média, 342 alunos, distribuídos em 11 turmas. Com 34 professores, a razão média de professor por aluno era de 0,10, aproximadamente, e cada professor era responsável por 1,64 disciplina e 6,22 turmas. Em média, 93,50% dos docentes tinham curso superior completo, mas apenas 46,07% possuíam formação qualificada. Com relação à reprovação e ao abandono, observa-se que as taxas médias eram de 10,60% e 6,87%, respectivamente. Por fim, o número médio de computadores disponíveis para os discentes era de 27,4, o que corresponde a uma razão de 0,08 microcomputador por aluno.

Os resultados mostrados na Tabela 5 evidenciam que as médias dos indicadores qualitativos são coerentes, quando se consideram os níveis da qualidade do ensino médio da rede pública. Um exemplo disso pode ser observado no que se refere à existência de laboratório de ciências. Os resultados revelam que, em relação às escolas de nível baixo com esse tipo de equipamento, o percentual daquelas de nível médio foi 70% maior e o das escolas de alto nível foi 264% superior. No que tange aos indicadores qualitativos (Tabela 6), a mesma tendência foi percebida para qualificação docente (percentuais de docentes com curso superior completo e de docentes com formação qualificada) e porte médio da escola (número de alunos matriculados e total de turmas). Com relação ao número médio de disciplinas por professor e à taxa de abandono, observou-se tendência contrária, ou seja, a diminuição mostrou-se associada ao aumento do nível de qualidade do ensino médio das escolas.

TABELA 6
Número de escolas, média e desvio padrão dos indicadores quantitativos, segundo níveis de qualidade do ensino médio, com base no desempenho médio no Enem Brasil – 2016-2018

Níveis de qualidade do ensino médio e indicadores quantitativos	Escolas	Média	Desvio padrão
Nível baixo de qualidade do ensino médio			
Média de disciplinas por professor	2.327	2,26	1,81
Média de turmas por professor	2.327	6,43	3,02
Percentual de docentes com curso superior completo	2.327	88,65	18,55
Razão professor aluno	2.327	0,13	0,17
Total de turmas	2.327	8,20	8,33
Número de matriculados	2.327	235,24	245,50
Taxa de reprovação (%)	2.311	8,56	8,97
Taxa de abandono (%) – total	2.311	9,24	8,91
Razão de microcomputadores por aluno	2.327	0,09	0,25
Percentual de docentes com formação qualificada	2.327	33,74	19,30
Nível intermediário de qualidade do ensino médio			
Média de disciplinas por professor	13.051	1,57	0,76
Média de turmas por professor	13.051	6,17	2,56
Percentual de docentes com curso superior completo	13.051	94,09	8,91
Razão professor aluno	13.051	0,10	0,08
Total de turmas	13.051	11,28	8,20
Número de matriculados	13.051	352,90	294,40
Taxa de reprovação (%)	13.047	10,95	8,43
Taxa de abandono (%) – total	13.047	6,91	6,64
Razão de microcomputadores por aluno	13.051	0,08	0,14
Percentual de docentes com formação qualificada	13.051	47,55	17,87
Nível alto de qualidade do ensino médio			
Média de disciplinas por professor	2.174	1,43	0,53
Média de turmas por professor	2.174	6,31	2,46
Percentual de docentes com curso superior completo	2.174	95,13	6,89
Razão professor aluno	2.174	0,10	0,08
Total de turmas	2.174	12,45	9,49
Número de matriculados	2.174	393,62	345,87
Taxa de reprovação (%)	2.171	10,63	8,18
Taxa de abandono (%) – total	2.171	4,12	5,28
Razão de microcomputadores por aluno	2.174	0,09	0,14
Percentual de docentes com formação qualificada	2.174	53,21	16,44
Total			
Média de disciplinas por professor	17.552	1,64	0,98
Média de turmas por professor	17.552	6,22	2,62
Percentual de docentes com curso superior completo	17.552	93,50	10,69
Razão professor aluno	17.552	0,10	0,10
Total de turmas	17.552	11,01	8,47
Número de matriculados	17.552	342,34	298,62
Taxa de reprovação (%)	17.529	10,60	8,51
Taxa de abandono (%) – total	17.529	6,87	6,96
Razão de microcomputadores por aluno	17.552	0,08	0,16
Percentual de docentes com formação qualificada	17.552	46,42	18,66

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

Os efeitos dos indicadores na incidência relativa das categorias dos níveis de qualidade do ensino podem ser observados nos resultados do modelo RLM, apresentados na Tabela 7.⁷ O modelo logístico para as escolas públicas de ensino médio regular apresentou grau de ajuste significativo, evidenciado pelo teste de razão de verossimilhança ($p < 0,0001$), assim como significativos foram os efeitos de praticamente todos os indicadores que atuaram na explicação da variável dependente, seja em relação ao submodelo das escolas classificadas com baixo nível, em comparação ao das escolas com alto nível, seja em ambos. A exceção foi o percentual de docentes com nível superior completo. Segundo o modelo, esse indicador não afetou diferenciadamente a razão de chances de ocorrência entre as categorias de análise (baixo nível e alto nível) e a de referência (nível intermediário). Esse resultado corrobora a inexistência de diferenças significativas desse percentual entre as escolas de diferentes níveis de qualidade de ensino médio. Não foram observados, ainda, efeitos significativos dos indicadores existência de laboratório de informática e razão de microcomputadores por alunos para a prevalência do nível baixo de qualidade das escolas de ensino médio. De modo similar, não foram observados efeitos significativos dos indicadores existência de sala de professores, existência de laboratório de informática e existência de quadra de esportes coberta para a maior ocorrência do nível alto de qualidade das escolas de ensino médio.

Na análise das razões de chances dos indicadores, cujos efeitos foram significativos para probabilidade de ocorrência das categorias de análise, deve-se levar em conta seus valores absolutos, uma vez que valores próximos da unidade tende a ter menores efeitos marginais na variação da probabilidade de ocorrência das categorias da variável dependente do RLM. Desse modo, caso significativo, razões de chances maiores do que a unidade sugerem que o indicador possui efeitos positivos e, quanto maior o valor do indicador, maior será o incremento da probabilidade de ocorrência da categoria de análise, em relação à categoria de referência. Em sentido oposto, razões de chances inferiores à unidade sugerem que o indicador possui efeitos negativos e, quanto maior o valor do indicador, menor será o incremento da probabilidade de ocorrência da categoria de análise em relação à categoria de referência. Isso posto, os resultados da Tabela 7 mostram que o nível baixo de qualidade de ensino médio tendeu a ser mais recorrente, quando comparado ao nível intermediário, em escolas rurais, sem sala de professores, sem laboratório de informática, sem quadras de esporte cobertas – ou seja, com baixa infraestrutura; para esse nível também foram observados efeitos positivos dos indicadores média de disciplinas por professor, taxa de abandono total e razão professor por aluno; e efeitos negativos dos indicadores taxa de reprovação e, principalmente, percentual de docentes com formação qualificada. Em relação ao nível alto de qualidade de ensino médio, foi possível constatar, em detrimento ao nível intermediário, que essa condição é mais comum em escolas urbanas, equipadas

⁷ Optou-se, para fins de análise do modelo, em apresentar no corpo do texto apenas parâmetros selecionados da análise de regressão logística multinomial; os demais podem ser conferidos nas tabelas 4 e 5 do Anexo.

com biblioteca e laboratório de ciências, com professores se dedicando a um número menor de disciplinas, lecionando para um número maior de turmas, porém de tamanhos relativamente menores; para esse nível também foram observados efeitos negativos dos indicadores tamanho da escola, taxa de abandono total e taxa de reprovação; bem como efeitos positivos dos indicadores razão de microcomputadores por aluno e, principalmente, percentual de docentes com formação qualificada.

TABELA 7
Razões de chances e intervalos de confiança do modelo de regressão logística multinomial, por nível de qualidade do ensino, segundo indicadores das escolas públicas do ensino médio regular Brasil – 2016-2018

Indicadores	Baixo nível		Alto nível	
	Razão de chances	Intervalo de confiança (95%)	Razão de chances	Intervalo de confiança (95%)
Escola em área urbana	0,430**	(0,377-0,490)	1,483**	(1,189-1,850)
Escola com sala de professores	0,849*	(0,724-0,995)	0,877	(0,674-1,142)
Escola com laboratório de informática	0,683**	(0,601-0,777)	1,046	(0,879-1,244)
Escola com laboratório de ciências	0,957	(0,847-1,080)	2,015**	(1,824-2,225)
Escola com quadra de esportes coberta	0,570**	(0,511-0,637)	1,046	(0,913-1,198)
Escola com biblioteca e/ou sala de leitura	0,606**	(0,531-0,692)	1,463**	(1,176-1,819)
Média de disciplinas do ensino médio por professor	1,152**	(1,081-1,228)	0,783**	(0,679-0,903)
Média de turmas do ensino médio por professor	0,952**	(0,925-0,979)	1,117**	(1,084-1,150)
Percentual de docentes com curso superior completo do ensino médio	1,003	(0,999-1,008)	0,995	(0,989-1,002)
Razão professor por aluno – ensino médio	2,755**	(1,572-4,827)	2,333*	(1,064-5,116)
Total de turmas no ensino médio	0,960**	(0,939-0,983)	1,031**	(1,013-1,050)
Tamanho da escola medido por número de matriculados no ensino médio	1,001*	(1,000-1,002)	0,999**	(0,999-1,000)
Taxa de reprovação no ensino médio – total	0,980**	(0,974-0,986)	0,993*	(0,988-0,999)
Taxa de abandono no ensino médio – total	1,038**	(1,031-1,044)	0,904**	(0,895-0,913)
Razão de microcomputadores por aluno	1,166	(0,824-1,652)	1,674**	(1,161-2,413)
Percentual de docentes com formação qualificada	0,060**	(0,042-0,087)	6,805**	(4,564-10,145)

Fonte: Inep. Exame Nacional do Ensino Médio de 2016, 2017 e 2018; Censo Escolar de 2017.

*p < 0,05; **p < 0,01. Teste de razão de verossimilhança: $\chi^2 = 3840,25$; p < 0,0001.

Nota: Categoria de referência: nível intermediário.

Quando se comparam as chances de a escola ser classificada como baixo ou alto nível de qualidade, alguns fatores chamam a atenção. Estar localizada em área urbana aumenta em 3,4 vezes a chance de a escola ter alto nível de qualidade. Dentre os fatores escolares relacionados ao alto nível de qualidade, destacam-se, também, ter biblioteca ou sala de leitura, que amplia as chances em 2,4 vezes; ter laboratório de ciências (2,1 vezes); ter quadra coberta (1,8 vez); laboratório de informática (1,5 vez) e equipamentos de informática (1,4 vez). Os resultados corroboram os achados de outros estudos, que já indicavam que a melhor infraestrutura contribui para a melhoria da qualidade da escola

(CERQUEIRA; SAWYER, 2007; RIANI; RIOS-NETO, 2008). Apesar de os fatores de infraestrutura serem relevantes, no presente estudo nenhum se mostrou mais significativo do que a qualificação docente, que aumenta em 113,4 vezes a chance de uma escola ser classificada como de alto nível, em relação ao baixo nível de qualidade. Não é novidade a importância da qualificação docente no ensino brasileiro. Riani e Rios-Neto (2008) mostram que essa característica tem importante efeito na qualidade do ensino fundamental. As escolas com maior percentual apresentam melhor desempenho, ou melhor qualidade, segundo as definições desse trabalho. Essa constatação revela a resposta à última pergunta do trabalho.

Considerações finais

O presente estudo procurou responder a questões objetivas sobre o sistema educacional de ensino médio no país, em especial a respeito do sistema público. Foram utilizados os microdados do Censo Escolar 2017 e das edições de 2016, 2017 e 2018 do Exame Nacional do Ensino Médio e aplicadas técnicas de mineração de dados, que permitiram a definição dos indicadores empregados posteriormente na modelagem estatística. Acredita-se que, diante dos resultados e das análises sobre o desempenho escolar, em seus diversos recortes, bem como da seleção dos indicadores-chave e da modelagem da qualidade de ensino das escolas públicas, lograram-se reflexões que vão ao encontro das respostas às indagações desse trabalho.

A utilização da mineração de dados, por meio dos procedimentos do método KDD, possibilitou que se investigassem, no âmbito do sistema educacional de ensino médio no país, relações pouco exploradas, entre uma gama considerável de indicadores escolares, e identificar, dentre os mais relevantes, o mais efetivo para a qualidade desse nível de ensino. Essa pesquisa corrobora as afirmações de Baker, Isotani e de Carvalho (2011), de que a mineração de dados educacionais é um dos campos mais promissores, tanto para a produção de conhecimentos, quanto para o subsídio de políticas públicas na área da educação.

No que tange ao desempenho médio escolar das três classes de ensino médio (regular, técnico e profissionalizante integrado) oferecidas pelos quatro tipos de estabelecimentos escolares no país (federal, estadual municipal e privado), o regular estadual – o mais comum dentre eles – obteve um grau considerável, ficando a apenas 23 pontos do desempenho médio das escolas privadas de ensino técnico. Quando se observa apenas o desempenho escolar da rede pública de ensino médio, surpreendentemente, não se encontram grandes diferenciais interestaduais. Em termos percentuais, verificou-se uma variação de 4,9%, para mais ou para menos, em torno da média nacional, redundando em um diferencial, entre a UF de menor desempenho (Maranhão) e a de maior (Brasília), de apenas 50 pontos. Tais resultados induzem à seguinte afirmação: o ensino médio público regular no Brasil possui uma estrutura básica que é capaz de conferir aos seus alunos um desempenho escolar razoável – entre 465 e 513 pontos no Enem –, desde que eles estejam cursando regularmente o terceiro ano do ensino médio.

Dentro desse escopo, evidenciou-se que, no geral, as escolas de ensino médio regular da rede pública sofrem efeitos de seus indicadores de infraestrutura. A relação desses indicadores com a qualidade do ensino fundamental já havia sido apontada por Cerqueira e Sawyer (2007), Rodrigues, Rios-Neto e Pinto (2011) e Alves e Xavier (2018). A partir dos resultados desse trabalho, pode-se afirmar o mesmo com relação ao ensino médio: o investimento em infraestrutura é um facilitador do melhor desempenho dos alunos.

Análise semelhante pode ser feita no que se refere às taxas de rendimento escolar, frequentemente consideradas quando se quer inferir sobre o desempenho escolar no ensino fundamental (FERNANDES; NATENZON, 2007; ALBERNAZ FERREIRA; FRANCO, 2002, NASCIMENTO, CRUZ-JUNIOR; FAGUNDES, 2018) e no superior (COSTA; CAZELLA; RIGO, 2007). No presente estudo, as taxas de rendimento escolar do ensino médio revelaram importante relação com o rendimento médio dos alunos no Enem. Embora a afirmação possa parecer óbvia, talvez possa-se remeter aos trabalhos que pontuam a importância da formação das turmas para a melhoria do desempenho médio dos alunos (ALVES; SOARES, 2007; LAROS; MARCIANO, 2008; PALERMO; SILVA; NOVELLINO, 2014), ou mesmo a adoção de práticas pedagógicas semelhantes às de escolas de destaque, em termos de desempenho, como apontam Soares e Andrade (2008).

Por fim, dentre os fatores analisados, destacam-se os indicadores vinculados à prática e adequação docentes, que se mostraram eficazes na determinação de maiores níveis de desempenho escolar. Esse resultado encontra respaldo em artigos mais recentes sobre a qualidade do ensino fundamental (FONSECA; NAMEN, 2016; NASCIMENTO, CRUZ-JUNIOR; FAGUNDES, 2018). Levando-se em conta que o grande peso do ensino médio é dado por escolas públicas, é importante ressaltar que cabem muito mais ao poder público do que à escola a promoção da capacitação, a correta alocação – de acordo com a área de especialização – e a criação de condições para que docentes qualificados permaneçam nas escolas públicas.

Referências

- AGAOLU, M. Predicting instructor performance using data mining techniques in higher education. *IEEE*, v. 4, p. 2379-2387, 2016.
- ALVES, M. T. G.; SOARES, J. S. Efeito-escola e estratificação escolar: o impacto da composição de turmas por nível de habilidade dos alunos. *Educação em Revista*, v. 45, p. 25-58, 2007.
- ALVES, M. T. G.; XAVIER, F. P. Indicadores multidimensionais para avaliação da infraestrutura escolar do ensino fundamental. *Cadernos de Pesquisa*, v. 48, n. 169, p. 708-746, 2018.
- ALBERNAZ, A.; FERREIRA, F. H. G.; FRANCO, C. **Qualidade e equidade na educação fundamental brasileira**. Rio de Janeiro: PUC-Rio, 2002. (Texto para discussão, n. 455).
- ANDRADE, R. J.; SOARES, J. F. O efeito da escola básica brasileira. *Estudos em Avaliação Educacional*, v. 19, n. 41, p. 379-406, 2008.
- BAKER, R.; ISOTANI, S.; DE CARVALHO, A. Mineração de dados educacionais: oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 19, n. 2, p. 3-13, 2011.

BARBOSA, M. E. F.; FERNANDES, C. Modelo multinível: uma aplicação a dados de avaliação educacional. **Estudos em Avaliação Educacional**, v. 22, p. 135-154, 2000.

CATERINE, C. V.; LUCE, M. B.; ESPINAR, S. R. O desempenho acadêmico como indicador de qualidade da transição ensino médio-educação superior. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 22, n. 84, p. 635-670, 2014.

CERQUEIRA, C. A.; SAWYER, D. R. O. T. Tipologia dos estabelecimentos escolares brasileiros. **Revista Brasileira de Estudos de População**, v. 24, n. 1, p. 53-67, 2007.

CÉSAR, C. C.; SOARES, J. F. Desigualdades acadêmicas induzidas pelo contexto escolar. **Revista Brasileira de Estudos de População**, v. 18, n. 1, p. 97-110, 2001.

COLEMAN, J. S.; CAMPBELL, E. Q.; HOBSON, C. J.; MCPARTLAND, J.; MOOD, A. M.; WEINFELD, F. D.; YORK, R. L. **Equality of educational opportunity**. Washington, D.C.: U.S. Government Printing Office, 1966.

COSTA, S. S.; CAZELLA, S.; RIGO, S. J. Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: um estudo de caso sobre evasão escolar na UNA-SUS. **Novas Tecnologias na Educação**, v. 12, n. 2, p. 1-11, 2014.

CUNHA, J. M. P.; PEREZ, J. R. R.; AIDAR, T. Proposta metodológica de elaboração de indicador educacional sintético para os municípios. **Revista Brasileira de Estudos de População**, v. 18, n. 1, p. 131-159, 2001.

DUTRA, R. S.; DUTRA, G. B. M.; PARENTE, P. H. N.; PARENTE, L. O. S. S. Determinantes do desempenho educacional dos institutos federais do Brasil no Exame Nacional do Ensino Médio. **Pesquisa e Educação**, v. 45, p. 1-23, 2019.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, 1996.

FERNANDES, R.; NATENZON, P. E. A evolução recente do rendimento escolar das crianças brasileiras: uma reavaliação dos dados do Saeb. **Estudos em Avaliação Educacional**, n. 28, p. 3-22, 2007.

FERNANDEZ, G. C. Discriminant analysis, a powerful classification technique in data mining. *In*: SAS USERS INTERNATIONAL CONFERENCE. **Proceedings** [...]. Paper 247-27, 2002.

FISHER, R. A. **The design of experiments**. Edinburgh: Oliver & Boyd, 1935.

FONSECA, S. O.; NAMEN, A. A. Mineração em bases de dados do INEP: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133-157, 2016.

FREITAS-JÚNIOR, O. G.; RODRIGUES, W. R. M.; BARBIRATO, J. C. C.; BARROS COSTA, E. Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de Maceió. **Novas Tecnologias na Educação**, v. 17, n. 1, p. 296-305, 2019.

GATTI, B. A. Estudos quantitativos em educação. **Educação e Pesquisa**, v. 30, n. 1, p. 11-30, 2004.

HANUSHEK, E. The economics of schooling: production and efficiency in public schools. **Journal of Economic Literature**, v. 24, p. 1141-1177, 1986.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. 2. ed. New York: Wiley, 2000.

IMBERNÓN, F. **Qualidade do ensino e formação do professorado: uma mudança necessária**. São Paulo: Cortez, 2016.

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Resumo técnico:** Censo da Educação Básica 2017. Brasília, 2017. Disponível em: http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf. Acesso em: 02 nov. 2019.

JESUS, G. R. de; LAROS, J. A. Eficácia escolar: regressão multinível com dados de avaliação em larga escala. **Avaliação Psicológica**, v. 3, n. 2, p. 93-106, 2004.

KAMPPF, A. J. C.; REATEGUI, E. B.; LIMA, J. V. Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. **Novas Tecnologias na Educação**, v. 6, n. 2, p. 1-8, 2008.

LAROS, J. A.; MARCIANO, J. L. Índices educacionais associados à proficiência em língua portuguesa: um estudo multinível. **Avaliação Psicológica**, v. 7, n. 3, p. 371-389, 2008.

LARRUBIA, L. M. *et al.* Uma função de produção educacional para o ensino de 2º grau, no Rio Grande do Sul. **Revista do Centro de Ciências Rurais**, v. 6, n. 3, p. 315-322, 1976.

LOURENÇO, R. L.; NASCIMENTO, J. C. H. B.; SAUERBRONN, F. F.; MACEDO, M. A. S. Determinantes sociais e pedagógicos das notas do IDEB. **Revista Pensamento Contemporâneo em Administração**, v. 11, n. 4, p. 27-43, 2017.

MARTELETO, L.; ANDRADE, F. The educational achievement of Brazilian adolescents: cultural capital and the interaction between families and schools. **Sociology of Education**, v. 87, n. 1, p. 16-35, 2007.

NAMEN, A. A.; BORGES, S. X. A.; SADALA, M. G. S. Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas. **Revista Brasileira de Estudos Pedagógicos**, v. 94, n. 238, p. 677-700, 2013.

NASCIMENTO, R. L. S.; CRUZ-JUNIOR, G. G.; FAGUNDES, R. A. A. Mineração de dados educacionais: um estudo sobre indicadores da educação em bases de dados do INEP. **Novas Tecnologias na Educação**, v. 16, n. 1, p. 1-11, 2018.

PALERMO, G. A.; SILVA, D. B. N.; NOVELLINO, M. S. F. 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**, v. 31, n. 2, p. 367-394, 2014.

RIANI, J. L. R.; RIOS-NETO, E. L. G. Análise do dividendo demográfico na matrícula escolar no Brasil numa abordagem hierárquica e hierárquica-espacial. **Revista Brasileira de Estudos de População**, v. 24, n. 1, p. 69-90, 2007.

RIANI, J. L. R.; RIOS-NETO, E. L. G. Background familiar *versus* perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? **Revista Brasileira de Estudos de População**, v. 25, n. 2, p. 251-269, 2008.

RIOS-NETO, E. L. G.; CÉSAR, C. C.; RIANI, J. L. R. Estratificação educacional e progressão escolar por série no Brasil. **Pesquisa e Planejamento Econômico**, v. 32, n. 3, p. 395-416, 2002.

RODRIGUES, C. R.; RIOS-NETO, E. L. G.; PINTO, C. C. X. Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005. **Revista Brasileira de Estudos de População**, v. 28, n. 1, p. 5-36, 2011.

SILVA-JUNIOR, W. S.; GONÇALVES, F. O. Evidências da relação entre a frequência no ensino infantil e o desempenho dos alunos do ensino fundamental público no Brasil. **Revista Brasileira de Estudos de População**, v. 33, n. 2, p. 283-301, 2016.

SMITH, G. Step away from stepwise. **Journal of Big Data**, v. 5, article 32, 2018.

SOARES, J. F.; ALVES, M. T. G.; OLIVEIRA, R. M. O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000. **Estudos em Avaliação Educacional**, v. 24, p. 69-118, 2001.

SOARES, T. M.; FERNANDES, N. S.; NÓBREGA, M. C.; NICOLE, A. C. N. Fatores associados ao abandono escolar no ensino médio público de Minas Gerais. **Educação e Pesquisa**, v. 41, n. 3, p. 757-772, 2015.

SOUSA, C. A.; LIRA-JUNIOR, M. A.; FERREIRA, R. L. C. Avaliação de testes estatísticos de comparações múltiplas de médias. **Revista Ceres**, v. 59, n. 3, p. 350-354, 2012.

TRAVITZKI, R.; CALERO, J.; BOTO, C. What does the National High School Exam (ENEM) tell Brazilian society? **Cepal Review**, v. 113, p. 157-174, 2014.

TRAVITZKI, R. Avaliação da qualidade do Enem 2009 e 2011 com técnicas psicométricas. **Estudos em Avaliação Educacional**, v. 28, n. 67, p. 256-288, 2017.

TRAVITZKI, R.; FERRÃO, M. E.; COUTO, A. P. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: uma visão a partir da análise de dados do ENEM. **Arquivos Analíticos de Políticas Educativas**, v. 24, n. 74, p. 1-36, 2016.

VIANNA, H. M. Avaliações nacionais em larga escala: análises e propostas. **Estudos em Avaliação Educacional**, v. 27, p. 41-76, 2003.

Sobre os autores

Ricardo Alexandrino Garcia é professor associado do Departamento de Geografia da Universidade Federal de Minas Gerais (UFMG).

Eduardo Luiz Gonçalves Rios-Neto é professor titular aposentado do Departamento de Demografia e pesquisador do Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG).

Adriana de Miranda-Ribeiro é professora adjunta do Departamento de Demografia e pesquisadora do Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG).

Endereço para correspondência

Ricardo Alexandrino Garcia

Universidade Federal de Minas Gerais, Instituto de Geociências
Av. Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Eduardo Luiz Gonçalves Rios-Neto

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas
Av. Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Adriana de Miranda-Ribeiro

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas
Av. Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Abstract

School performance, infrastructure and teaching practice effects on secondary education in Brazil

This paper identifies and discusses school factors that most influence the quality of secondary education in Brazil: school performance, infrastructure, and teaching practice. We used data from the School Census and the National High School Exam (ENEM). The data mining technique has raised key characteristics for good school performance. A logistic regression model identified school characteristics that most influence the school performance of public high school students in Brazil. School performance was also investigated according to the type of school and type of education; and interstate differentials, considering the regular public high school system. Results show that students from state schools are at a disadvantage and that those from private and federal schools have similar performance when considering only the type of school. When considering the type of education as well, regular school students from federal schools have the best performance, about 1.3 times higher than that of state schools students. There are no major interstate disparities, although the results point to regional differentiation. Among school factors, teacher qualification proved to be the most impacting factor on school performance.

Keywords: Education quality. School performance. High school. Data mining. ENEM. Brazil.

Resumen

Efectos del desempeño escolar, la infraestructura y la práctica docente sobre la calidad de la educación secundaria en Brasil

El presente artículo identifica y discute los factores escolares que más influyen en la calidad de la educación secundaria en Brasil: desempeño escolar, infraestructura o práctica docente. Para ello se utilizaron datos del censo escolar y del examen nacional de secundaria. La técnica de minería de datos ha planteado características clave para un buen desempeño escolar, y un modelo de regresión logística identificó los factores escolares que más influyen en el desempeño escolar de los estudiantes de la escuela secundaria pública en Brasil. También se investigaron las diferencias de desempeño escolar de los estudiantes de secundaria según el tipo de escuela y el tipo de educación, y diferenciales interestatales, considerando el sistema público de la escuela secundaria regular. Los resultados muestran que los estudiantes de las escuelas estatales están en desventaja y que los de las escuelas privadas y federales tienen un desempeño similar al considerar solo el tipo de escuela, pero al considerar también el tipo de educación los estudiantes de escuelas regulares de las escuelas federales tienen el mejor desempeño, aproximadamente 1,3 veces más alto que el de los estudiantes de las escuelas estatales. No hay grandes disparidades interestatales, aunque los resultados apuntan a una diferenciación regional. Entre los factores escolares, la calificación de los maestros resultó ser el factor más impactante en el desempeño escolar.

Palabras clave: Calidad de la educación. Rendimiento escolar. Educación secundaria. Minería de datos. ENEM. Brasil.

Recebido para publicação em 17/03/2020

Aceito para publicação em 12/05/2021

ANEXO

TABELA 1
Resultados não significativos do teste de Fisher (LSD) para comparações múltiplas de pares de médias amostrais do Enem, por perfil de escola
Brasil – 2016-2018

(I) X perfil	(J) X perfil	Diferença média (I-J)	Erro padrão	Sig.	Intervalo de confiança (95%)	
					Limite inferior	Limite superior
Escolas municipais de ensino profissionalizante (técnico)	Escolas municipais de ensino médio e profissionalizante	-0,34072	0,65589	0,603	-1,6263	0,9448
Escolas municipais de ensino médio e profissionalizante	Escolas municipais de ensino profissionalizante (técnico)	0,34072	0,65589	0,603	-0,9448	1,6263
Escolas estaduais de ensino médio e profissionalizante	Escolas privadas de ensino profissionalizante (técnico)	-0,28065	0,28841	0,33	-0,8459	0,2846
Escolas privadas de ensino profissionalizante (técnico)	Escolas estaduais de ensino médio e profissionalizante	0,28065	0,28841	0,33	-0,2846	0,8459

Fonte: Inep. Censo Escolar 2017; Enem, 2016 a 2018.

TABELA 2
Resultados não significativos do teste de Fisher (LSD) para comparações múltiplas de pares de médias amostrais do Enem
Unidades da federação selecionadas – 2016-2018

(I) X perfil	(J) X perfil	Diferença média (I-J)	Erro padrão	Sig.	Intervalo de confiança (95%)	
					Limite inferior	Limite superior
Minas Gerais	Rio de Janeiro	-0,34072	0,65589	0,603	-1,6263	0,9448
Rio de Janeiro	Minas Gerais	0,34072	0,65589	0,603	-0,9448	1,6263
Ceará	Paraná	-0,28065	0,28841	0,33	-0,8459	0,2846
Paraná	Ceará	0,28065	0,28841	0,33	-0,2846	0,8459

Fonte: Inep. Censo Escolar 2017; Enem, 2016 a 2018.

TABELA 3
Resultado do ajuste do modelo de regressão logística múltipla (Pseudo R² para 16 indicadores)

	Pseudo R ²
Cox and Snell	0,197
Nagelkerke	0,254
McFadden	0,147
N	17529

Fonte: Inep. Censo Escolar 2017; Enem, 2016 a 2018.

TABELA 4
Resultado dos parâmetros estimados do modelo de regressão logística múltipla com 16 indicadores
Brasil – 2016-2018

Níveis de qualidade EM	Indicadores	Parâmetros estimados						Intervalo de confiança (95%)	
		B	Erro padrão	Wald	G.L.	Sig.	Exp(B)	Limite inferior	Limite superior
Baixo (N=2.311)	Intercepto	0,951	0,218	19,12	1	0			
	Escola em área urbana	-0,845	0,066	161,34	1	0	0,430	0,377	0,490
	Escola com sala de professores	-0,164	0,081	4,08	1	0,043	0,849	0,724	0,995
	Escola com laboratório de informática	-0,381	0,066	33,58	1	0	0,683	0,601	0,777
	Escola com laboratório de ciências	-0,044	0,062	0,51	1	0,474	0,957	0,847	1,080
	Escola com quadra de esportes coberta	-0,561	0,057	98,39	1	0	0,570	0,511	0,637
	Escola com biblioteca e/ou sala de leitura	-0,501	0,068	54,60	1	0	0,606	0,531	0,692
	Média de disciplinas do EM por professor	0,142	0,032	19,07	1	0	1,152	1,081	1,228
	Média de turmas do EM por professor	-0,050	0,014	11,78	1	0,001	0,952	0,925	0,979
	Percentual de docentes EM com superior completo	0,003	0,002	2,86	1	0,091	1,003	0,999	1,008
	Razão professor por aluno – ensino médio	1,013	0,286	12,54	1	0	2,755	1,572	4,827
	Total de turmas no ensino médio	-0,041	0,012	12,02	1	0,001	0,960	0,939	0,983
	Tamanho da escola medido por número de matriculados no ensino médio	0,001	0,000	6,43	1	0,011	1,001	1,000	1,002
	Taxa de reprovação no ensino médio – total	-0,020	0,003	40,29	1	0	0,980	0,974	0,986
	Taxa de abandono no ensino médio – total	0,037	0,003	136,01	1	0	1,038	1,031	1,044
	Razão de microcomputadores por aluno	0,154	0,177	0,75	1	0,386	1,166	0,824	1,652
Percentual de docentes com formação qualificada	-2,811	0,187	225,72	1	0	0,060	0,042	0,087	

(Continua)

(Continuação)

Níveis de qualidade EM	Indicadores	Parâmetros estimados							Intervalo de confiança (95%)	
		B	Erro padrão	Wald	G.L.	Sig.	Exp(B)	Limite inferior	Limite superior	
Alto (N=2.171)	Intercepto	-3,228	0,378	72,89	1	0				
	Escola em área urbana	0,394	0,113	12,22	1	0	1,483	1,189	1,850	
	Escola com sala de professores	-0,131	0,135	0,95	1	0,329	0,877	0,674	1,142	
	Escola com laboratório de informática	0,045	0,089	0,25	1	0,614	1,046	0,879	1,244	
	Escola com laboratório de ciências	0,700	0,051	191,06	1	0	2,015	1,824	2,225	
	Escola com quadra de esportes coberta	0,045	0,069	0,41	1	0,52	1,046	0,913	1,198	
	Escola com biblioteca e/ou sala de leitura	0,380	0,111	11,67	1	0,001	1,463	1,176	1,819	
	Média de disciplinas do EM por professor	-0,245	0,073	11,38	1	0,001	0,783	0,679	0,903	
	Média de turmas do EM por professor	0,110	0,015	53,67	1	0	1,117	1,084	1,150	
	Percentual de docentes EM com superior completo	-0,005	0,004	1,77	1	0,183	0,995	0,989	1,002	
	Razão professor por aluno – ensino médio	0,847	0,401	4,47	1	0,034	2,333	1,064	5,116	
	Total de turmas no ensino médio	0,030	0,009	10,96	1	0,001	1,031	1,013	1,050	
	Tamanho da escola medido por número de matriculados no ensino médio	-0,001	0,000	8,43	1	0,004	0,999	0,999	1,000	
	Taxa de reprovação no ensino médio – total	-0,007	0,003	5,11	1	0,024	0,993	0,988	0,999	
	Taxa de abandono no ensino médio – total	-0,101	0,005	378,77	1	0	0,904	0,895	0,913	
	Razão de microcomputadores por aluno	0,515	0,187	7,61	1	0,006	1,674	1,161	2,413	
	Percentual de docentes com formação qualificada	1,918	0,204	88,53	1	0	6,805	4,564	10,145	

Fonte: Inep: Censo Escolar 2017; Enem, 2016 a 2018.