

# Cor/Raça nas bases educacionais brasileiras: uma proposta metodológica para redução de dados faltantes

Joyce S. R. Petrus<sup>1</sup>

Universidade Federal de Minas Gerais  
Belo Horizonte, MG, Brasil

[joyceufmg@gmail.com](mailto:joyceufmg@gmail.com)

Flavia P. Xavier

Universidade Federal de Minas Gerais  
Belo Horizonte, MG, Brasil

[flaviapx@ufmg.com](mailto:flaviapx@ufmg.com)

Valéria C. de Oliveira

Universidade Federal de Minas Gerais  
Belo Horizonte, MG, Brasil

[valcrisoli@ufmg.br](mailto:valcrisoli@ufmg.br)

## Resumo

Este estudo analisa a qualidade da variável cor/raça e propõe uma metodologia para reduzir o percentual de estudantes sem declaração racial nas principais bases de dados educacionais brasileiras, o Censo Escolar e o Sistema de Avaliação da Educação Básica (Saeb), conduzidos pelo Inep. Diante da relevância dessa variável para compreender as desigualdades educacionais, especialmente as raciais, o estudo combina informações dessas duas fontes para reduzir o percentual de não declaração, que chega a 30%. A estratégia prioriza a autodeclaração feita no Saeb e complementa com dados do Censo Escolar nos casos de omissão. O procedimento permitiu recuperar quase 80% das informações ausentes, reduzindo a taxa de não declaração de 30,2% para 7,0%. Os resultados reforçam a importância de qualificar os dados sobre cor/raça para subsidiar políticas educacionais mais equitativas, especialmente diante da condicionalidade do Valor Aluno Ano por Resultado (VAAR).

Palavras-chave: cor; raça; desigualdades educacionais; dados faltantes; bases educacionais.

---

<sup>1</sup> A autora agradece à CAPES.

## 1 Introdução

A variável cor/raça é um dos marcadores mais relevantes para a análise das desigualdades educacionais no Brasil. Inúmeras pesquisas evidenciam disparidades sistemáticas nos resultados educacionais segundo grupos raciais (ALVES; SOARES; XAVIER, 2016; CARNOY; RODRIGUES, 2024; BOTELHO; MADEIRA; RANGEL, 2015; SOARES; ALVES; FONSECA, 2021; CARVALHO, 2005). Parte da literatura também discute as tensões entre autoidentificação e heteroidentificação étnico-racial (OSÓRIO, 2003; BASTOS; TERRA; NUNES, 2014; ROCHA; ROSEMBERG, 2007). Embora o preconceito de marca seja um traço significativo no Brasil, o que destacaria a centralidade da heteroidentificação, as experiências sociais e subjetivas dos indivíduos também desempenham um papel fundamental na definição da identidade racial, reforçando a importância da autodeclaração. Segundo Bastos, Terra e Nunes (2014), “a autoidentificação está pautada muito mais na questão social e econômica do que em atributos de pigmentação da pele ou o fato de ter cabelo crespo”.

Osório (2003), a partir de dados da pesquisa Datafolha (1995), da Pesquisa Nacional de Demografia e Saúde (PNDS, 1996) chega a dois achados centrais: (i) há alta convergência entre auto e heteroidentificação, embora persista uma tendência de embranquecimento nos registros feitos por terceiros; e (ii) tanto a auto quanto a heteroidentificação são processos subjetivos, e optar por um deles implica escolher entre a percepção do próprio indivíduo e a de um observador externo. Ademais, Rocha e Rosemberg (2007) demonstraram que a partir dos nove anos de idade os estudantes são plenamente capazes de responder ao quesito cor/raça, reforçando a qualidade dos dados provenientes da autodeclaração, como ocorre nos questionários contextuais do Saeb.

Esse debate não é meramente conceitual ou teórico: ele se materializa diretamente na produção das estatísticas educacionais, afetando a qualidade dos dados sobre cor/raça e a comparabilidade entre diferentes bases. As escolhas metodológicas sobre como coletar e tratar essa variável — se pela autodeclaração, pela heteroidentificação ou pela combinação de fontes — não são neutras, pois impactam tanto o diagnóstico das desigualdades quanto o monitoramento de políticas públicas. Nos últimos anos, isso se tornou ainda mais central diante da implementação do Valor Aluno Ano por Resultado (VAAR), mecanismo previsto no novo Fundeb, que condiciona o repasse de recursos federais adicionais à melhoria dos indicadores educacionais e à redução de desigualdades, inclusive as de raça/cor (BRASIL, 2020).

Contudo, as principais bases censitárias da educação básica ainda apresentam elevados percentuais de registros sem preenchimento da variável cor/raça, além de divergências entre fontes e ao longo do tempo, como o fenômeno da reclassificação racial. Tais limitações impõem desafios importantes para o monitoramento de políticas públicas voltadas à equidade. Diante desse cenário, este trabalho propõe uma estratégia metodológica de reconstrução dessa variável, combinando informações do Censo Escolar e do Saeb, com vistas à redução do percentual de dados faltantes e à qualificação da informação disponível para análise de desigualdades educacionais.

## 2 Metodologia

A pesquisa utilizou os microdados identificados do Sistema de Avaliação da Educação Básica (Saeb) e do Censo Escolar, disponibilizados pelo Inep por meio do Serviço de Acesso a Dados Protegidos (Sedap), de 2013 a 2023. Foram analisadas seis edições do Saeb e onze edições do Censo, totalizando 38.627.634 estudantes distintos, com registros relativos ao 5º e 9º anos do EF e à 3ª série (ou 4ª) do EM.

Considerando que no Saeb, 30% dos estudantes (avaliados ou não) não possuem cor/raça declarada, o principal objetivo de trazer os dados do Censo Escolar, foi o de recuperar informações sobre a cor/raça declarada, tornando os estudos e as políticas públicas baseadas em evidências, mais precisas. Para a construção da variável combinada, foi adotada uma hierarquia de fontes e anos, priorizando a autodeclaração registrada no Saeb (do mais recente para o mais antigo). Na ausência da informação, recorreu-se ao dado mais recente disponível no Censo Escolar, também de forma retroativa (de 2023 a 2013).

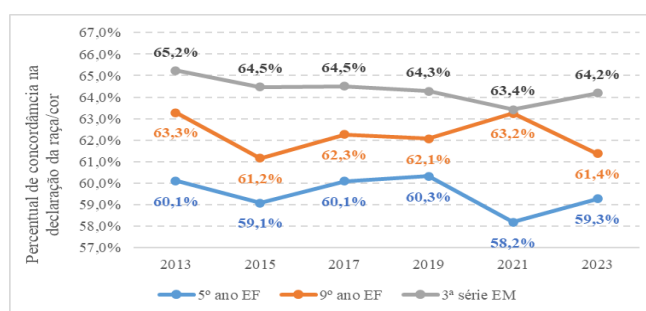
A estratégia visou garantir que cada estudante tivesse uma única declaração de cor/raça associada ao seu registro. O procedimento permitiu recuperar informações ausentes, reduzir inconsistências e construir uma variável mais estável ao longo do tempo. Além da reconstrução da variável, foram realizadas duas análises complementares: (i) uma comparação entre as declarações de cor/raça nas duas fontes (Saeb e Censo Escolar), no mesmo ano e para os mesmos estudantes, a fim de verificar a concordância ou divergência entre as bases; (ii) uma análise longitudinal com dados do Saeb, em que observamos a reclassificação racial, acompanhando quatro coortes de estudantes entre o 5º e o 9º ano do EF.

### 3 Resultados

#### 3.1 Divergências entre pesquisas e entre anos

As análises evidenciaram dois padrões de inconsistência na variável cor/raça: (1) divergências entre o que é declarado no Saeb e no Censo para os mesmos estudantes no mesmo ano; (2) reclassificações entre diferentes anos do Saeb, dos mesmos estudantes. No primeiro caso, a proporção de estudantes pretos é sistematicamente maior no Saeb (autodeclaração), 12,6%, do que no Censo Escolar (heteroidentificação ou declaração de terceiros), 4,9%. Já a proporção de estudantes brancos é sempre superior no Censo Escolar, sugerindo um viés sistemático de embranquecimento. A discordância média entre as fontes gira em torno de 38%, variando por etapa e edição (ver gráfico 1).

**Gráfico 1: Percentual de concordância entre declaração de cor/raça do Censo e do Saeb, por ano letivo (apenas estudantes que se declararam em ambas as bases)**

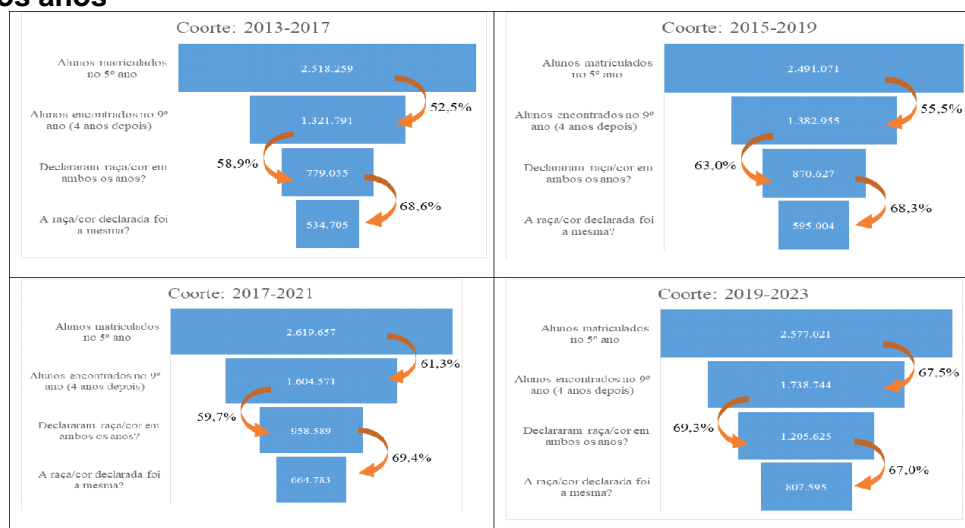


Fonte: Elaboração própria a partir dos microdados do Censo Escolar e Saeb/Sedap/Inep, 2013 a 2023.

Nota: A escala do eixo vertical (eixo y) foi alterada para facilitar a leitura do gráfico.

No segundo caso, analisando quatro coortes longitudinais entre 2013 e 2023, verificou-se que cerca de 32% dos estudantes mudaram sua declaração racial entre o 5º e o 9º ano do EF. As taxas de reclassificação se mantêm relativamente estáveis nas quatro coortes analisadas (ver figura 1). O fenômeno pode refletir tanto mudanças de identidade quanto pressões contextuais de identificação.

**Figura 1: Estudantes do Saeb no 5º ano do Ensino Fundamental, encontrados 4 anos depois, no 9º ano do Ensino Fundamental, com informação de cor/raça declarada em ambos os anos**

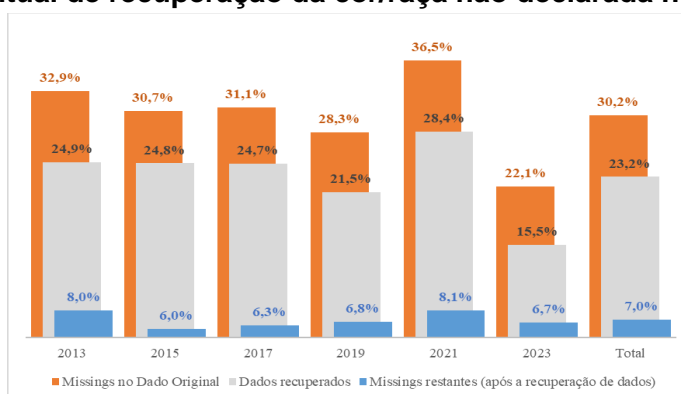


Fonte: Elaboração própria a partir dos microdados do Saeb/Sedap/Inep, 2013 e 2023. Nota: trata-se dos mesmos estudantes identificados em ambas as etapas escolares, com distância de 4 anos.

### 3.2 Recuperação dos dados faltantes

A aplicação da metodologia reduziu expressivamente o percentual de não declaração no Saeb, de 30,2% para 7,0%, mais de 8,97 milhões de registros (ver gráfico 2).

**Gráfico 2: Percentual de recuperação da cor/raça não declarada no Saeb, 2013- 2023**



Fonte: Elaboração própria a partir dos microdados do Censo Escolar e Saeb/Sedap/Inep, 2013 a 2023.

### 3.3 Confiabilidade e validade da nova variável

A nova variável apresenta maior confiabilidade, ao reduzir discrepâncias entre bases e ao longo do tempo, permitindo um retrato mais estável da identificação racial dos estudantes. Além disso, os resultados indicam ganhos de validade em duas dimensões: validade de construto, por priorizar a autodeclaração, aproximando-se

das definições normativas de pertencimento racial e da distribuição demográfica observada no Censo de 2022; e validade preditiva, pois análises exploratórias mostram que a nova variável demonstra as diferenças de desempenho entre grupos raciais nos testes do Saeb.

## **4 Discussão**

Ao priorizar a autodeclaração na construção da variável cor/raça, os resultados obtidos reforçam o entendimento de que o método de coleta não é um detalhe técnico, mas uma dimensão central na produção das políticas públicas que se apoiam nesses dados. O achado reforça os argumentos de Bastos, Terra e Nunes (2014), para quem a autodeclaração não apenas capta a identidade fenotípica, mas também as vivências sociais e econômicas dos indivíduos. A tendência de sub-representação de estudantes pretos e, em menor grau, pardos nos registros do Censo Escolar, quando comparada aos dados do Saeb, dialoga diretamente com os achados de Osório (2003) e Rocha e Rosemberg (2007). Ademais, observa-se que, embora exista uma convergência razoável entre auto e heteroidentificação, persiste uma tendência estrutural de embranquecimento quando a classificação é realizada por terceiros.

A proposta metodológica apresentada constitui uma contribuição relevante tanto para o campo acadêmico quanto para a produção oficial de estatísticas educacionais. A hierarquização das fontes — priorizando a autodeclaração no Saeb e, na sua ausência, utilizando o dado mais recente do Censo Escolar — oferece uma solução técnica viável, transparente e replicável, capaz de melhorar significativamente a integralidade e a validade da variável cor/raça nas bases do Inep.

## **5 Conclusões e Considerações Finais**

A proposta metodológica desenvolvida neste trabalho contribui de forma significativa para qualificar a variável cor/raça nas bases da educação básica, enfrentando dois desafios centrais: o elevado percentual de não declaração e as divergências entre registros provenientes de diferentes fontes. Ao priorizar a autodeclaração do Saeb e complementar com informações do Censo Escolar, a estratégia apresentou resultados expressivos na redução dos registros sem informação, diminuindo a taxa de não declaração de 30,2% para 7,0%, ou seja, uma recuperação de quase 9 milhões de registros entre 2013 e 2023.

Do ponto de vista metodológico, a proposta tem algumas vantagens: é simples, transparente, replicável e pode ser incorporada institucionalmente tanto pelo Inep, quanto pelas redes de ensino. Além de qualificar as análises acadêmicas, sua adoção fortalece a capacidade do Estado de monitorar as desigualdades raciais na educação, o que ganha especial relevância diante da implementação do VAAR e de outras políticas que condicionam o financiamento à redução das desigualdades.

Entretanto, é importante reconhecer que a metodologia não resolve todos os problemas associados à variável cor/raça. Persistem limitações, especialmente nos casos em que não há qualquer registro da variável nas duas bases utilizadas, além de não eliminar o fenômeno da reclassificação racial, que permanece em patamares elevados ao longo da trajetória escolar dos estudantes.

Apesar desses limites, a metodologia trazida representa um avanço, na medida em que melhora tanto a cobertura quanto a consistência da variável cor/raça, ampliando sua validade para análises sobre desigualdades educacionais. Além disso, abre-se uma agenda de pesquisa aplicada que inclui tanto a expansão da análise para outras bases, como o Enem e o Censo da Educação Superior, quanto a aplicação complementar de métodos de imputação de dados.

## 6 Referências

ALVES, M. T. G.; SOARES, J. F.; XAVIER, F. P. Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais. *Revista Brasileira de Sociologia*, São Cristóvão, v. 4, n. 7, p. 49-81, 2016. Disponível em:

<https://rbs.sbsociologia.com.br/index.php/rbs/article/view/181/84>. Acesso em: 3 mar. 2023.

BASTOS, T. G.; TERRA, A. D. G.; NUNES, L. C. E. S. Hetero ou autoidentificação: qual melhor mecanismo para utilização nas cotas? *Congresso Internacional Interdisciplinar em Sociais e Humanidades – CONINTER 3*, Salvador, n. 3, v. 17, p. 362-379, 2014.

BOTELHO, F.; MADEIRA, R. A.; RANGEL, M. A. Racial discrimination in grading: evidence from Brazil. *American Economic Journal: Applied Economics*, v. 7, n. 4, p. 37-52, 2015. Disponível em: <https://doi.org/10.1257/app.20140352>. Acesso em: 29 mar. 2024.

BRASIL. Lei nº 14.113, de 25 de dezembro de 2020. Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), de que trata o art. 212-A da Constituição Federal; revoga dispositivos da Lei nº 11.494, de 20 de junho de 2007; e dá outras providências. *Diário Oficial da União: seção 1*, Brasília, DF, 25 dez. 2020.

CARNOY, M.; RODRIGUES, E. C. Achievement gains and social inequality: an analysis of Brazilian school districts, 2007-2017. *International Journal of Educational Development*, v. 107, 2024. Disponível em: <https://doi.org/10.1016/j.ijedudev.2024.103049>. Acesso em: 29 mar. 2024.

CARVALHO, M. P. de. Quem é negro, quem é branco: desempenho escolar e classificação racial de alunos. *Revista Brasileira de Educação*, n. 28, p. 77-95, 2005. Disponível em: <https://doi.org/10.1590/S1413-24782005000100007>. Acesso em: 29 nov. 2024.

OSÓRIO, R. G. O sistema classificatório de cor ou raça do IBGE. Brasília, DF: Ipea, 2003. (Texto para Discussão, n. 996). Disponível em: <https://repositorio.ipea.gov.br/handle/11058/2958>. Acesso em: 29 nov. 2024.

ROCHA, E. J.; ROSEMBERG, F. Autodeclaração de cor e/ou raça entre escolares paulistanos(as). *Cadernos de Pesquisa*, São Paulo, v. 37, n. 132, p. 759-799, set./dez. 2007. Disponível em: <https://www.scielo.br/j/cp/a/5NbCwJ6ShDZ6sq9FPgpBRxc/abstract/?lang=pt>. Acesso em: 29 nov. 2024.

SOARES, J. F.; ALVES, M. T. G.; FONSECA, J. A. Trajetórias educacionais como evidência da qualidade da educação básica brasileira. *Revista Brasileira de Estudos de População*, v. 38, p. e0167, 2021.