Supporting Information for

Racial classification during childhood and adolescence and its implications for education statistics

Guilherme Lichand, Lucas Klotz, Carlos A Dória, Lêticia Lopes, Elizabeth Kozleski, and Jason D Yeatman

Guilherme Lichand.
E-mail: glichand@stanford.edu

This PDF file includes:

¹² **Adapted questions in race survey.** Beginning in April 2024, Equidade.info's survey instrument incorporated a supporting
¹³ feature for students who struggled to understand the color categories in the race question, whereby surveyors present images
¹⁴ to help them decide which race or color they identified with. Concretely, students who answer *I don't know, I prefer to see*
¹⁵ *images to decide* are shown AI-generated families for each race category. Respondents then select the figure with which they
¹⁶ identify the most. Those who do not want to see supporting images can simply answer *I don't know.* Figure S1 illustrates the
instrument with AI-generated images of white, black, and yellow families.

(a) White Family

(b) Black Family

(c) Yellow Family

Fig. S1. Supporting images for adapted question in race survey, by category

Notes: Examples of AI-generated images for the adapted question in the race survey. Panel (a) shows the image of a white family; Panel (b), of a black family; and Panel (c), of a yellow family. Survey instrument from Equidade.info (April 2024 to March 2025).

¹⁷

Guilherme Lichand, Lucas Klotz, Carlos A Dória, Lêticia Lopes, Elizabeth Kozleski, and Jason D Yeatman

Adapted questions in disability survey. When it comes to sensory disabilities, enumerators asked students if they had difficulty with the following: *hearing the teacher speak,* **even when close to them and with the rest of the class silent**; and *seeing what is written on the board or in their notebook,* **even when close to it or wearing glasses**. Regarding motor disabilities, enumerators asked students if they had difficulty with *walking, running, jumping, writing, or holding objects.* Students could respond with the following options: *No difficulty*; *Yes, some difficulty*; *Yes, a lot of difficulty*; *Yes, cannot do it at all.*

In the case that the enumerator felt the student did not understand the question, they would select a fifth option: *I did not understand, I prefer to see pictures/images to decide.* Figure S2 shows the images depicting each disability category. In each case, students chose either the imagine that meant 'Yes', *I experience this difficulty* or 'No', *I do not experience this difficulty.* Those who answered 'Yes' were classified as having the corresponding disability. Less than ten students were assessed based on the adapted questionnaire.

(a) Visual disability

(b) Hearing disability

(c) Motor disability
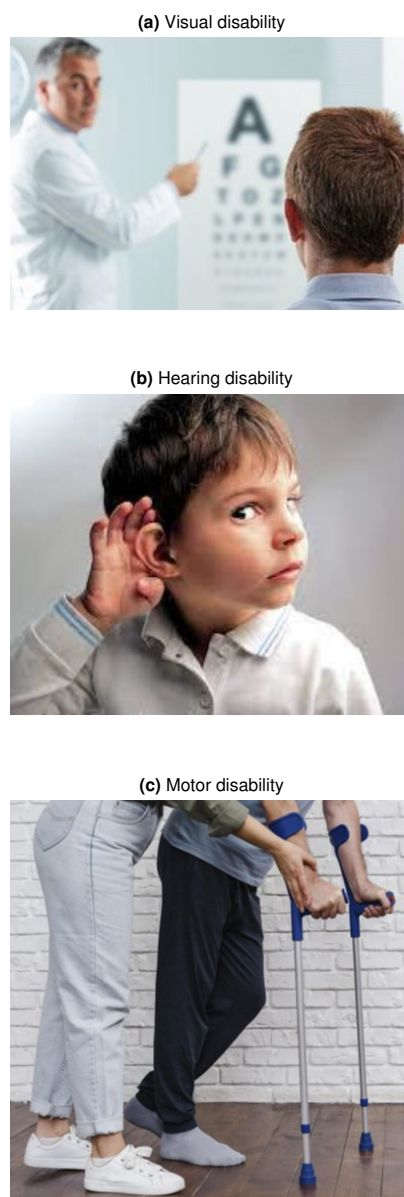
Fig. S2. Supporting images for adapted questions in disability survey, by category

Notes: Images that illustrate different disability categories in the adapted questionnaire. Panel (a) shows the image for experiencing visual difficulties; Panel (b), hearing difficulties; and Panel (c), motor difficulties. Survey instrument from Equidade.info (April and May 2024).

28 **Portuguese word clouds.** . Figure S3 displays word clouds based on the original answers (in Brazilian Portuguese) to the
29 open-ended self-reported race question by education level. The number of unique answers in each case is higher than the one
30 reported in the main text because some of the terms used to describe race or color can vary by gender in Brazilian Portuguese.

(a) Elementary school

(b) Middle school

(c) High school

Fig. S3. Word clouds for the answers to the open-ended self-reported race (in original Brazilian Portuguese)

Notes: Panel (a) shows word cloud for elementary school students, comprising 121 unique answers from 1101 students; Panel (b), that for middle school students, comprising 45 unique answers from 807 students; Panel (c), that for high school students, comprising 28 unique answers from 503 students. Word cloudsgenerated using an online tool (see https://www.wordclouds.com/). Data from students' open-ended self-reported race question from 2,412 students surveyed by Equidade.info (Jun-Jul 2023 and Feb-Mar 2025).

Guilherme Lichand, Lucas Klotz, Carlos A Dória, Lêticia Lopes, Elizabeth Kozleski, and Jason D Yeatman

Additional results for Study 1: Discrepancies between school- and self-reported race. We document the joint distribution of students' self-reported race and that reported by schools, spanning grades K-12, without sample weights (Tables S1 and S2). The diagonal (in blue) features the share of observations with matching race classifications; off-diagonal cells (in red) feature the share of observations with inconsistent classifications in each case.

Table S3 displays the share of students who report their race classification as 'other', 'do not know', or 'do not want to declare' – we do not include them in the main analyses as schools were constrained to report student race as one of the five categories in the table. We compute the proportions on the aggregated and by education level. The 'Total' column shows the total percentage of students in these unclassified categories considering the whole sample, and not only the ones who did not self-report one of the five categories shown in the main text.

Table S1. School-reported race (rows) vs. Self-reported race (columns), simple averages

	White	Black	Brown	Yellow	Indigenous	Total
White	**28.82%**	0.58%	10.93%	2.25%	0.30%	**42.88%**
Black	0.44%	**11.58%**	5.91%	0.24%	0.59%	**18.77%**
Brown	3.78%	3.47%	**22.50%**	0.99%	0.85%	**31.59%**
Yellow	0.81%	0.22%	1.59%	**0.61%**	0.08%	**3.30%**
Indigenous	0.07%	0.12%	0.22%	0.06%	**2.98%**	**3.45%**
Total	**33.92%**	**15.97%**	**41.15%**	**4.15%**	**4.80%**	**100%**

Notes: Joint distribution of students' self-reported race (columns) and that reported by schools (rows), spanning grades K-12. The shares of *Don't Know*, *Other*, and *Do not want to declare* were not included as schools were constrained to report student race as one of the five categories in the table (see Table S3). Share of matching classification by race shown in blue; mismatches shown in red. Data from 20,206 students across all 27 Brazilian States.

Table S2. School-reported race (rows) vs. Self-reports (columns), simple averages by education level

(a) Elementary school

	White	Black	Brown	Yellow	Indigenous	Total
White	**28.41%**	0.91%	10.28%	2.96%	0.38%	**42.96%**
Black	0.78%	**10.57%**	5.47%	0.33%	0.67%	**17.83%**
Brown	4.84%	4.39%	**21.31%**	1.41%	1.14%	**33.09%**
Yellow	0.94%	0.24%	1.40%	**0.63%**	0.10%	**3.32%**
Indigenous	0.10%	0.12%	0.12%	0.07%	**2.37%**	**2.80%**
Total	**35.07%**	**16.23%**	**38.58%**	**5.40%**	**4.66%**	**100%**

(b) Middle school

	White	Black	Brown	Yellow	Indigenous	Total
White	**27.38%**	0.21%	12.26%	1.80%	0.33%	**41.98%**
Black	0.19%	**13.26%**	6.89%	0.14%	0.76%	**21.24%**
Brown	2.69%	2.21%	**23.85%**	0.63%	0.74%	**30.12%**
Yellow	0.62%	0.22%	2.02%	**0.46%**	0.08%	**3.40%**
Indigenous	0.03%	0.19%	0.35%	0.03%	**2.67%**	**3.27%**
Total	**30.91%**	**16.09%**	**45.37%**	**3.06%**	**4.58%**	**100%**

(c) High school

	White	Black	Brown	Yellow	Indigenous	Total
White	**31.94%**	0.40%	10.39%	1.28%	0.05%	**44.06%**
Black	0.05%	**11.39%**	5.46%	0.19%	0.14%	**17.22%**
Brown	3.00%	3.28%	**23.24%**	0.54%	0.33%	**30.40%**
Yellow	0.78%	0.17%	1.37%	**0.78%**	0.02%	**3.12%**
Indigenous	0.05%	0.02%	0.24%	0.07%	**4.82%**	**5.20%**
Total	**35.82%**	**15.26%**	**40.70%**	**2.86%**	**5.36%**	**100%**

Notes: Joint distribution of students' self-reported race (columns) and that reported by schools (rows). Panel (a) refers to elementary school (K–5), Panel (b) to middle school (grades 6–9), and Panel (c) to high school (grades 10–12). The shares of *Don't Know*, *Other*, and *Do not want to declare* were not included as schools were constrained to report student race as one of the five categories in the table (see Table S3). Entries on the main diagonal (in blue) indicate consistent classification across sources. Off-diagonal entries (in red) reflect mismatches. Data from 20,206 students across all 27 Brazilian States.

Table S3. Share of unclassified race categories

(a) Weighted averages

Group	Other	Not Declared	Don't Know	Total
All Education Levels	1.68%	1.54%	10.07%	**13.29%**
Elementary School	2.25%	2.44%	14.39%	**19.08%**
Middle School	0.66%	0.84%	5.63%	**7.13%**
High School	0.44%	0.29%	1.57%	**2.30%**

(b) Simple averages

Group	Other	Not Declared	Don't Know	Total
All Education Levels	0.97%	1.43%	9.85%	**12.25%**
Elementary School	1.31%	2.27%	14.95%	**18.53%**
Middle School	0.85%	0.74%	5.97%	**7.56%**
High School	0.25%	0.23%	2.05%	**2.53%**

Notes: Share of students who report their race classification as 'other', 'do not know', or 'do not want to declare', on aggregate and by education level. Each row represents a group, showing the share of students in each category. Total column shows the total percentage of students in these unclassified categories related to whole sample. Panel (a) shows weighted averages, while Panel (b) displays simple averages. Elementary school refers to K–5 students; middle school is grades 6 to 9; and high school is grades 10 to 12. Weighted averages based on sample weights provided by Equidade.info, computed based on student and school characteristics to match the universe of students and schools in the 2023 school census. Data from 23,028 students across all 27 Brazilian States.

Guilherme Lichand, Lucas Klotz, Carlos A Dória, Lêticia Lopes, Elizabeth Kozleski, and Jason D Yeatman

40 **Additional results for Study 2.** Table S4 shows the regression results discussed in the main text.

Table S4. Effect of political incumbency (2016 election) on student racial mix (9th graders in 2019 vs. 5th graders in 2015)

	Panel (a) Absolute change in black enrollment based on school census (school reports)	Panel (b) Absolute change in black enrollment based on SAEB (self-reports)
Incumbent Reelected	-0.005**	-0.004
	(0.002)	(0.005)
N (Left / Right)	6,665 / 5,603	2,233 / 2,034
Effective Obs. (Left / Right)	4,429 / 3,120	1,184 / 840
Optimal bandwidth	0.152	0.102
Robust 95% CI	$[-0.010, 0.000]$	$[-0.017, 0.004]$
% black students (5th graders in 2015)	0.025	0.085

Notes: Regression discontinuity estimates of incumbent political party win in close elections on the change in the municipal share of black students between 2015 (5th graders) and 2019 (9th graders). Panel (a) computes that share based on school-reported race, using school census data. Panel (b) computes that share based on self-reported race, using Prova Brasil data. Robust p-values and confidence intervals based on bias-corrected standard errors. Optimal bandwidth estimated following (1).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

41 Panel (a) documents that mayoral reelection causally decreases noise in the municipal black enrollment share based on
42 school reports. The effect size is large – 20% of the baseline black enrollment share among 5th graders in 2015 – and precisely
43 estimated (p < 0.10). In contrast, Panel (b) documents no causal effect of reelection on that share when based on student
44 self-reports. The effect size is not only imprecisely estimated, but effectively much smaller than that in Panel (a) – only 4.7%
45 of the baseline black enrollment share among 5th graders in 2015. The reason is that the share of students who self-report to
46 be black is much larger than that in school reports (8.5% instead of 2.5%), consistent with Equidade.info survey data.

47 ***Remarks on sampling, SAEB's participation eligibility and results' public release.*** This subsection aims to clarify the sample size used in
48 the analyses of Study 2. In addition to the municipality-level restriction related to close elections in 2016, other factors further
49 reduce the sample. These restrictions help explain the drop in the number of observations when linking race in the Brazilian
50 school census and in Prova Brasil (in the context of SAEB).
51 To analyze school-level data for the 2015–2019 cohort, we restrict our sample to schools that offer both 5th and 9th grades.
52 This requirement introduces an additional constraint, as it is common for a school to offer only elementary education (which
53 includes 5th grade) but not secondary education (which includes 9th grade). Indeed, looking at the 2015 school census data,
54 overall there were 119,913 schools offering elementary education, 62,644 offering secondary education, and only 46,648 offering
55 both.
56 Moreover, not all schools are eligible to participate in SAEB (2), which imposes another layer of restriction. Specifically, for
57 the composition of SAEB's reference population, the following are excluded:

58 I – Schools with fewer than 10 students enrolled in elementary or high school levels;

59 II – Multi-grade classrooms;

60 III – Remedial education classes (age-grade distortion correction);

61 IV – Youth and Adult Education (EJA) classes;

62 V – Secondary-level teacher training programs (Ensino Médio Normal/Magistério);

63 VI – Special education classes, schools, or services not integrated into regular education;

64 VII – Indigenous schools that do not offer Portuguese as the first language of instruction.

65 Lastly, even if a school participates in SAEB, its results are not always publicly available. For results to be disclosed, the
66 school must: i) register at least 10 students from the assessed grade level present at the time of the assessment; and ii) achieve
67 a participation rate of at least 80% of enrolled students in the assessed grade level, based on the data reported by the school to
68 the 2019 school census, using the final enrollment data and excluding students who transferred during the period. An entire
69 municipality may also be excluded from public disclosure if it fails to meet these same criteria.

References

1. S Calonico, MD Cattaneo, MH Farrell, R Titiunik, rdrobust: Software for regression discontinuity designs. *Stata J.* **17**, 372–404 (2017).
2. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Portaria nº 366, de 29 de abril de 2019 (Diário Oficial da União) (2019) Estabelece diretrizes, normas e procedimentos para a realização do SAEB 2019.

Guilherme Lichand, Lucas Klotz, Carlos A Dória, Lêticia Lopes, Elizabeth Kozleski, and Jason D Yeatman