AER: Insights 2024, 6(1): 20–37
https://doi.org/10.1257/aeri.20220292

Race and the Mismeasure of School Quality†

By Joshua Angrist, Peter Hull, Parag A. Pathak, and Christopher R. Walters*

In large urban districts, schools enrolling more White students tend to have higher performance ratings. We use an instrumental variables strategy leveraging centralized school assignment to explore this relationship. Estimates from Denver and New York City suggest that the correlation between school performance ratings and White enrollment shares reflects selection bias rather than causal school value added. In fact, value added in these two cities is essentially unrelated to White enrollment shares. A simple regression adjustment is shown to yield school ratings uncorrelated with race while predicting value added as well as or better than the corresponding unadjusted measures. (*JEL* H75, I21, I24, I28, J15)

In the fall of 2021, U.S. News & World Report released long-anticipated rankings of American middle and elementary schools based on test scores and other measures of student achievement. These and other school ratings—such as those of GreatSchools.org, Niche.com, and various state accountability offices—meet a growing demand for information on school quality from parents and policymakers. The intense public interest in school performance is also clear on real estate sites like Zillow and Redfin, which feature school ratings prominently. Such ratings affect families' choices of where to live and where to enroll (Bergman and Hill 2018; Hasan and Kumar 2019), as well as district decisions to restructure schools (Rockoff and Turner 2010; Abdulkadiroğlu et al. 2016; Cohodes, Setren, and Walters 2021).

Do highly sought-after school ratings serve the public interest? Journalists like Barnum and LeMee (2019) have brought recent attention to the strong correlation between widely reported rankings and the racial makeup of schools. In urban districts enrolling large numbers of non-White students, highly rated schools tend to enroll disproportionate shares of White and Asian students. For example, the student

*Angrist: MIT and NBER (email: angrist@mit.edu); Hull: Brown University and NBER (email: peter_hull@brown.edu); Pathak: MIT and NBER (email: ppathak@mit.edu); Walters: University of California, Berkeley and NBER (email: crwalters@econ.berkeley.edu). Rohini Pande was coeditor for this article. Our title is a tribute to Gould (1981), who immeasurably shaped the twentieth-century discussion of race in social science. Financial support from the Walton Foundation (grant 00101546) and the National Science Foundation (grants 1426541 and 2119849 for Peter Hull) is gratefully acknowledged. We thank the staff of GreatSchools.org for many helpful conversations. Thanks also to Eryn Heying, Anna Vallee, and Talia Gerstle for invaluable administrative support. Jimmy Chin, Ray Han, Russell Legate-Yang, and Ini Umosen provided expert research assistance. The research described here was carried out under data-use agreements between MIT and the NYC and Denver Public Schools school districts.

†Go to https://doi.org/10.1257/aeri.20220292 to visit the article page for additional materials and author disclosure statement(s).

body enrolled at the top five New York City middle schools as ranked by U.S. News is 80 percent White and Asian, compared with the 35 percent White and Asian share in the district as a whole.[1] Statistics like these suggest that links between published school ratings and racial composition may contribute to ongoing racial segregation (National Fair Housing Alliance 2006; Yoshinaga and Kamenetz 2016).

The correlation between school ratings and student race may reflect an uncomfortable truth: Black and White students have long attended schools of differing quality, a fact first brought to economists' attention by Welch (1973). Improvements in the quality of predominately Black schools account for much of the reduction in Black-White wage gaps seen from the 1950s through the 1970s (Card and Krueger 1992a, b). This progress notwithstanding, school attendance remains highly segregated within districts (Monarrez 2023). The higher achievement and graduation rates found at schools that enroll more White students may reflect these schools' greater impact on learning—a view reflected in decades of argument over access to selective enrollment high schools like Boston Latin School and New York's Stuyvesant, Brooklyn Tech, and Bronx Science (Jonas 2021).

However, the link between school rankings and schools' racial makeup may also be an artifact of selection bias. Higher-income and nonminority students tend to have better educational outcomes for reasons other than the quality of the schools they attend. School ratings based on student achievement levels are therefore likely to conflate school quality with the background of enrolled students. More sophisticated ratings that adjust for student demographics and lagged achievement, like conventional value-added models for teachers (e.g., Chetty, Friedman, and Rockoff 2014; Rothstein 2010, 2017) and schools (e.g., Deming 2014; Beuermann and Jackson 2022), may similarly be biased by unobserved differences in student composition. Recent research suggests that such selection bias is pervasive (Abdulkadiroğlu et al. 2020). Biased rating schemes are likely to direct households to low-minority rather than high-quality schools, while penalizing schools that improve achievement for less-advantaged groups.

This paper investigates the relationship between widely used public school ratings and student racial composition, drawing broader implications for school assessment systems. Our analysis focuses on two properties of a school rating: *predictive accuracy*, defined as the rating's r-squared in a regression of a school's true causal effect on achievement, and *racial imbalance*, defined as the slope in a regression of school ratings on White enrollment shares. If schools that enroll more White students tend to be better, in the sense of having higher causal value added, those wishing to inform the public about school quality appear to face an unavoidable trade-off between predictive accuracy and racial imbalance.

Our findings show that the trade-off between predictive accuracy and racial imbalance can be much smaller than the observed correlation between school ratings and racial composition suggests. This conclusion is reached in two steps. First, we derive a simple but novel characterization of the theoretical link between accuracy and imbalance, based on unobserved school quality. Second, we estimate

[1] The list of top New York middle schools can be found at https://www.usnews.com/education/k12/middle-schools/new-york. Demographic shares are calculated for the 2018–2019 school year using the administrative data described in Section I.

the components of this trade-off formula by using the random variation in school attendance generated by centralized school assignment systems (Abdulkadiroğlu et al. 2017, 2022). Specifically, we adopt the instrumental variables value-added model (IV VAM) approach of Angrist et al. (2021) to study how random shifts in the racial composition and ratings of a student's school affect her achievement. This method yields feasible estimators of the relationships between causal value added, student race, and school ratings.

We study the trade-off between predictive accuracy and racial imbalance for middle school students in New York City and Denver. Both districts allocate seats using a centralized match that generates partially randomized variation in school assignment, yielding the instruments needed for IV VAM. These two districts are also central to discussions of segregation and school access: New York is America's largest district, with a long history of de facto segregation, while Denver is a majority Hispanic district with a unified enrollment match combining charter and traditional public schools.

School performance ratings based on achievement levels and achievement growth are both highly correlated with schools' racial composition in New York and Denver. Our analysis substantiates the view that this correlation is largely an artifact of selection bias. IV VAM estimates show that causal value added is unrelated to racial composition in both cities. In view of our theoretical characterization, this result suggests that adjusting school ratings to reduce racial imbalance may come at little cost.

We confirm this prediction by showing that a conventional progress-based rating adjusted to be uncorrelated with student race has predictive accuracy slightly better than that of the corresponding unadjusted measure. Moreover, in both New York and Denver, this racially balanced progress rating essentially coincides with an optimal rating constructed to best predict causal value added as a function of conventional progress ratings, student race, and school sector. Racially balanced ratings may thus represent a rare "free lunch" for school accountability policy: a simple adjustment to existing ratings, requiring only data on student race, eliminates racial imbalance while also improving the ratings' value as predictors of true school quality.

I. Settings and Data

Our Denver analysis includes students applying for sixth-grade seats at any middle school in the Denver Public Schools school district between the 2012–2013 and 2018–2019 school years. Our New York analysis includes sixth-grade applicants to New York City (NYC) middle schools for the 2016–2017 through 2018–2019 school years. We observe the school preferences and priorities submitted by each applicant and the subsequent assignments generated by each district's centralized school assignment system. We also have data on subsequent school enrollment, student demographics, and achievement scores.[2] Denver outcomes are from the Colorado Student Assessment Program (CSAP) and Colorado Measures of Academic Success (CMAS) standardized tests. New York outcomes come from state achievement tests for New York State. The main outcome for our analysis combines scaled math and

[2] The samples analyzed here are derived from those used in Angrist et al. (2021).

English language arts (ELA) scores in sixth grade, standardized to have mean zero and standard deviation one in each city and year. Combining math and ELA scores helps to align our outcome with ratings reported by GreatSchools.org, school districts, and states.

Students in Denver rank up to five schools in the district. Admissions priorities are based on criteria like sibling status and the applicant's residential neighborhood. The deferred acceptance (DA) algorithm with a single lottery tiebreaker assigns students to schools. New York school applicants rank up to 12 academic programs; while schools may host more than one program, our analysis aggregates multiple programs to the school level. The New York DA algorithm features a variety of tiebreakers, with "unscreened" schools using a common random lottery number and "screened" schools using nonrandom tiebreakers such as past test scores and grades.

Our empirical strategy leverages the randomness embedded in each city's school assignment mechanism. We follow Abdulkadiroğlu et al. (2017, 2022) in computing each applicant's risk (i.e., probability) of assignment to each school as a function of the applicant's school preferences and priorities. Assignment risk for Denver applicants is computed using the propensity score formula derived by Abdulkadiroğlu et al. (2017). This formula is an analytical large-market approximation to the school assignment probability for DA with a lottery tiebreaker.[3] Assignment risk for New York applicants is computed as described in Abdulkadiroğlu et al. (2022). New York assignment risk depends in part on bandwidths for screened school tiebreakers, similar to those used in standard regression discontinuity designs.[4] Score conditioning yields a stratified randomized trial: conditional on assignment risk, school assignment is independent of applicant characteristics, both observed and unobserved—an application of the Rosenbaum and Rubin (1983) propensity score theorem.

Our analysis of school ratings focuses on two achievement-based measures of school quality that replicate widely disseminated state ratings for Colorado and New York State. *Levels* ratings consist of the share of students scored as proficient on state assessments, averaged across math and ELA tests. *Progress* ratings are based on year-to-year improvement in the average math and ELA achievement percentiles of enrolled students. This mirrors the student growth percentiles reported by many states and districts, as well as the GreatSchools.org Student Progress Rating. Our interest in progress ratings is partly motivated by previous findings that growth-type measures more accurately predict school quality (Angrist et al. 2017, 2021). Ratings are computed separately for every school and year and are standardized to be mean zero with a standard deviation matching our estimated standard deviation of school value added, detailed in Section II. Online Appendix B.1 details the school ratings construction.

Online Appendix Table A1 describes the students and schools in the Denver and New York samples, separately for all enrolled students and for the subsample of applicants for whom school assignment has a random component. We refer to the latter group as the sample with risk.[5] As is typical in large urban districts, most Denver and

[3] The Denver score is computed using the *formula score* described in Abdulkadiroğlu et al. (2017).

[4] The New York score is the *local DA score* described in Section 4.2 of Abdulkadiroğlu et al. (2022). Bandwidths used here are computed as suggested by Calonico et al. (2019).

[5] Formally, applicants in this sample (indexed by i) have a propensity score p_{ij} strictly between zero and one for at least one school j. Roughly a quarter of the students in each sample face some assignment risk.

New York students are from disadvantaged backgrounds, with over 70 percent eligible for a subsidized lunch. In both districts, the demographic characteristics, enrollment status, and baseline scores of applicants with assignment risk are similar to those of the full sample of sixth-grade students. Our New York sample includes 1,584 school-year observations with a median enrollment of 83 students. Our Denver sample includes 435 school-year observations with a median enrollment of 81 students.[6]

Table 1 validates the natural experiment from centralized assignment by comparing the characteristics of students offered seats at higher-rated and lower-rated schools (these comparisons are based on the progress rating). Uncontrolled comparisons show large differences in characteristics between those offered seats at high- and low-rated schools, but these differences vanish when we adjust for assignment risk. The fact that risk adjustment balances observed characteristics suggests that unobserved characteristics are likely balanced as well.[7]

Figure 1 shows that both the levels and progress ratings are highly correlated with the racial composition of schools. Specifically, the figure plots average school ratings computed conditional on share White in bins of width 0.1, along with the corresponding regression line fit to school-level data. Evidence of racial imbalance is especially strong for levels ratings. In New York, a regression of levels ratings on share White yields a slope coefficient of 0.70 with a robust standard error of 0.03. The standard deviation of each rating equals roughly 0.2, so this coefficient implies that a ten percentage point increase in share White is associated with a rating increase of about 0.35 standard deviations. The corresponding slope is smaller for progress, falling to 0.22, but the relationship remains clear and statistically precise. Evidence of racial imbalance for Denver is similar, with coefficients of 0.85 for the levels rating and 0.38 for the progress rating (both precisely estimated).

II. Econometric Framework

The distinction between causal value added and selection bias is cast here in terms of a constant-effects causal model of education production. Consider a population of students, each attending one of J schools in a district. Student i's potential academic achievement at school $j \in \{1, \ldots, J\}$, denoted Y_{ij}, is given as

$$Y_{ij} = \beta_j + \varepsilon_i, \tag{1}$$

where β_j gives the contribution of attendance at school j to achievement; we refer to this as school j's *quality* or *value added*. The random variable ε_i reflects other factors that influence a student's academic achievement, such as family background, motivation, and ability.

[6] The tenth (ninetieth) percentile of school-year enrollment is 36 (279) in New York and 19 (141) in Denver. Online Appendix Table A2 further summarizes the samples of schools in both settings.

[7] Balance checks regress student characteristics on the progress rating of the school where applicants are offered a seat, along with a dummy indicating whether the applicant was offered a seat anywhere. The school progress rating is set to zero for nonoffered students. Risk controls consist of the expected progress rating and the probability of receiving any offer. The former is computed as a score-weighted average of the school quality measure, following Borusyak and Hull (forthcoming). Online Appendix Table A3 further shows that differential attrition is unlikely a concern in this sample: follow-up rates for key outcomes are unrelated to assigned school ratings, conditional on assignment risk.

Table 1—Statistical Tests for Balance

	NYC		Denver	
	Uncontrolled (1)	Controlled (2)	Uncontrolled (3)	Controlled (4)
Panel A. Demographics				
Hispanic	−0.169	0.037	−0.481	0.033
	(0.008)	(0.025)	(0.014)	(0.045)
Black	−0.540	−0.013	0.020	0.006
	(0.007)	(0.023)	(0.010)	(0.033)
Asian	0.357	−0.030	−0.006	0.012
	(0.006)	(0.016)	(0.005)	(0.014)
White	0.360	−0.002	0.443	−0.043
	(0.005)	(0.013)	(0.013)	(0.036)
Female	0.020	0.034	−0.051	0.006
	(0.008)	(0.025)	(0.015)	(0.046)
Free or reduced-price lunch	−0.274	0.045	−0.519	0.037
	(0.007)	(0.020)	(0.014)	(0.041)
Special education	−0.092	0.003	−0.054	0.000
	(0.006)	(0.020)	(0.009)	(0.027)
English language learner	0.017	0.031	−0.282	0.019
	(0.005)	(0.017)	(0.014)	(0.046)
Panel B. Baseline scores				
Math (standardized)	1.02	0.020	0.858	0.090
	(0.015)	(0.046)	(0.030)	(0.092)
ELA (standardized)	0.759	−0.016	0.780	0.013
	(0.015)	(0.048)	(0.029)	(0.088)
Observations	184,760	46,095	37,089	8,100

Notes: This table reports balance statistics, estimated by regressing baseline covariates on the estimated progress rating of the offered school and an indicator for any offer. Rows report the estimated coefficient on the former. Estimates in columns 2 and 4 control for expected progress rating, any offer risk, and running variable controls in the New York sample. Expected progress rating is computed as a score-weighted average of the school quality measure, following Borusyak and Hull (forthcoming). Robust standard errors are reported in parentheses.

Equation (1) is a constant-effects model because ε_i is assumed to vary across students but not schools. For any two schools, j and k, and any applicant, i, $Y_{ij} - Y_{ik} = \beta_j - \beta_k$ gives the causal effect of attending j rather than k. This constant-effects setup allows us to focus on selection bias rather than treatment effect heterogeneity.[8]

The outcome observed for student i, denoted Y_i, equals the potential outcome associated with the school he or she attends. Letting D_{ij} be an indicator for student i's enrollment at school j, we have

$$Y_i = \sum_j Y_{ij} D_{ij} = \sum_j \beta_j D_{ij} + \varepsilon_i. \tag{2}$$

[8] Angrist et al. (2017, 2021) find little evidence of effect heterogeneity in lottery-based analyses of school value added in the cities studied here. This conclusion is supported by estimates that allow school effects to vary with student characteristics.

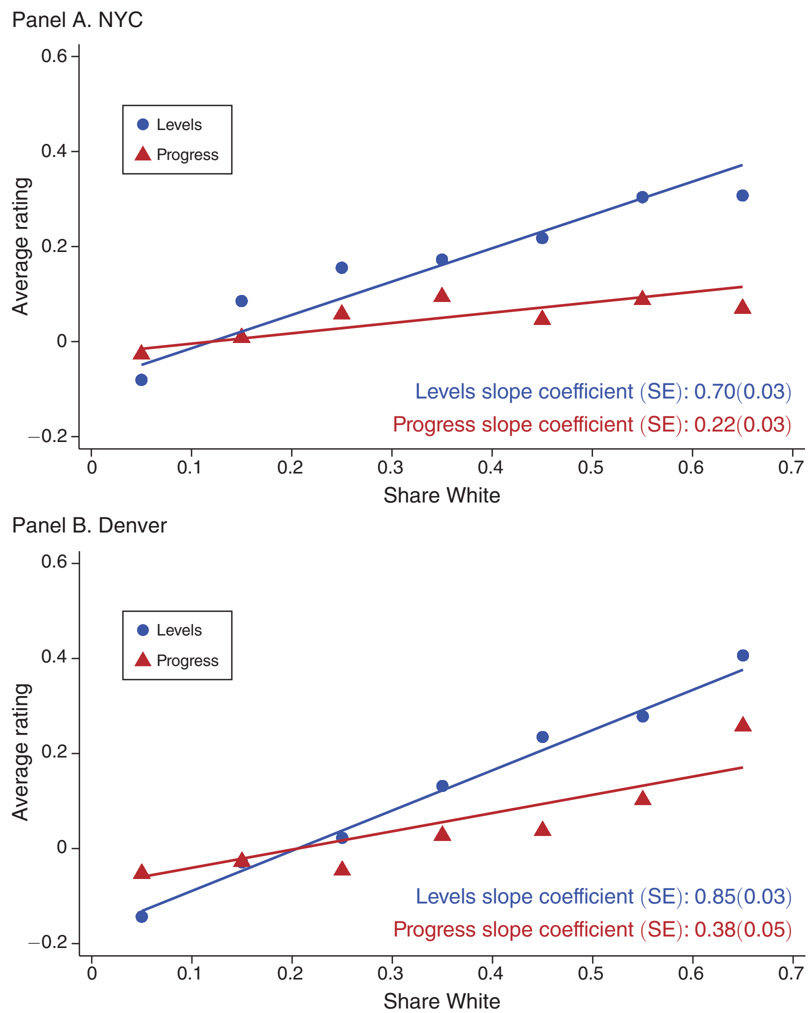

Figure 1. Levels, Progress, and Race

Notes: These binned scatterplots depict average levels and progress ratings conditional on the share of students at a school that are White. Bins are defined by 0.1 increments in share White with the last bin grouping schools with share White $\geq$ 0.6. The levels rating is the mean share of students deemed proficient in math and ELA, based on sixth-grade state assessment scores. The progress rating is computed using the student growth percentile models described in online Appendix B.1. Ratings are mean zero and scaled to have standard deviation equal to the standard deviation of school quality across schools in the district, which equals roughly 0.2 in both cities.

The average outcome at school j is given by $E\left[Y_i | D_{ij} = 1\right]$. School attendance is not randomly assigned, so these average outcomes may be a poor guide to causal effects. In particular, for any school j, $E\left[Y_i | D_{ij} = 1\right] = \beta_j + E\left[\varepsilon_i | D_{ij} = 1\right]$, which differs from β_j when schools are chosen based on factors that are correlated with ε_i.

Schools are also distinguished by the demographic composition of their student bodies. Let W_j denote the share of students enrolled in school j designated as White—that is, $W_j = E\left[w_i | D_{ij} = 1\right]$, where w_i indicates student i's race. Correlation between share White and school ratings may arise because of a relationship between

W_j and β_j, in which case the rating accurately reveals a demographic gap in school quality. Alternatively, this correlation may arise because D_{ij} is correlated with (w_i, ε_i)—a case of selection bias.

A. *Racial Imbalance and Predictive Accuracy*

Because β_j is unobserved, educational authorities report an imperfect rating, R_j, computed as a function of student achievement. As in earlier work on value added (e.g., Angrist et al. 2016, 2017), we treat school-level characteristics—here, ratings, quality, and share White—as random variables. Our investigation of the relationship between school ratings and racial composition considers the following two aspects of the distribution of school ratings:

DEFINITION: *The predictive accuracy of school rating R_j is defined as $\rho_R = \text{cov}(\beta_j, R_j)^2 / [\text{var}(\beta_j)\text{var}(R_j)]$. The racial imbalance of school rating R_j is given by $\mathcal{I}_R = \text{cov}(W_j, R_j)/\text{var}(W_j)$.*

The predictive accuracy of a rating scheme is the r-squared from a regression of school quality on ratings. Parents or policymakers seeking to identify effective schools should prefer ratings with higher ρ_R. A rating scheme's racial imbalance is the slope coefficient from a regression of R_j on W_j. These features are defined for any choice of R_j, so that $\mathcal{I}_\beta$ denotes the slope coefficient from a regression of β_j on W_j.[9]

Racially imbalanced rating schemes may favor schools with a higher share White regardless of school quality. To ameliorate this, race-balanced ratings can be constructed as the residual from a regression of R_j on W_j:

$$R_j = \gamma + \lambda W_j + \tilde{R}_j, \tag{3}$$

where $\lambda = \mathcal{I}_R$. By construction, $\tilde{R}_j$ is uncorrelated with W_j and thus has racial imbalance $\mathcal{I}_{\tilde{R}} = 0$.

Although racial imbalance is easily eliminated, this may come at the cost of reduced predictive accuracy. To describe this trade-off, consider first the coefficients on ratings in the following two regressions of school quality:

$$\beta_j = \mu + \varphi R_j + \nu_j; \tag{4}$$

$$\beta_j = \tilde{\mu} + \tilde{\varphi} R_j + \tau W_j + \tilde{\nu}_j. \tag{5}$$

Predictive accuracy is the r-squared for (4) and is therefore proportional to φ^2, while $\tilde{\varphi}$ coincides with the coefficient from a regression of β_j on the ratings residual $\tilde{R}_j$. We refer to φ and $\tilde{\varphi}$ as *forecast coefficients*, quantifying the relationship between school quality and imperfect ratings.

[9] In practice, the school quality distributions we study, like school ratings, are year specific. See online Appendix B.1 for details.

Suppose that schools with a higher share of White students tend to be rated higher, as in Figure 1: that is, $\mathcal{I}_R > 0$. The two forecast coefficients are then related as follows:

PROPOSITION 1: *Suppose that $\mathcal{I}_R > 0$. Then $\tilde{\varphi} > \varphi$ if and only if $\tau < 0$.*

PROOF:
By the omitted variables bias formula, $\varphi = \tilde{\varphi} + \tau \frac{\mathrm{cov}(R_j, W_j)}{\mathrm{var}(R_j)}$. So $\tilde{\varphi} > \varphi$ if and only if $\tau < 0$ when $\mathrm{cov}(R_j, W_j) > 0$. ∎

Proposition 1 shows that given the gradient in Figure 1, race-adjusted ratings generate a larger forecast coefficient whenever the coefficient on share White in the long forecast regression (5) is negative. $\tau < 0$ corresponds to a scenario in which schools with a higher share White tend to have value added below that of other schools with the same rating. This pattern arises, for example, with a rating scheme that rewards share White in a school system where race predicts ε_i but not school quality.

The effect of racial adjustment on predictive accuracy is given by the ratio of the forecast coefficients defined by (4) and (5), along with τ and the racial imbalance in school quality:

PROPOSITION 2: *Suppose that $\mathcal{I}_R > 0$ and $\tilde{\varphi} > 0$. Then $\rho_{\tilde{R}} > \rho_R$ if and only if $\mathcal{I}_\beta < -\tau(\varphi/\tilde{\varphi})$.*

PROOF:
Predictive accuracy for R_j and $\tilde{R}_j$ is given by $\rho_R = \frac{\varphi^2 \mathrm{var}(R_j)}{\mathrm{var}(\beta_j)}$ and

$$\rho_{\tilde{R}} = \frac{\tilde{\varphi}^2 \mathrm{var}(\tilde{R}_j)}{\mathrm{var}(\beta_j)} = \frac{\tilde{\varphi}^2 \left[\mathrm{var}(R_j) - \lambda^2 \mathrm{var}(W_j)\right]}{\mathrm{var}(\beta_j)},$$

respectively, where the latter expression uses fact that the fitted values and residuals in regression (3) are uncorrelated. The change in r-squared after residualizing is therefore proportional to

$$(6) \qquad \begin{aligned} (\rho_{\tilde{R}} - \rho_R)\mathrm{var}(\beta_j) &= \tilde{\varphi}^2\left[\mathrm{var}(R_j) - \lambda^2 \mathrm{var}(W_j)\right] - \varphi^2 \mathrm{var}(R_j) \\ &= (\tilde{\varphi} - \varphi)(\tilde{\varphi} + \varphi)\mathrm{var}(R_j) - \tilde{\varphi}^2 \lambda^2 \mathrm{var}(W_j) \\ &= -\tau\lambda \frac{\mathrm{var}(W_j)}{\mathrm{var}(R_j)}(\tilde{\varphi} + \varphi)\mathrm{var}(R_j) - \tilde{\varphi}^2 \lambda^2 \mathrm{var}(W_j) \\ &= -\left[\tau(\tilde{\varphi} + \varphi) + \tilde{\varphi}^2 \lambda\right]\lambda \mathrm{var}(W_j), \end{aligned}$$

using the fact that $\tilde{\varphi} - \varphi = -\tau\lambda \frac{\text{var}(W_j)}{\text{var}(R_j)}$ by the proof of Proposition 1 and the definition of $\lambda = \text{cov}(W_j, R_j)/\text{var}(W_j)$. Since $\lambda = \mathcal{I}_R$ by definition and $\mathcal{I}_R > 0$, equation (6) shows that when $\tilde{\varphi} > 0$, $\rho_{\tilde{R}} > \rho_R$ if and only if

$$\tau + \tilde{\varphi}\lambda < -\tau\frac{\varphi}{\tilde{\varphi}}. \tag{7}$$

By the omitted variables bias formula, $\tau + \tilde{\varphi}\lambda = \mathcal{I}_\beta$. ∎

This result is especially sharp in a scenario where school quality is unrelated to race, so $\mathcal{I}_\beta = 0$. In this case, if ratings are racially imbalanced ($\mathcal{I}_R > 0$) but still informative, then $\tau < 0$ and $\rho_{\tilde{R}} > \rho_R$.[10] More generally, Proposition 2 shows that when τ is negative, racial adjustment increases the predictive value of ratings as long as race is a sufficiently weak predictor of school quality. In this case, Proposition 2 shows that racial adjustment offers a free lunch: boosting predictive accuracy by eliminating racial imbalance.

An analyst solely interested in maximizing predictive accuracy might combine information on racial makeup with ratings data, with a rating given by the fitted value from (5):

$$\beta_j^* = \tilde{\mu} + \tilde{\varphi}R_j + \tau W_j. \tag{8}$$

This best linear predictor of school quality may improve and cannot reduce predictive accuracy relative to R_j and $\tilde{R}_j$ since the extra regressor, W_j, cannot reduce r-squared.[11] The question of whether β_j^* mitigates racial imbalance is addressed by the following result:

PROPOSITION 3: *The racial imbalance of the fitted values from regression* (5) *and the racial imbalance of causal value added coincide:* $\mathcal{I}_{\beta^*} = \mathcal{I}_\beta$.

PROOF:

$\text{cov}(W_j, \tilde{\nu}_j) = 0$, so $\frac{\text{cov}(W_j, \beta_j)}{\text{var}(W_j)} = \frac{\text{cov}(W_j, \beta_j^* + \tilde{\nu}_j)}{\text{var}(W_j)} = \frac{\text{cov}(W_j, \beta_j^*)}{\text{var}(W_j)}$. ∎

This result formalizes the intuition that any racial imbalance in school quality is captured by the coefficient on W_j in the model generating β_j^*.

In summary, Propositions 1–3 show that the trade-off between the predictive power and racial imbalance of a school rating scheme depends on two forecast coefficients φ and $\tilde{\varphi}$; the coefficient τ in equation (5); and the racial imbalance of value

[10] If $\mathcal{I}_\beta = 0$, then τ is proportional to $\text{cov}(\beta_j, W_j - \alpha R_j) = -\alpha\text{cov}(\beta_j, R_j)$, where α is the coefficient from a regression of W_j on R_j. When $\mathcal{I}_R > 0$, $\alpha > 0$, so $\tau < 0$ when $\varphi \propto \text{cov}(\beta_j, R_j) > 0$.

[11] To see this for $\tilde{R}_j$, let $\hat{R}_j$ be the fitted values from (3) and write equation (8) as

$$\beta_j^* = \tilde{\mu} + \tilde{\varphi}\hat{R}_j + (\tilde{\varphi}\tilde{R}_j + \tau W_j).$$

The term in parentheses on the right-hand side is orthogonal to the balanced rating, $\tilde{R}_j$, so the variance of β_j^* exceeds the variance of $\tilde{R}_j$.

added, $\mathcal{I}_\beta$. The challenge in applying these results is that school quality parameters, β_j, are unobserved. To surmount this challenge, we estimate the determinants of predictive accuracy and racial imbalance for alternative ratings using the IV VAM empirical strategy in Angrist et al. (2021). Specifically, we use instruments to estimate the coefficients in (4) and (5): $\varphi, \tilde{\varphi}$, and τ. IV VAM also yields a measure of $\mathcal{I}_\beta$, the slope from a regression of school quality on share White, and an estimate of the total variance of β_j, which is used to calculate the predictive accuracy of each rating.

B. *Identification and Estimation*

The IV VAM approach starts with an augmented version of regression (5) that incorporates additional predictors of school quality. The augmented model can be written

$$\beta_j = \mathbf{M}_j'\psi + \xi_j, \tag{9}$$

where $\mathbf{M}_j$ denotes a vector of quality predictors. $\mathbf{M}_j$ includes a constant, school ratings, share White, and school sector dummies. Forecast regression (9) is a linear projection, so $E[\mathbf{M}_j \xi_j] = 0$ by definition of the forecast residual ξ_j. Substituting this projection into the causal model (2) yields

$$\begin{aligned} Y_i &= \sum_j \left(\mathbf{M}_j'\psi + \xi_j\right) D_{ij} + \varepsilon_i \\ &= \mathbf{M}_{j(i)}'\psi + \xi_{j(i)} + \varepsilon_i, \end{aligned} \tag{10}$$

where $\mathbf{M}_{j(i)} = \sum_j \mathbf{M}_j D_{ij}$ and $\xi_{j(i)} = \sum_j \xi_j D_{ij}$ denote the school characteristics and forecast residual for student i's school, indexed by $j(i)$. Equation (9) is a linear projection, but equation (10) need not be: selection bias makes it likely that elements of $\mathbf{M}_{j(i)}$ are correlated with ε_i. IV VAM therefore uses centralized school assignment offers, denoted Z_{ij} for school j, as instruments for the school characteristics in $\mathbf{M}_{j(i)}$.[12]

The IV VAM estimating equation includes a vector of individual-level control variables, $\mathbf{X}_i$, including school assignment risk and other applicant characteristics. Controlling for the latter isn't necessary for identification but may boost precision.[13] Let θ denote the coefficient from a regression of the composite residual $\xi_{j(i)} + \varepsilon_i$ on $\mathbf{X}_i$, with associated residual η_i. The IV VAM estimating equation can then be written

$$Y_i = \mathbf{M}_{j(i)}'\psi + \mathbf{X}_i'\theta + \eta_i, \tag{11}$$

where $E[\mathbf{X}_i \eta_i] = 0$ by definition of θ.

[12] An alternative approach would be to instrument the school enrollment indicators in equation (2), thereby estimating the β_j parameters directly. Such direct estimation is infeasible here, however, because some schools are undersubscribed. Angrist et al. (2021) address the identification problem arising from the fact that we have fewer instruments than schools.

[13] Additional controls are functions of fifth-grade math and ELA scores, the demographic variables listed in online Appendix Table A1, and year fixed effects interacted with lagged scores and demographic characteristics. Risk controls for New York include local linear functions of the relevant screened school tiebreakers; see Abdulkadiroğlu et al. (2022) for details.

The addition of risk controls to the covariate vector in a linear model is sufficient to ensure that offer instruments Z_{ij} are uncorrelated with unobserved applicant background and ability, ε_i. Importantly, however, residual η_i in (11) depends on a school component, $\xi_{j(i)}$, as well as applicant heterogeneity, ε_i. The former reflects determinants of value added not explained by the included endogenous variables and can be thought of as arising from violations of the IV exclusion restriction that underpins identification in this context. Angrist et al. (2021) formulate sufficient conditions for IV VAM estimates to be consistent in the face of such violations. Intuitively, these conditions require the relationship between individual school offers and residual school quality to average to zero over schools.

The IV VAM exclusion restriction is made more plausible by including likely strong predictors of school quality in $\mathbf{M}_j$. Intuitively, adding such mediators reduces and perhaps even eliminates variation in residual school quality, ξ_j. In our implementation, $\mathbf{M}_j$ includes the levels and progress ratings, share White, a dummy for charter schools (in Denver), and a dummy for screened schools (in New York). By instrumenting the average test score levels and growth measures, we avoid mechanical biases from simply regressing outcomes on outcome averages.

We estimate the parameters in (11) by two-stage least squares (2SLS). This yields estimates of ψ in equation (9), defined as the regression of β_j on the full vector of school characteristics, $\mathbf{M}_j$. Coefficients in shorter projections of β_j on subsets of $\mathbf{M}_j$ can then be generated by application of the omitted variables bias formula. For example, the coefficients in (5) are obtained from a partition such that $\mathbf{M}_j = \left(\mathbf{M}_{1j}', \mathbf{M}_{2j}'\right)'$, with $\mathbf{M}_{1j} = \left(1, R_j, W_j\right)'$ and $\psi = \left(\psi_1', \psi_2'\right)'$ partitioned correspondingly. We then have

$$(\tilde{\mu}, \tilde{\varphi}, \tau)' = \psi_1 + E\left[\mathbf{M}_{1j}\mathbf{M}_{1j}'\right]^{-1} E\left[\mathbf{M}_{1j}\mathbf{M}_{2j}'\right]\psi_2. \tag{12}$$

This two-step approach uses 2SLS estimates of (11) as the common foundation for forecast regressions of any shorter length. As a by-product, the minimized 2SLS minimand (an overidentification test statistic) generates a quadratic form proportional to the variance of β_j. This variance is used in the formula for predictive accuracy.[14]

III. Results

School quality is unrelated to the share of enrolled students who are White in the sample of New York schools. This can be seen in the first column of panel A in Table 2, which reports estimates of the projection of β_j on share White and a screened school indicator for schools in New York.[15] The full set of IV VAM

[14] Specifically, the variance of ξ_j is estimated by $\frac{(\mathbf{Y} - \mathbf{Q}\hat{\phi})'\mathbf{P}_{\tilde{\mathbf{Z}}}(\mathbf{Y} - \mathbf{Q}\hat{\phi})}{tr(\hat{\mathbf{\Pi}}'\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}}\hat{\mathbf{\Pi}})}$, where $\mathbf{Y}$ is the vector of outcomes, $\mathbf{Q}$ is the matrix of endogenous regressors and covariates, $\mathbf{P}_{\tilde{\mathbf{Z}}}$ is the projection matrix for the instruments $\tilde{\mathbf{Z}}$ after partialling out covariates, $\hat{\phi}$ is the vector of 2SLS coefficient estimates, and $\hat{\mathbf{\Pi}}$ is the matrix of first-stage coefficient estimates. Supplemental Appendix I of Angrist et al. (2021) derives this formula. The version used here omits bias-correction terms that yield qualitatively similar results; see Angrist et al. (2021) for details.

[15] Online Appendix Tables A4 and A5 report results from models that replace share White with share White or Asian and with the share of students not eligible for free or reduced-price lunch. These variations yield results similar to those in Table 2.

Table 2—Projections of School Quality and School Ratings on School Characteristics

		Test score levels			Test score progress		
	Value added projection (derived)	Value added projection (derived)	Rating projection (OLS)	Value added projection (derived)	Value added projection (derived)	Rating projection (OLS)	Value added projection (derived)
Dependent variable:	School quality (β)	School quality (β)	Test score levels (R)	School quality (β)	School quality (β)	Test score progress (R)	School quality (β)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. NYC							
Predictors							
Test score levels		0.214		0.391			
		(0.053)		(0.060)			
Test score progress					0.757		0.785
					(0.037)		(0.037)
Screened school dummy	−0.052		0.101	−0.092		−0.034	−0.025
	(0.035)		(0.014)	(0.035)		(0.016)	(0.032)
Share White	0.004		0.687	−0.265		0.222	−0.171
	(0.061)		(0.024)	(0.069)		(0.026)	(0.057)
First-stage F				23.2			
Observations (school-year)				1,501			
Panel B. Denver							
Predictors							
Test score levels		0.468		1.28			
		(0.124)		(0.207)			
Test score progress					0.859		0.975
					(0.084)		(0.099)
Charter school dummy	0.095		0.098	−0.031		0.141	−0.043
	(0.037)		(0.011)	(0.046)		(0.020)	(0.040)
Share White	0.188		0.881	−0.941		0.433	−0.235
	(0.135)		(0.027)	(0.225)		(0.051)	(0.135)
First-stage F				15.1			
Observations (school-year)				373			

Notes: Estimates in columns 1, 2, 4, 5, and 7 are from projections of school quality on the predictors listed at left. These estimates are derived from the long IV VAM coefficient estimates reported in online Appendix Table A6, computed via the omitted variables bias formula as described in the text. Estimates in columns 3 and 6 are from models that predict ratings. These come from OLS regressions of school ratings on share White and a school sector dummy. Robust standard errors are reported in parentheses.

estimates underlying these results appears in online Appendix Table A6.[16] Both of the derived coefficient estimates in column 1 of Table 2 are small and significantly insignificant. The share White coefficient is estimated precisely enough to rule out a racial imbalance as large as 0.124 on the basis of 95 percent confidence interval coverage. The large racial imbalance estimate of 0.687 in column 3, by contrast, shows that share White is highly predictive of school ratings based on test score levels—as we saw in Figure 1. Together, the results in columns 1 and 3 imply that the strong

[16]The first-stage F-statistics for these estimates, computed as Kleibergen and Paap (2006) robust Wald test statistics, are above the rule-of-thumb threshold of ten commonly used to diagnose weak instrument bias. The 2SLS estimates in the table are also close to just-identified IV estimates reported in Table A6, from models where weak instrument bias is unlikely a concern. This just-identified estimator replaces individual school offer dummies as instruments with values of the mediator at the offered school, one for each mediator. Overidentified limited information maximum likelihood and the bias-corrected IV estimator in Kolesár et al. (2015) are likewise similar to the 2SLS estimates reported here.

relationship between school ratings and share White in New York reflects selection bias and not school quality.[17]

Levels ratings are weakly related to school quality in New York: the estimated forecast coefficient in column 2 of Table 2 (panel A) shows that a one standard deviation improvement in test score levels is associated with a 0.21 standard deviation increase in causal value added.[18] Column 4 reports estimates of $\tilde{\varphi}$ and τ in forecast equation (5), computed by adding share White and screened school status to ratings as predictors of school quality. Estimated coefficients on the screened school dummy and share White are both negative and significantly different from zero. This conforms to the pattern discussed above: schools that enroll more White students, as well as highly sought-after screened schools, are of lower quality than other similarly rated schools.

Column 5 of Table 2 shows that progress ratings predict New York school quality with a forecast coefficient of about 0.76—a marked improvement relative to the levels rating. But progress ratings are compromised by selection bias too. Column 6 in panel A reports an estimated share White coefficient of 0.22 in a regression of progress that controls for a screened school dummy. Column 7 shows that the progress coefficient remains high when quality is predicted by progress and share White, but share White is again negatively related to quality. Like the estimates in column 4, this pattern reflects the fact that quality and share White are unrelated, so that disproportionately White and screened schools are, on average, overrated. The fact that progress ratings exhibit modest selection bias, while improving markedly over the predictive accuracy of the levels rating, is consistent with past findings on bias in school value-added models (Angrist et al. 2017, 2021). The fact that racial imbalance decreases in a more accurate rating is consistent with our main finding that school quality itself is uncorrelated with student racial composition.

Analogous results for Denver, reported in panel B of Table 2, are qualitatively similar to those for New York, though these smaller-district estimates are less precise. Column 1 shows a statistically insignificant relationship between school quality and share White, while Denver's many charter schools generate a precisely estimated achievement gain of about 0.10 standard deviations.[19] As in New York, share White predicts levels more than progress (compare columns 3 and 6 in panel B), but both predictive relationships for ratings are strong. Also as in New York, multivariate quality projections for Denver yield negative (though more imprecise) estimated coefficients on share White when ratings are included as an explanatory variable; see columns 4 and 7 of panel B.

[17] Online Appendix Table A7 tests the equality of IV estimates of the racial imbalance of school quality and OLS estimates of the racial imbalance of either the levels ratings or the progress ratings. Columns 1 and 2 use a Hausman (1978) test, while columns 3 and 4 use a test based on the joint estimation of IV and OLS models. Both tests reject equality of IV and OLS decisively in New York.

[18] As detailed in online Appendix B.1, each rating is scaled to have the same standard deviation as estimated for value added so that the forecast coefficient can be interpreted as the standard deviation gain in causal value added associated with a one standard deviation increase in the rating.

[19] Denver estimates are imprecise enough to not be able to rule out moderate degrees of racial imbalance. Notably, however, the online Appendix Table A7 tests find significant selection bias in the racial imbalance of both levels and progress ratings (though test rejections are more marginal for progress).

Table 3—Predictive Accuracy and Racial Imbalance

	NYC		Denver	
	Predictive accuracy (ρ)	Racial imbalance ($\mathcal{I}$)	Predictive accuracy (ρ)	Racial imbalance ($\mathcal{I}$)
	(1)	(2)	(3)	(4)
Test score levels	0.046	0.702	0.219	0.846
		(0.026)		(0.027)
Test score progress	0.573	0.217	0.738	0.384
		(0.026)		(0.050)
Race-balanced progress	0.596	0.000	0.751	0.000
		–		–
Best linear predictor	0.598	−0.004	0.783	0.154
		(0.061)		(0.134)

Notes: This table reports predictive accuracy (ρ_R) and racial imbalance ($\mathcal{I}_R$) for alternative school ratings. Predictive accuracy is derived from IV VAM regressions of causal school quality on ratings. In rows 1, 2, and 4, racial imbalance is the bivariate OLS coefficient from a regression of ratings on share White. Test score levels and progress are estimated as described in online Appendix B.1. The best linear predictor is the fitted value obtained from model (8) augmented with a sector dummy. Race-balanced progress is the residual from a regression of progress on share White. Robust standard errors are reported in parentheses.

Online Appendix Figure A1 highlights implications of the results in Table 2 by plotting alternative ratings against share White in New York and Denver. The figure shows the estimated conditional expectation function (CEF) for three ratings, computing in ten-point bins, along with a regression fit to the underlying school-level data. As in Figure 1, the relationship between the progress rating and share White for New York schools is positive (the y-axis range in online Appendix Figure A1 is half that in the first figure). Race-balanced progress, computed as the residual from a regression of progress on share White, generates a flat regression fit by construction. The best linear predictor of New York school quality given the progress rating, share White, and screened school status (the fitted value from the model generating column 7 of Table 2) yields a very similar CEF.

IV VAM estimates suggest that ratings for Denver are less compromised by selection bias than the corresponding estimates for New York, with larger forecast coefficients for both levels and progress. Share White is also more strongly predictive of progress ratings in Denver than in New York (compare the estimates for the two cities in column 6 of Table 2). Consistent with these estimates, the CEF for the best linear predictor of Denver school quality plotted in panel B of online Appendix Figure A1 is weakly dependent on share White. Even so, the best linear predictor for Denver school quality rises much less steeply in share White compared to the CEF of the raw progress rating.

Table 3 summarizes our investigation with estimates of predictive accuracy and racial imbalance for alternative ratings. In both New York and Denver, progress ratings are far more accurate than levels ratings while also being much more weakly correlated with share White. This improvement notwithstanding, progress remains substantially correlated with race. Race-balanced ratings boost predictive accuracy in both cities. The best linear predictor of school quality given progress ratings, share White, and a sector dummy has predictive accuracy only slightly better than that of race-balanced progress. This is explained by the fact that the best linear predictor of school quality depends little, if at all, on race.

IV. Conclusions

This paper uses the random assignment embedded in centralized school assignment mechanisms to study the relationship between school ratings, school quality, and student race. In Denver and New York middle schools, the fact that schools with more White students are highly rated reflects selection bias rather than educational quality. As a result, ratings purged of their correlation with race predict school quality as well as or better than standard measures.[20]

Denver and New York are just two districts, of course, but the differences between them are noteworthy. Denver enrolls many more Hispanic students and runs a unified admissions system that includes charter schools. It's also worth noting that the correlation between race and widely disseminated accountability measures documented in Figure 1 is visible in districts nationwide. Across all US schools in 2018, regressing GreatSchools's levels school ratings on share White yields a coefficient of 0.632, while the corresponding regression for the GreatSchools progress measure is only 0.310 (see online Appendix Table A8; these regressions control for district fixed effects and charter status).[21] Larger differences in correlation appear in New York State and Colorado, the states containing our study districts. Such differences suggest that the association between race and achievement levels in the typical urban district is primarily due to selection bias. Although this is a conjecture rather than a finding, the growing importance of centralized assignment should allow a wider validation in the not-too-distant future.[22] An equally important question for future work, requiring empirical methods distinct from those used here, is whether our findings extend to racial imbalance *across* districts.

Our analysis leaves open the question of how racially balanced school ratings might affect household decision-making. Households appear to respond to school performance ratings (Hastings and Weinstein 2008; Bergman and Hill 2018; Bergman, Chan, and Kapor 2020; Houston and Henig 2023; Campos and Kearns 2021). Credible racially balanced quality information may therefore increase the demand for high-quality schools with lower White enrollment. At the same time, school choice may respond more to peer characteristics than to value added (Rothstein 2006; Abdulkadiroğlu et al. 2020). We hope to study the extent to which households respond to improved measures of school quality in future work.

REFERENCES

Abdulkadiroğlu, Atila, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak. 2016. "Charters without Lotteries: Testing Takeovers in New Orleans and Boston." *American Economic Review* 106 (7): 1878–920.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica* 85 (5): 1373–432.

[20] Other efforts in this direction, inspired by similar concerns with possibly misleading racial imbalance, include the GreatSchools Equity Rating (https://www.greatschools.org/gk/ratings-methodology/#methodology-equity-rating).

[21] Levels is GreatSchools's Test Score Rating, and progress is GreatSchools's Student Progress Rating when available and Academic Progress Rating otherwise.

[22] A review of enrollment portals in the 100 largest districts shows that more than a third of urban students attend schools in districts that assign seats centrally, while 83 percent attend schools in districts with at least some random assignment. See online Appendix Table A9 for these statistics.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2022. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Econometrica* 90 (1): 117–51.

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review* 110 (5): 1502–39.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2016. "Interpreting Tests of School VAM Validity." *American Economic Review* 106 (5): 388–92.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *Quarterly Journal of Economics* 132 (2): 871–919.

Angrist, Joshua, Peter Hull, Parag A. Pathak, and Christopher Walters. 2021. "Credible School Value-Added with Undersubscribed School Lotteries." *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01149.

Angrist, Joshua, Peter Hull, Parag A. Pathak, and Christopher R. Walters. 2023. "Replication data for: Race and the Mismeasure of School Quality." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E182002V1.

Barnum, Matt, and Gabrielle LaMarr LeMee. 2019. "Looking for a Home? You've Seen GreatSchools Ratings. Here's How They Nudge Families toward Schools with Fewer Black and Hispanic Students." *Chalkbeat*, December 5. https://www.chalkbeat.org/2019/12/5/21121858/looking-for-a-home-you-ve-seen-greatschools-ratings-here-s-how-they-nudge-families-toward-schools-wi.

Bergman, Peter, Eric W. Chan, and Adam Kapor. 2020. "Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment." NBER Working Paper 27209.

Bergman, Peter, and Matthew J. Hill. 2018. "The Effects of Making Performance Information Public: Regression Discontinuity Evidence from Los Angeles Teachers." *Economics of Education Review* 66: 104–13.

Beuermann, Diether W., and C. Kirabo Jackson. 2022. "The Short- and Long-Run Effects of Attending the Schools that Parents Prefer." *Journal of Human Resources* 57 (3): 725–46.

Borusyak, Kirill, and Peter Hull. Forthcoming. "Non-Random Exposure to Exogenous Shocks." *Econometrica*.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *Review of Economics and Statistics* 101 (3): 442– 51.

Campos, Christopher, and Caitlin Kearns. 2021. "The Impacts of Neighborhood School Choice: Evidence from Los Angeles's Zones of Choice Program." Blueprint Labs Discussion Paper 2021.16.

Card, David, and Alan B. Krueger. 1992a. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100 (1): 1–40.

Card, David, and Alan B. Krueger. 1992b. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107 (1): 151–200.

Castellano, Katherine E., and Andrew D. Ho. 2013. *A Practitioner's Guide to Growth Models*. Washington, DC: Council of Chief State School Officers.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–632.

Cohodes, Sarah R., Elizabeth M. Setren, and Christopher R. Walters. 2021. "Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector." *American Economic Journal: Economic Policy* 13 (1): 138–67.

Colorado Department of Education. 2019. *Colorado Growth Model*. Denver: Colorado Department of Education.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104 (5): 406–11.

Gould, Stephen Jay. 1981. *The Mismeasure of Man*. New York: W. W. Norton.

Hasan, Sharique, and Anuj Kumar. 2019. "Digitization and Divergence: Online School Ratings and Segregation in America." Unpublished.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–414.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251–71.

Houston, David M., and Jeffrey R. Henig. 2023. "The 'Good' Schools: Academic Performance Data, School Choice, and Segregation." *AERA Open* 9. https://doi.org/10.1177/23328584231177666.

Jonas, Michael. 2021. "Are Exam Schools Really an Academic Promised Land?" *CommonWealth*, July 15. https://commonwealthmagazine.org/education/are-exam-schools-really-an-academic-promised-land/.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133 (1): 97–126.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and Inference with Many Invalid Instruments." *Journal of Business & Economic Statistics* 33 (4): 474–84.

Monarrez, Tomás E. 2023. "School Attendance Boundaries and the Segregation of Public Schools in the United States." *American Economic Journal: Applied Economics* 15 (3): 210–37.

National Fair Housing Alliance. 2006. *Unequal Opportunity—Perpetuating Housing Segregation in America*. Washington, DC: National Fair Housing Alliance.

New York State Education Department. 2020. *2018/19 Growth Model for Institutional Accountability*. Albany, NY: New York State Education Department.

Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy* 2 (4): 119–47.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review* 96 (4): 1333–50.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125 (1): 175–214.

Rothstein, Jesse. 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review* 107 (6): 1656–84.

Welch, Finis. 1973. "Black-White Differences in Returns to Schooling." *American Economic Review* 63 (5): 893–907.

Yoshinaga, Kendra, and Anya Kamenetz. 2016. "Race, School Ratings, and Real Estate: A 'Legal Gray Area.'" National Public Radio, October 10. https://www.npr.org/sections/ed/2016/10/10/495944682/race-school-ratings-and-real-estate-a-legal-gray-area.