Race and the Mismeasure of School Quality: Supplemental Appendix

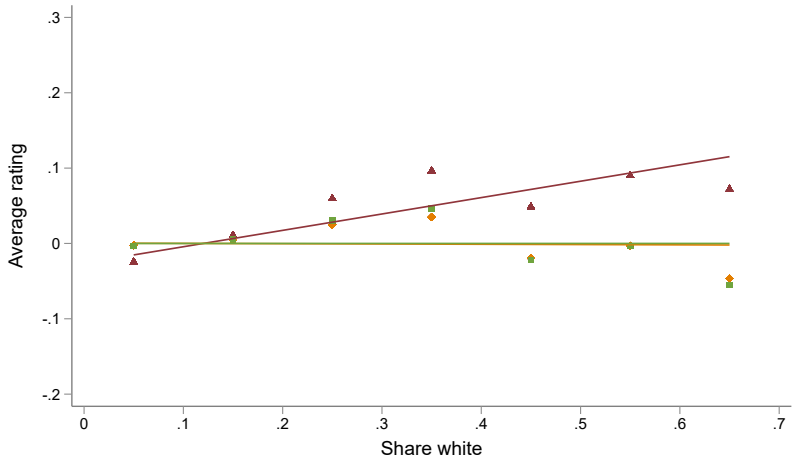
Joshua Angrist, MIT & NBER
Peter Hull, Brown University & NBER
Parag Pathak, MIT & NBER
Christopher Walters, UC Berkeley & NBER

September 2023

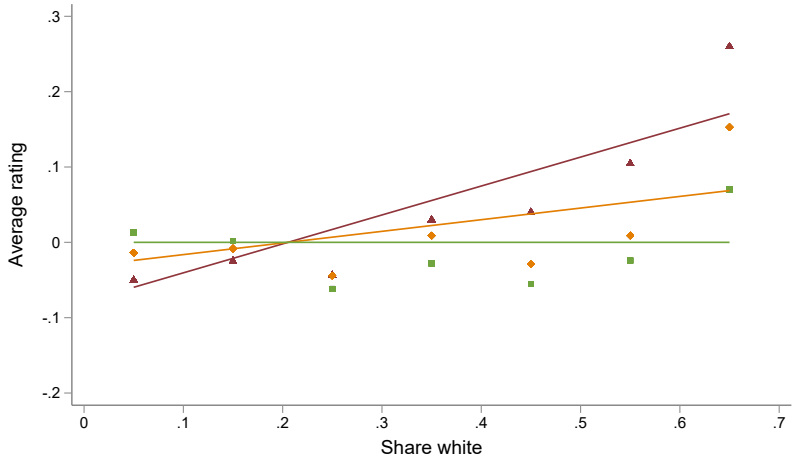
Appendix Figures and Tables

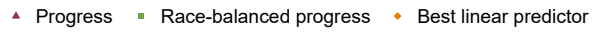
Figure A1. Adjusted Ratings and Race

A. NYC



B. Denver





Notes: These binned scatterplots depict the relationship between three sorts of progress ratings and the share of students at a school that are white. Red triangles correspond to the benchmark progress rating, while green squares correspond to the racially-balanced progress rating obtained as the residual from equation (3). Orange diamonds correspond to the best linear predictor of school value-added, obtained as the fitted values from equation (8) augmented with a sector dummy. Bins are defined by 0.1 increments in share White with the last bin grouping schools with share white ≥ 0.6. Ratings are mean zero and scaled to have standard deviation equal to the standard deviation of school quality across schools in the district.

Table A1. Descriptive Statistics

	NYC		Denver	
	All	With risk	All	With risk
	(1)	(2)	(3)	(4)
Demographics				
Hispanic	0.413	0.445	0.592	0.581
Black	0.231	0.254	0.125	0.140
Asian	0.184	0.171	0.032	0.033
White	0.154	0.110	0.210	0.201
Female	0.494	0.484	0.493	0.494
Free/reduced price lunch	0.731	0.763	0.723	0.703
Special education	0.201	0.215	0.102	0.087
English language learner	0.113	0.113	0.393	0.416
Baseline scores				
Math (standardized)	0.000	-0.063	0.000	0.077
ELA (standardized)	0.000	-0.055	0.000	0.070
Enrollment				
Screened	0.067	0.044	0.000	0.000
Lottery	0.933	0.956	1.000	1.000
Share non-compliant	0.268	0.324	0.300	0.291
Share not offered	0.149	0.134	0.182	0.048
Students	184,760	46,095	37,089	8,100
Schools	624	594	80	75
Lotteries (schools with risk)		448		67

Notes: This table describes the Denver and New York student samples used to compute ratings and estimate school quality. Column 1 show statistics for New York middle school students enrolled in 6th grade in the 2016-17 through 2018-19 school years. Column 3 shows descriptive statistics for Denver students enrolled in 6th grade in the 2012-13 through 2018-19 school years. Columns 2 and 4 describe the corresponding samples of applicants with assignment risk at at least one school. Baseline characteristics and lagged scores are from 5th grade. Baseline scores are standardized to be mean zero and standard deviation one in the student-level test score distribution, separately by year. Screened schools are defined as schools without any lottery programs. The share non-compliant is defined as the proportion of students who enroll other than where offered a seat; this includes students receiving no offers.

Table A2. School Counts

	NYC				Denver		
	TPS						
	Non-screened	Screened	Charter	All schools	TPS	Charter	All schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. School-year counts						
In sample	1359	142		1501	223	150	373
Not in sample	80	3		83	52	10	62
Total	1439	145		1584	275	160	435
	Panel B. School counts (2016)						
In sample	433	47	90	570	31	22	53
Not in sample	17	0	28	45	9	2	11
Total	450	47	118	615	40	24	64

Notes: This table describes the schools in the IV estimation sample. These schools enroll at least one student with non-degenerate risk. The columns labelled "TPS" indicate traditional public schools. Screened schools in New York are schools that offer only screened programs. In New York, student-level charter enrollment is only observed in the 2016-2017 school year. In Panel A, charter school-years are counted as non-screened observations.

Table A3. Tests for Differential Attrition

	NYC	Denver
	(1)	(2)
Offered progress	0.032	0.022
	(0.019)	(0.038)
N	53,094	9,234
Mean follow-up rate	0.898	0.896

Notes: This table reports differential attrition estimates. These estimates come from regressions of a follow-up indicator on the estimated progress rating of the offered school, controlling for expected progress rating and running variable controls in the New York sample. Robust standard errors are reported in parentheses.

Table A4. Projections of School Quality and School Ratings on Share White and Asian

		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		0.164		0.536			
		(0.055)		(0.071)			
Test score progress					0.738		0.812
					(0.037)		(0.038)
Screened school dummy	-0.047		0.101	-0.101		-0.037	-0.017
	(0.035)		(0.012)	(0.035)		(0.016)	(0.032)
Share white and Asian	-0.046		0.541	-0.336		0.199	-0.207
	(0.046)		(0.013)	(0.062)		(0.016)	(0.045)
First-stage F				15.1			
N (school-year)				1501			
				Panel B. Denver			
Predictors							
Test score levels		0.482		1.37			
		(0.148)		(0.221)			
Test score progress					0.843		0.945
					(0.089)		(0.097)
Charter school dummy	0.100		0.099	-0.033		0.139	-0.033
	(0.036)		(0.011)	(0.045)		(0.020)	(0.038)
Share white and Asian	0.175		0.834	-0.977		0.405	-0.210
	(0.126)		(0.025)	(0.219)		(0.049)	(0.122)
First-stage F				9.09			
N (school-year)				373			

Notes: This table reports estimates from projections of levels and progress school ratings and causal value added on school characteristics, including the share white and Asian. The models and derivation procedure used to compute these estimates are as the estimates in Table 2. Robust standard errors are reported in parentheses.

Table A5. Projections of School Quality and School Quality on Share Non-FRPL

		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		0.232		0.451			
		(0.052)		(0.063)			
Test score progress					0.761		0.774
					(0.037)		(0.037)
Screened school dummy	-0.050		0.060	-0.077		-0.040	-0.019
	(0.035)		(0.013)	(0.035)		(0.016)	(0.032)
Share non-FRPL	0.018		0.656	-0.278		0.144	-0.094
	(0.050)		(0.018)	(0.059)		(0.024)	(0.046)
First-stage F				20.4			
N (school-year)				1501			
				Panel B. Denver			
Predictors							
Test score levels		0.443		1.29			
		(0.147)		(0.213)			
Test score progress					0.851		0.941
					(0.083)		(0.096)
Charter school dummy	0.087		0.066	-0.011		0.124	-0.018
	(0.036)		(0.012)	(0.041)		(0.020)	(0.037)
Share non-FRPL	0.151		0.745	-0.842		0.344	-0.178
	(0.112)		(0.023)	(0.188)		(0.044)	(0.109)
First-stage F				10.9			
N (school-year)				373			

Notes: This table reports estimates from projections of levels and progress school ratings and causal value added on school characteristics, including the share not eligible for a free or reduced-price lunch. The models and derivation procedure used to compute these estimates are as the estimates in Table 2. Robust standard errors are reported in parentheses.

Table A6. IV VAM Regressions

	Over-identified (school assignment instruments)		Just-identified (offered mediator instruments)	
	NYC	Denver	NYC	Denver
	(1)	(2)	(3)	(4)
Mediators				
Test score levels	-0.140	0.417	-0.234	-0.006
	(0.064)	(0.230)	(0.102)	(0.437)
Test score progress	0.839	0.847	1.10	1.05
	(0.044)	(0.116)	(0.064)	(0.151)
Screened school dummy	-0.009		0.011	
	(0.033)		(0.037)	
Charter school dummy		-0.066		0.010
		(0.044)		(0.063)
Share white	-0.087	-0.547	-0.051	-0.129
	(0.064)	(0.217)	(0.079)	(0.340)
First-stage F	23.2	15.1	608	31.7
Value-added std. dev.	0.194	0.217		
N	46,095	8,100	46,095	8,100

Notes: This table reports IV VAM parameter estimates. These estimates are used to obtain the estimates reported in Table 2. The set of listed mediators is treated as endogenous. Columns 1 and 2 use individual school assignment offer dummies as instruments for 2SLS estimation. Columns 3 and 4 use values of the mediator at the offered school as instruments. All models control for school assignment risk and year fixed effects fully interacted with the demographic variables listed in Appendix Table A1 and cubic functions of 5th grade math and ELA scores. New York models also include local linear functions of the relevant screened-school tie-breakers. Ratings are demeaned and scaled to have variance matching that of school quality across schools in the district. Robust standard errors are reported in parentheses.

Table A7. Tests for Equality of IV and OLS Estimates of Racial Imbalance

	Hausman		Joint estimation	
	Test score levels (1)	Test score progress (2)	Test score levels (3)	Test score progress (4)
	Panel A: NYC			
Racial imbalance				
IV (school quality)	0.004		0.004	
	(0.061)		(0.062)	
OLS	0.687	0.222	0.687	0.222
	(0.024)	(0.026)	(0.024)	(0.026)
IV - OLS	-0.683	-0.219	-0.683	-0.219
	(0.055)	(0.055)	(0.066)	(0.068)
	[0.000]	[0.000]	[0.000]	[0.001]
	Panel B: Denver			
Racial imbalance				
IV (school quality)	0.188		0.188	
	(0.135)		(0.122)	
OLS	0.881	0.433	0.881	0.433
	(0.027)	(0.051)	(0.027)	(0.051)
IV - OLS	-0.693	-0.246	-0.693	-0.246
	(0.132)	(0.125)	(0.125)	(0.131)
	[0.000]	[0.049]	[0.000]	[0.060]

Notes: This table reports tests for equality between the IV estimates of the racial imbalance of school quality and OLS estimates of the racial imbalance of either the levels rating or the progress rating. Columns 1 and 2 use a Hausman (1978) test which takes as the covariance between the IV and OLS estimators the variance of the OLS estimator. Columns 3 and 4 compute the covariance between the IV and OLS estimators by jointly estimating these models. Standard errors, clustered by school-year, are reported in parentheses. P-values for the test of IV and OLS equality are reported in brackets.

Table A8. Comparison of Racial Imbalance in GreatSchools' Levels and Progress Ratings

	Test score levels (1)	Test score progress (2)
Panel A: USA		
Predictors		
Charter school dummy	0.019	0.015
	(0.005)	(0.006)
Share white	0.632	0.310
	(0.004)	(0.006)
N (schools)	72573	61247
Panel B: New York		
Predictors		
Charter school dummy	-	-
Share white	0.625	0.095
	(0.022)	(0.030)
N (schools)	3979	3099
Panel C: Colorado		
Predictors		
Charter school dummy	0.019	0.015
	(0.005)	(0.006)
Share white	0.735	0.302
	(0.022)	(0.031)
N (schools)	1210	1474

Notes: This table reports racial imbalance regressions for GreatSchools levels and progress ratings in the 2018 school year. Panel A includes all public schools in the United States with GreatSchools ratings, while Panels B and C restrict the sample to schools in New York state and Colorado, respectively. Ratings are standardized by state to have mean zero and standard deviation 0.2, which is roughly the standard deviation of school quality in both NYC and Denver. All models include district fixed effects, which absorb charter school indicators in New York. Levels is GreatSchools' Test Score Rating, and progress is GreatSchools' Student Progress Rating when available and Academic Progress Rating otherwise. See Appendix B.1 and `https://www.greatschools.org/gk/ratings-methodology/` for more information on GreatSchools ratings.

Table A9. Centralized Assignment in Large Public School Districts

	All (1)	Minority (2)
All districts		
Enrollment (% of all districts)	100%	91%
N	100	87
Centralized		
Enrollment (% of all districts)	36%	34%
N	26	24
Partially centralized		
Enrollment (% of all districts)	69%	65%
N	59	52
Any randomness		
Enrollment (% of all districts)	83%	77%
N	76	66

Notes: This table describes the student assignment mechanism for the 100 largest public school districts in the United States. Column 2 considers districts enrolling at least 30% Black and Hispanic students. Centralized districts employ mechanisms with quasi-random offer variation for traditional public schools. Partially centralized districts include those with a centralized aftermarket for school choice transfers away from neighborhood schools. Any randomness districts employ mechanisms with any random offer variation, for instance decentralized lotteries at non-traditional public schools. Further details on definitions and coding procedures are available on request. Enrollment data reflect fall 2019 figures from the NCES.

B Data Appendix

B.1 School Quality Measures

The measures used here are motivated by the "test score" and "progress" ratings published by GreatSchools.org. The test score rating is a levels measure that uses student proficiency rates as inputs. The progress rating uses state-reported estimates of student growth as inputs. Our progress ratings are based on models underlying the "growth" rating reported by Colorado and the student growth percentile estimates reported by New York.[23]

Our computation differs in a few ways from GreatSchools and state ratings because we are interested in sixth-grade ratings for specific years and outcomes; it's sometimes unclear which grades and years were used to compute published ratings. Also, GreatSchools ratings transform state-reported inputs into a discrete 1-10 rating; we omit this step. Like GreatSchools ratings, our computation is year-specific.[24]

Our *levels* rating averages the share of students who are proficient in math and the share of students who are proficient in English Language Arts (ELA), as measured by sixth-grade achievement tests. Formally, this is $R_j = (E[q_i^m \mid D_{ij} = 1] + E[q_i^e \mid D_{ij} = 1])/2$, where q_i^s indicates a student who is deemed proficient in subject s (math or ELA). Students are deemed proficient when their scores cross state-determined cutoffs.

Our *progress* rating is derived from estimates of student growth percentile models. Neither of these procedures involve simple difference-based measures of growth; rather, they adjust for lagged achievement. Nevertheless, the resulting measures are often called a "student growth percentile," or SGP (Castellano and Ho, 2013). The underlying models are described in New York State Education Department (2020) for New York and Colorado Department of Education (2019) for Colorado.

For purposes of our analysis, New York growth percentiles are computed by first estimating the regression:

$$Y_i^s = \delta^s + X_i'\Gamma^s + \eta_i^s,$$

for each subject $s \in \{m, e\}$. Here X_i is a control vector including 3rd, 4th, and 5th grade achievement scores. Missing lagged scores are coded to zero, with indicators for missing scores also included in X_i. From these regressions we compute the percentile rank, r_i^s, of the residual η_i^s in the city distribution of students. The progress rating is then the mean of the

[23]These ratings can be found through Colorado's Performance Snapshot (`https://www.cde.state.co.us/code/accountability-performancesnapshot`) and the "ACC EM Growth" table in New York's Report Card Database (`https://data.nysed.gov/downloads.php`).

[24]See `https://www.greatschools.org/gk/ratings-methodology/` for more information on the GreatSchools ratings computation.

school average math and ELA ranks: $R_j = (E[r_i^m \mid D_{ij} = 1] + E[r_i^e \mid D_{ij} = 1])/2$.

Student growth percentiles for Denver are computed using quantile regression. This procedure begins by using quantile regression to fit conditional quantiles as a function of the control vector, X_i, listed above. Quantile regression coefficients are computed for every percentile from 1-99. The Denver percentile rank is the quantile value, τ, that minimizes $Y_i^s - X_i'\hat{\Gamma}_\tau^s$, where $\hat{\Gamma}_\tau^s$ is the estimated vector of quantile regression coefficients for percentile τ. As in New York, subject-specific results are averaged to produce a single progress rating for each school and year.

B.2 Standardization of Outcomes and Ratings

The primary outcome for our analysis is constructed by first summing each student's scaled math and ELA sixth-grade test scores, then standardizing this sum to have mean zero and standard deviation one, separately by city and year. Year-specific school value added, β_j, is therefore measured in units of student-level test score standard deviations.

To facilitate comparisons of forecast coefficients across ratings, alternative ratings are scaled to have the same standard deviation as causal value-added. Specifically, we estimate the IV VAM model (11) and use the results to form an estimate $\hat{\sigma}_\beta$ of the standard deviation of causal value-added, as described in Angrist et al. (2021). For each year, we then multiply each rating (deviated from its mean) by the ratio of $\hat{\sigma}_\beta$ to its own standard deviation. This results in a rating with mean zero and standard deviation $\hat{\sigma}_\beta$. The forecast coefficients in Table 2 can therefore be interpreted as gains in standard deviations of causal value-added associated with a one standard-deviation increase in school ratings. A rating that accurately orders schools according to causal value-added should be expected to generate a forecast coefficient of roughly unity. It's worth noting, however, that the forecast coefficient may not be exactly one even for a rating that ranks schools exactly in order of β_j, since value-added and school ratings are measured in different units, even after rescaling.