Identifying High-Performing Schools for Historically Underserved Students: Exploring a Multistate Model

Technical Appendix

Erica Blom, Theresa Anderson, and Constance Lindsay
with Victoria Rosenboom, Semhar Gebrekristos, Macy Rainer, and Caroline Vilter

January 2020

The REMIQS (Robust and Equitable Measures to Inspire Quality Schools) analysis required a large amount of data compilation and cleaning and various analysis decisions. This appendix describes our analysis approach for a technical audience. It also describes the results of a separate-state model.

Data

The data we use in this study come from state longitudinal data systems (SLDS) in Kentucky, Massachusetts, and Virginia. These data systems are administrative datasets that link individual student records across multiple state agencies. The data in Kentucky, Massachusetts, and Virginia track students from the K–12 system through college and into the labor force, provided the students remain in the state.

We use ninth-grade cohorts beginning in 2009 through 2012. These data are recent enough to be relevant but old enough that we can see students enroll in college, provided they enrolled soon after high school completion.

The Importance of Individual-Level Data

The value of SLDS data is twofold. First, their longitudinal nature allows us to track students over time so we can control for individual characteristics before high school and follow them after high school. A second advantage is that the data are at the individual level. Similar studies could be done using aggregate (school-level) data, but these studies would suffer from the "ecological fallacy," the (incorrect) notion that inferences for individuals can be made from analyses of larger groups. By using individual-level data, we can link an individual student's characteristics to her own outcomes and obtain a less biased estimate of school effects.

Sample Selection

We focus on noncharter traditional public high schools (excluding vocational, alternative, or special education high schools) where at least 50 percent of students receive free and reduced-price lunch (FRPL) or at least 50 percent students are from historically underserved groups, defined as students who are American Indian, Alaska Native, Black, Hispanic, or Pacific Islander.[1] We further limit our analysis to schools with at least 100 students in the sample across the four years.

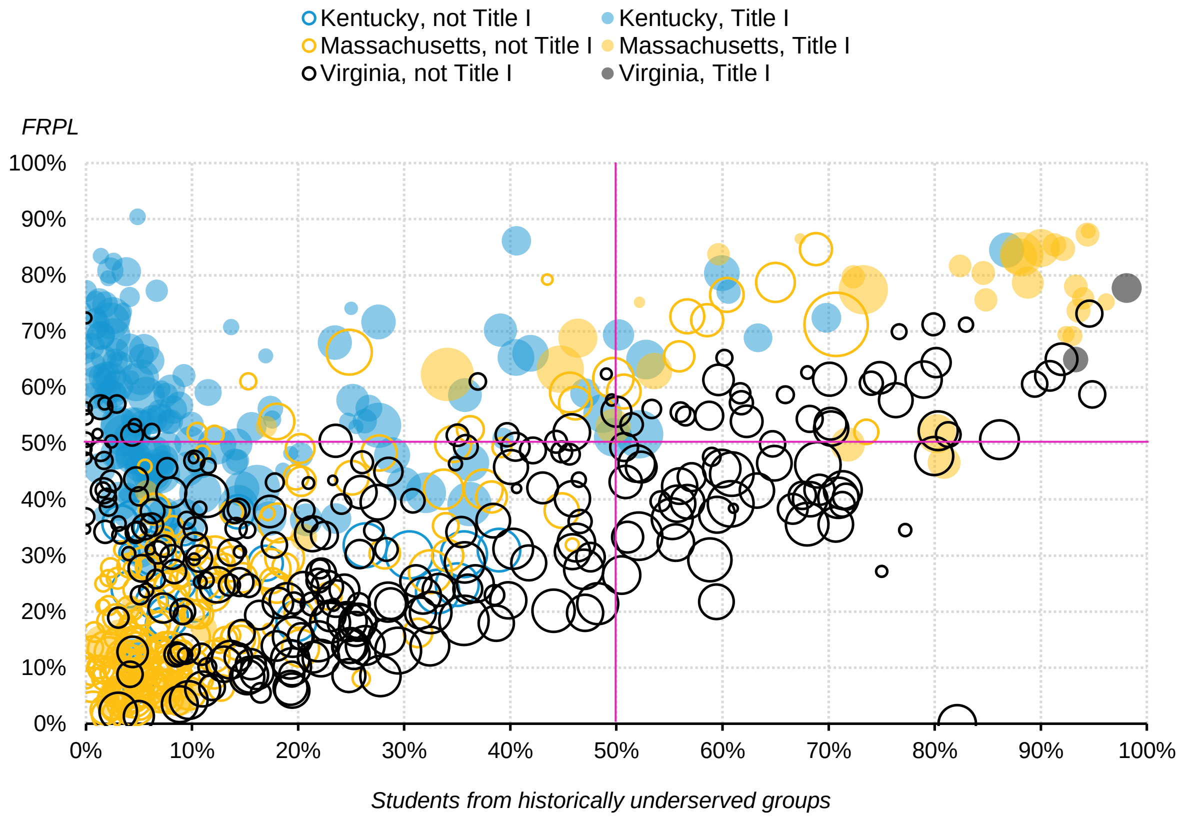
Defining the sample as schools with either a large low-income population or with a large population of historically underserved racial and ethnic groups accounts for vastly different demographics across states and includes a sufficient number of schools in each state that are meaningfully comparable. We explored but rejected other possibilities because they would have excluded too many schools we think should be included. For example, Title I eligibility—which is typically used to identify schools with high shares of students from low-income backgrounds—would have limited the analysis to only two schools in Virginia, whereas a focus only on schools with high shares of students from historically underserved groups would have limited the number of Kentucky schools we could include because many of them serve low-income White students.

Figure 1 displays the share of students receiving FRPL and the share of students from historically underserved groups at each school, as well as Title I status, by state. Bubble size indicates each school's relative size. Schools in Kentucky tend to serve poorer students (as proxied by FRPL eligibility) than other states and serve fewer students from historically underserved groups. Virginia is wealthier and more diverse, and Massachusetts lands in the middle. Our final sample includes all schools to the right of the vertical magenta line and all schools above the horizontal magenta line (excluding schools in the lower-left quadrant).

FIGURE 1

School Characteristics across States

2012



Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data.

Note: FRPL = free and reduced-price lunch.

Our final sample consists of 213 schools and 180,196 students. We limit our sample of schools before estimating the model, rather than after, which allows us to fit a better model for the populations of interest. An alternative approach is to include all high schools from the beginning and then focus on eligible schools after estimating the model. But we want to be sure high schools not in our population of interest do not influence the analysis.

Model Specification

In the concept paper for this project (Anderson et al. 2019), we laid out a detailed model to determine school value-add, given ideal data. In practice, even with the richness of each state's SLDS, the final model is simpler because of data constraints.

We use a variation on the value-add model common in education research. The model includes student-level characteristics, school-level characteristics, and district-level characteristics, and is as follows:

$$Y_{isdt} = \beta_0 + \delta_s + \gamma_t + \beta_1 X_{isdt} + \beta_2 X_{sdt} + \beta_3 X_{dt} + \epsilon_{isdt}$$

where Y_{isdt} is an outcome for student *i* in school *s* in district *d* in year *t*, X_{isdt} are individual student characteristics, X_{sdt} are school-level characteristics, X_{dt} are district-level characteristics, γ_t are year fixed effects, and ϵ_{isdt} are residuals. The school effects are captured by δ_s. By stacking four years of data, we can include school-level variables in the model and limit how much our "value-add" estimates capture noise, rather than permanent or transitory real value-add (Totty 2019). We run this model separately for each outcome Y.

Outcomes

Our focus is on long-term outcomes available across all three states and for all four cohorts. The main outcome we consider is college enrollment. College enrollment captures not only a student's ability to score well on a test but also deeper learning skills, such as goal-setting, responsibility, and confidence.

College enrollment is not a perfect measure. Students need not attend college to lead a fulfilling life. College enrollment can also be gamed; high schools bent on enrolling their students in college can make that happen. Enrollment is also not definitive; completing college is not a given after enrollment, especially for marginally prepared students. Nevertheless, it is the best long-term outcome available to us, and we prioritize this outcome over others.

For Virginia and Massachusetts, we use data from the National Student Clearinghouse (NSC), which tracks student enrollment nationally in postsecondary, Title IV, degree-granting institutions.[2] In Kentucky, we do not have NSC data and see only students who enroll in college in state. To best align with NSC institutions and the other states, we exclude less-than-two-year institutions. Furthermore, in Kentucky, we modify the outcome to include students who scored 24 or higher on the ACT. (Because the ACT is required in Kentucky, most Kentucky students in our sample take it.[3]) This is a conservative adjustment that captures high-achieving students attending selective out-of-state colleges but may miss students who live close to the state border and prefer to attend out-of-state but not necessarily selective institutions (Kentucky has tuition reciprocity arrangements with neighboring states.[4]) A score of 24 represents the 74th percentile on the ACT,[5] and counting students with this score or higher increases the college-going rate about 4.5 percent. For comparison, the Integrated Postsecondary

Education Data System suggests that 10 to 15 percent of college students from Kentucky go to college out of state, but the precise cutoff does not seem to affect results dramatically.

Table 1 shows the college-going rates at schools in our sample. Note that rates appear to decline with each subsequent cohort in two of the three states because we have less time to observe them.

TABLE 1

College-Going Rates for Schools in Our Sample

	Kentucky	Massachusetts	Virginia
2009	48%	60%	55%
2010	45%	59%	52%
2011	43%	58%	55%
2012	41%	55%	53%

Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia.

We also consider the following supplemental outputs:

- **High school attendance.** We measure this during the student's last year in the data, or as 0 if the student dropped out.
- **High school graduation.** We exclude students who died, transferred, had a long-term absence, or are unconfirmed, and we count only regular high school diplomas as completers. We do not include GEDs because this indicates the student dropped out of high school.

We do not consider the following outcomes:

- **College completion.** Our time frame is such that we cannot calculate six-year graduation rates even for the earliest cohort because they began high school in 2009 and the data extend to 2018. Requesting data for earlier cohorts poses constraints because data collection systems and standards have changed. In addition, results from cohorts from that long ago may not be meaningful to school administrators who may have gone through several turnovers in more than a decade.
- **Wages.** This turned out to be infeasible for two reasons. First, wage data are available only at the school level in Kentucky per state law, and Massachusetts could not provide individual-level records in time for the analysis. Second, our time frame is again insufficient. The general consensus is that wages begin to be a strong indicator of lifetime earnings about 10 years after high school graduation (Guvenen et al. 2015).

- **Civic participation (e.g., voting or community involvement), incarceration, or mental or physical health outcomes.** The administrative records, while rich, do not include these outcomes.

Because we focus on long-term outcomes, we omit many short-term outputs. We exclude the following:

- **College admission test scores.** Because we see dramatic differences across states in test-taking rates in our data (likely because of data constraints, as these differences do not appear to reflect actual test-taking rates reported elsewhere), we do not include this outcome.
- **Behavioral outputs (e.g., law referrals, suspensions, or expulsions).** We have behavioral data for only Kentucky and Massachusetts. Moreover, because states record these data differently, we do not to include these outputs. Individual models by state, however, could include these outputs, bearing in mind that these measures are less objective than other outputs given the possibility of discriminatory treatment (GAO 2018).

Inputs

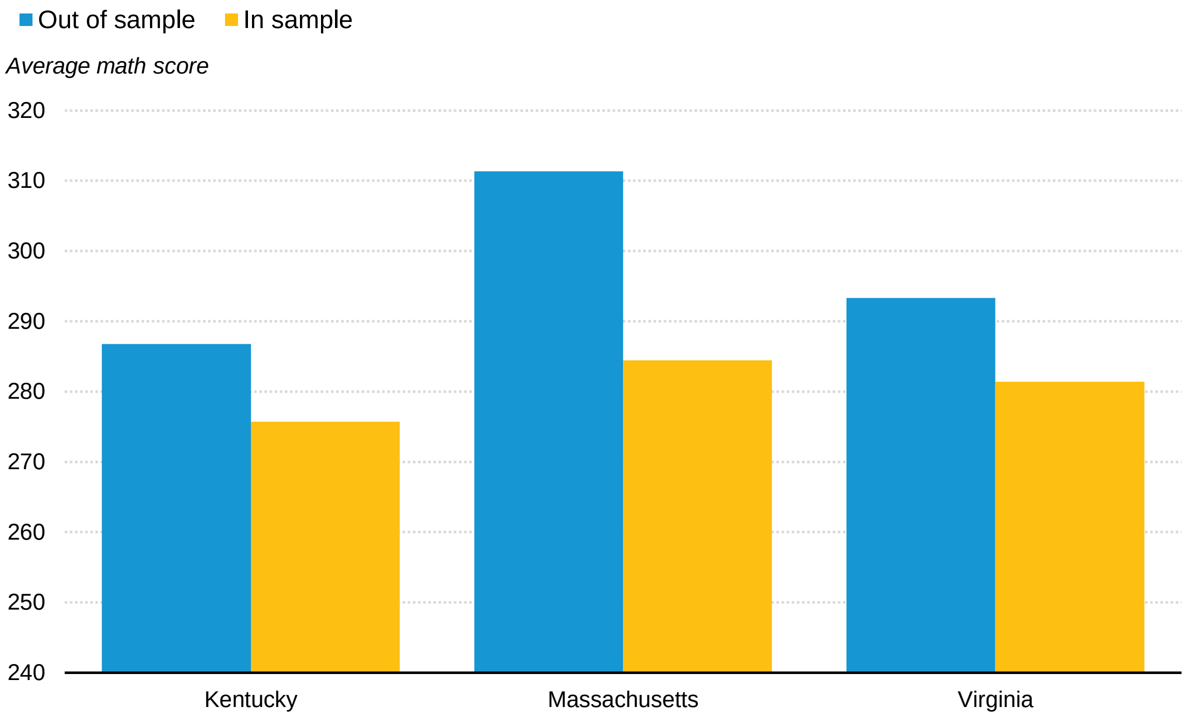
We include inputs at the student, school, and district levels. Student characteristics include race, gender, English language learner status, special education status, and cubics in both math and English language arts eighth-grade test scores. The inclusion of eighth-grade test scores is crucial, as it is important to not penalize high schools for incoming students' low levels of preparation or reward high schools for students who were already high achievers in eighth grade. Because eighth-grade assessments vary by state, we standardize them within each state. We then use state- and subject-specific National Assessment of Educational Progress (NAEP) means and standard deviations (interpolated for years when no NAEP assessments were given) to place them on a common scale. In other words, we take state-normed z-scores; multiply them by the state-, year-, grade-, and subject-specific NAEP standard deviations to obtain the correct distribution within each state; and then add the state-, year-, grade-, and subject-specific NAEP means to render cross-state comparisons meaningful.

Figures 2 and 3 compare these standardized eighth-grade test scores by state for schools in our sample (which serve historically underserved students) and schools out of our sample (which serve better-off students). In all states, particularly Massachusetts, the mean eighth-grade test score for the in-sample schools is less than for the out-of-sample schools, indicating that the in-sample schools tend to serve less-well-prepared students.

FIGURE 2

Eighth-Grade State Mathematics Assessment Scores

2012 cohort

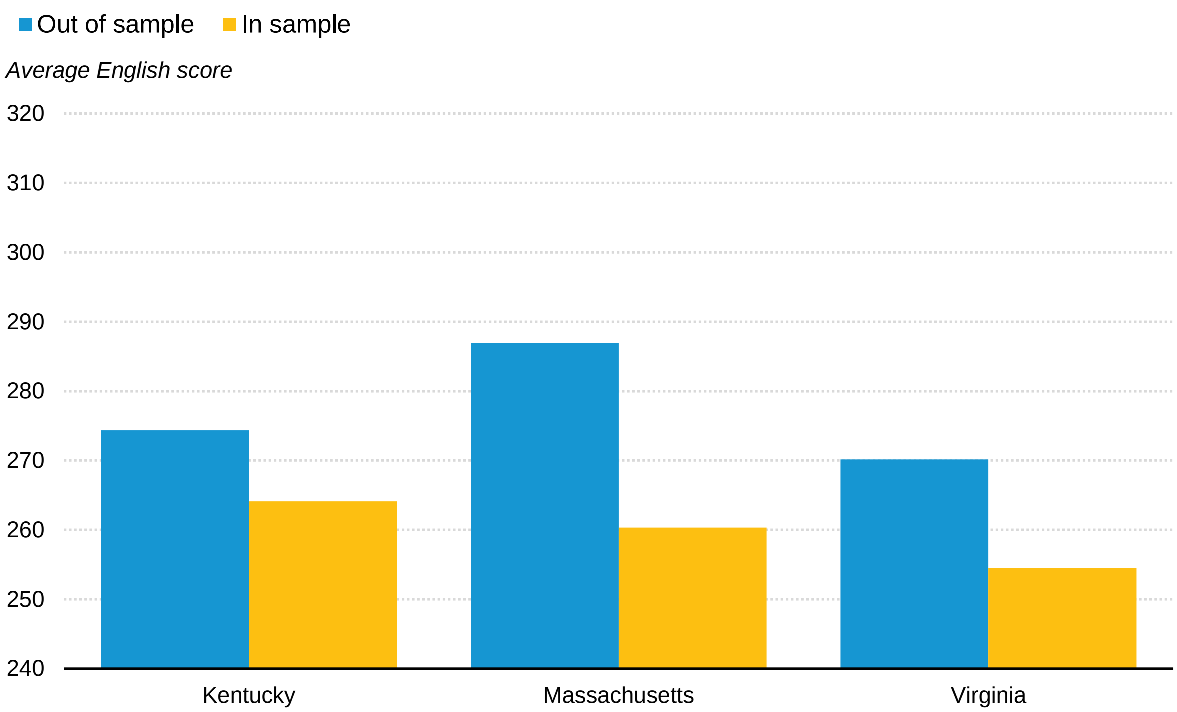


Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data and the National Assessment of Educational Progress.

FIGURE 3

Eighth-Grade State English Language Arts Assessment Scores

2012 cohort



Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data and the National Assessment of Educational Progress.

School characteristics include student-teacher ratio, a quadratic in enrollment, the share of students receiving free lunch, the share of students receiving reduced-price lunch, and the share of students that are black, Hispanic, Asian, or English language learners or enrolled in special education. Although two states provided student-level measures of economic need, the desire for consistency across all three states led us to include these measures at the school level instead. We include both individual-level and school-level measures of race or ethnicity and other student characteristics to reflect the fact that not only do a student's individual characteristics matter but so do those of her peers. For example, a school with a single English language learner may be different from a school with many English language learners, both for that student and others at that school. This is in keeping with the most recent literature (Totty 2019).

District characteristics include per student expenditures from the Common Core of Data (CCD). We also run the model without the student-teacher ratio or the district per student expenditures as a robustness check and obtain substantively identical results.

We do not include the following student-level inputs because of a lack of comparability across states: FRPL status (available in Massachusetts only; economically disadvantaged status is available in Virginia), homeless status (available in Virginia only), migrant status (available in Virginia only), and immigrant status (available in Massachusetts and Virginia only). Because of our model's specifications, we cannot include any school-level variables that do not vary by year, such as urbanicity, proximity to a college, or magnet status.

Model Results

Model results are displayed in table 2, with school fixed effects omitted for conciseness. Our model explains 23 percent of the variation in college enrollment, 16 percent of the variation in high school graduation, 19 percent of the variation in high school attendance, and 69 percent of the variation in standardized test scores. Students' eighth-grade test scores are the most important predictor, explaining 18 percent of the variation in college enrollment, 10 percent of the variation in high school graduation, 10 percent of the variation in high school attendance, and 65 percent of the variation in standardized test scores, when included as the only regressor (aside from year dummies).

Some of the results in the model have interesting values, such as the positive coefficients on English language learners and on students of color from all groups except American Indian or Alaska Native on three of the four outcomes. These come from the construction of the sample (which is the subset of schools serving low-income or historically underserved students), the controls for eighth grade test scores, and the controls for school-level demographics. When we run this analysis on all schools without pre-high school test scores and school-level control variables, the regression results look more similar to what is documented in other research, where and English language learners and students of color from all background except Asians tend to have worse educational outcomes on average.

TABLE 2

Regression Results

	College enrollment	High school graduation	High school attendance	Standardized test scores
Student-level variables				
Female	0.114***	0.034***	0.014***	-0.198
	0.002	0.002	0.001	0.170
Black	0.102***	0.062***	0.035***	-10.244***
	0.003	0.003	0.002	0.255
Hispanic	0.003	0.002	-0.018***	-7.195***
	0.004	0.003	0.002	0.298
Asian	0.124***	0.074***	0.058***	5.628***
	0.006	0.005	0.004	0.479
American Indian or Alaska Native	-0.003	-0.006	-0.002	-5.325**
	0.020	0.017	0.012	1.629
Native Hawaiian or other Pacific Islander	0.054	-0.006	0.006	-6.518*
	0.038	0.032	0.023	3.009
Two or more races	0.067***	0.030***	0.009*	-4.330***
	0.007	0.006	0.004	0.533
ELL	0.059***	0.053***	0.062***	-8.121***
	0.006	0.005	0.004	0.530
SPED	-0.047***	0.071***	0.034***	-16.038***
	0.003	0.003	0.002	0.295
Eighth-grade standardized test: English	-0.027***	-0.005*	0.010***	-0.493
	0.003	0.002	0.002	0.252
Eighth-grade standardized test: English, squared	0.000***	0.000***	0.000*	0.002*
	0.000	0.000	0.000	0.001
Eighth-grade standardized test: English, cubed	0.000***	0.000***	0.000	0.000
	0.000	0.000	0.000	0.000
Eighth-grade standardized test: Mathematics	-0.034***	-0.036***	-0.025***	-6.940***
	0.003	0.003	0.002	0.274
Eighth-grade standardized test: Mathematics, squared	0.000***	0.000***	0.000***	0.024***
	0.000	0.000	0.000	0.001
Eighth-grade standardized test: Mathematics, cubed	0.000***	0.000***	0.000***	0.000***
	0.000	0.000	0.000	0.000
School-level variables				
Student-teacher ratio	0.000	0.001*	0.001*	0.136*
	0.001	0.001	0.000	0.055
Share receiving free lunch	0.070*	0.076**	0.085***	7.818**
	0.033	0.028	0.020	2.749
Share receiving reduced-price lunch	0.387***	0.121	0.021	16.134*
	0.089	0.075	0.054	7.134
Enrollment (hundreds)	-0.006	-0.004	-0.001	-1.601***
	0.005	0.004	0.003	0.421
Enrollment (hundreds), squared	0.000	0.000	0.000	0.056***
	0.000	0.000	0.000	0.010
Black share[a]	-0.146**	0.058	-0.034	29.742***
	0.056	0.047	0.034	4.551
Hispanic share[a]	-0.188**	-0.038	-0.135***	11.539*
	0.068	0.057	0.041	5.374
Asian share[a]	0.075	0.160	0.067	41.786***
	0.136	0.114	0.082	10.759

	College enrollment	High school graduation	High school attendance	Standardized test scores
ELL share[a]	-0.029	-0.048	0.062	-10.546*
	0.064	0.054	0.039	5.156
SPED share[a]	0.017	0.003	-0.050	4.085
	0.050	0.042	0.031	4.125
District-level variables				
Per capita expenditures	0.000	0.000	0.000***	0.000
	0.000	0.000	0.000	0.000
Other variables				
Year 2010	-0.030***	-0.006	-0.005*	0.051
	0.003	0.003	0.002	0.285
Year 2011	-0.034***	0.007*	0.004	0.850**
	0.004	0.003	0.002	0.323
Year 2012	-0.050***	0.020***	0.013***	1.019**
	0.004	0.004	0.003	0.356
Constant	4.717***	3.078***	1.314***	1096.047***
	0.332	0.279	0.203	31.246
School fixed effects included?	Yes	Yes	Yes	Yes
N	180,196	180,196	175,267	141,389
R-squared	0.225	0.157	0.192	0.688

Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data.

Note: ELL = English language learner; SPED = special education.

a. The "share" variables are coded as 0.0-1.0, so that the increment on these independent variables in the regression runs from 0 percent to 100 percent.

$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

As we describe in the body of the report, the multistate model is constrained to treat all inputs the same across states. This may not be appropriate when working with diverse states. To explore how different the results appear when we run the model separately by state, we include results from separate-state models. Though we can build out the state-specific models to include additional covariates and broader outcomes, we constrain the state-specific models to the same specifications as the main model.

Because the state-specific models normalize the model within each state (i.e., the value-adds are all centered at 0 in each state), the results cannot be compared across states. Nonetheless, interesting patterns emerge when we compare each state's single-state model with the multistate model. Figure 4 shows how schools perform relative to what each single-state model would predict for college enrollment. The combined model suggests the expected college enrollment rates in each state have high overlap, but the state-specific models show low predicted college enrollment rates in Kentucky, high rates in Massachusetts, and rates roughly in the middle for Virginia, reflecting raw differences in underlying college-going rates by state. This might be because an important covariate, such as student

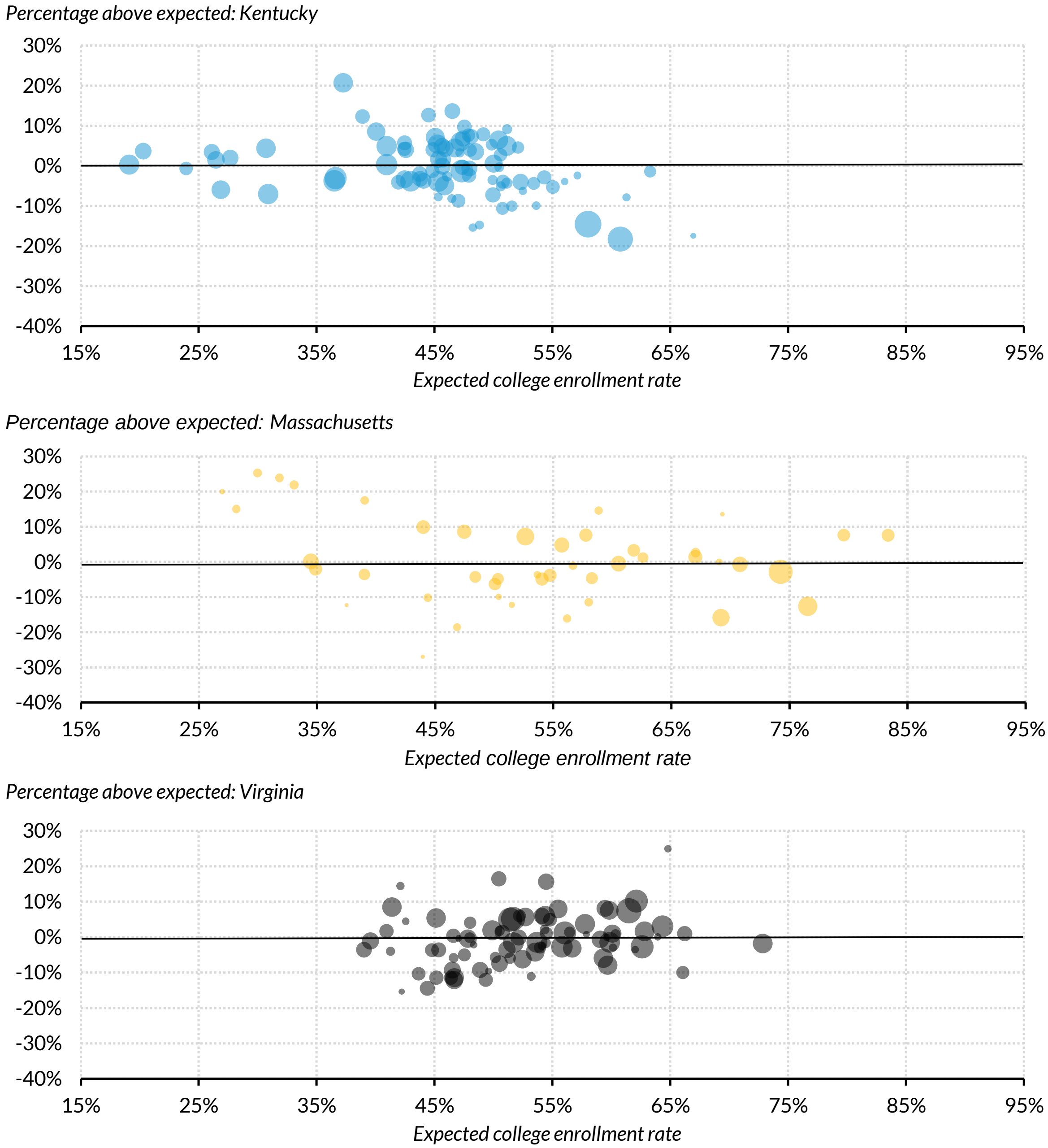
race or ethnicity, may operate differently across states in predicting college enrollment. For example, White (and Asian) students, on average, have higher college attendance rates than Black and Hispanic students, but compared with students in other states, White students in Kentucky have lower college attendance rates than White students elsewhere. As such, in the combined model, Kentucky is predicted to have higher college enrollment rates than it actually does. But how much the gap reflects omitted variables versus true differences in school quality across states is difficult to determine.

The states also have different patterns in the expected values of the three output indicators, especially high school graduation (figure 5) and attendance (figure 6). The expected values are more similar for tenth-grade test scores (figure 7), likely because the NAEP adjustment made the three states more comparable on this measure.

FIGURE 4

Percentage above or below Expected College Enrollment, by State and High School

2009–12



Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data and the National Assessment of Educational Progress.

Note: Bubble size indicates each high school's relative size.

FIGURE 5

Percentage above or below Expected High School Graduation, by State and High School

2009–12

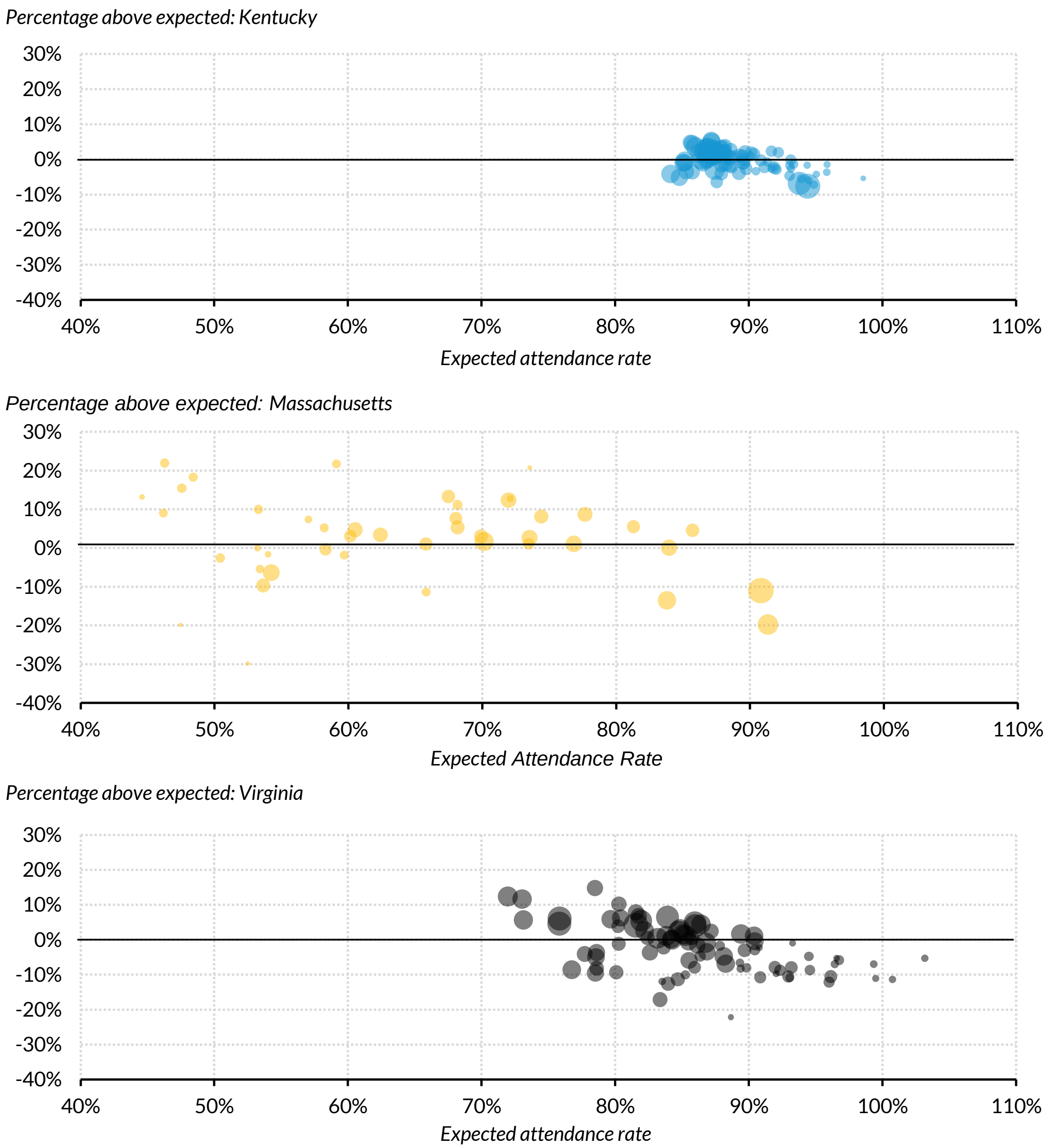
Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data and the National Assessment of Educational Progress.

Note: Bubble size indicates each high school's relative size.

FIGURE 6

Percentage above or below Expected High School Attendance, by State and High School

2009–12



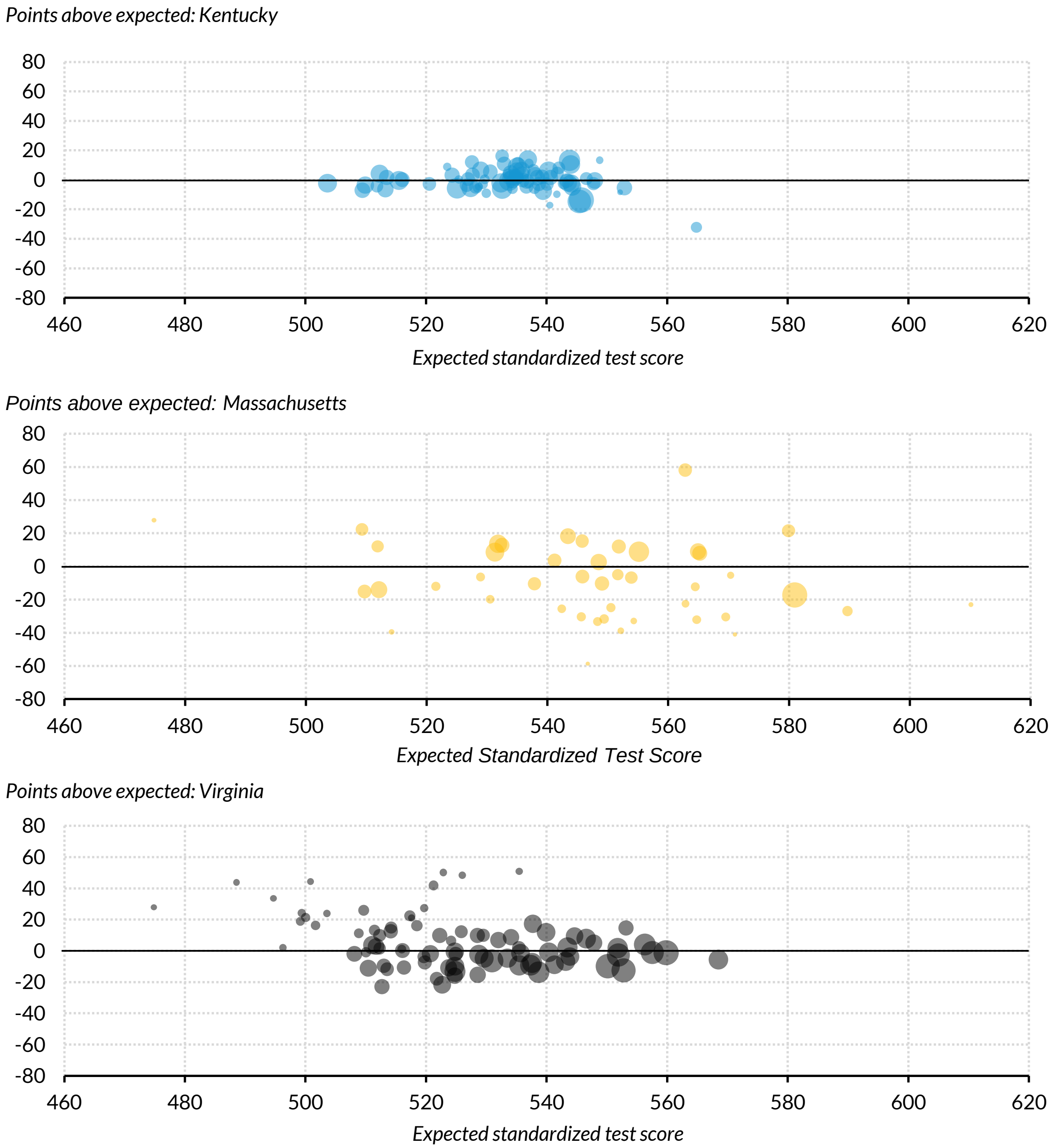
Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia and data from the Common Core of Data and the National Assessment of Educational Progress.

Note: Bubble size indicates each high school's relative size.

FIGURE 7

Points above or below Expected Tenth-Grade Standardized Test Scores, by State and High School

2009–12



Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia.

Note: Bubble size indicates each high school's relative size.

Notes

[1] These are “minorities” as defined in 20 U.S.C. § 1067k (2008).

[2] https://nscresearchcenter.org/workingwithourdata/

[3] Jordan Harris, “Kentucky Has Below Average ACT Scores, but There’s a Catch…,” Pegasus blog, August 15, 2018, https://www.pegasuskentucky.org/single-post/2018/08/15/Kentucky-has-below-average-ACT-scores-but-theres-a-catch-.

[4] “Policies: Tuition,” Kentucky Council on Postsecondary Education, accessed December 3, 2019, http://cpe.ky.gov/policies/tuition.html.

[5] “Historical ACT Percentiles for 2018, 2017, 2016, 2015, 2014, 2013, 2012, and 2011,” PrepScholar blog, January 27, 2019, https://blog.prepscholar.com/historical-act-percentiles-2014-2013-2012-2011.

References

Anderson, Theresa, David C. Blount, Constance A. Lindsay, Erica Blom, Semhar Gebrekristos, and Francisca Alba. 2019. "Robust and Equitable Measures to Identify Quality Schools." Washington, DC: Urban Institute.

Government Accountability Office (GAO). 2018. "Discipline Disparities for Black Students, Boys, and Students with Disabilities." GAO-18-258. Washington, DC: United States Government Accountability Office.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2015. "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?" Staff Report No. 710. New York, NY: Federal Reserve Bank of New York.

Totty, Evan. 2019. "High School Value-Added and College Outcomes." *Education Economics*.

Acknowledgments

This research was funded by grants from the Student-Centered Learning Research Collaborative, housed initially at Jobs for the Future and now part of KnowledgeWorks and the Nellie Mae Education Foundation. Supporting funders include the Barr Foundation, Carnegie Corporation of New York, and Oak Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

For more information on this project, see the main report.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.