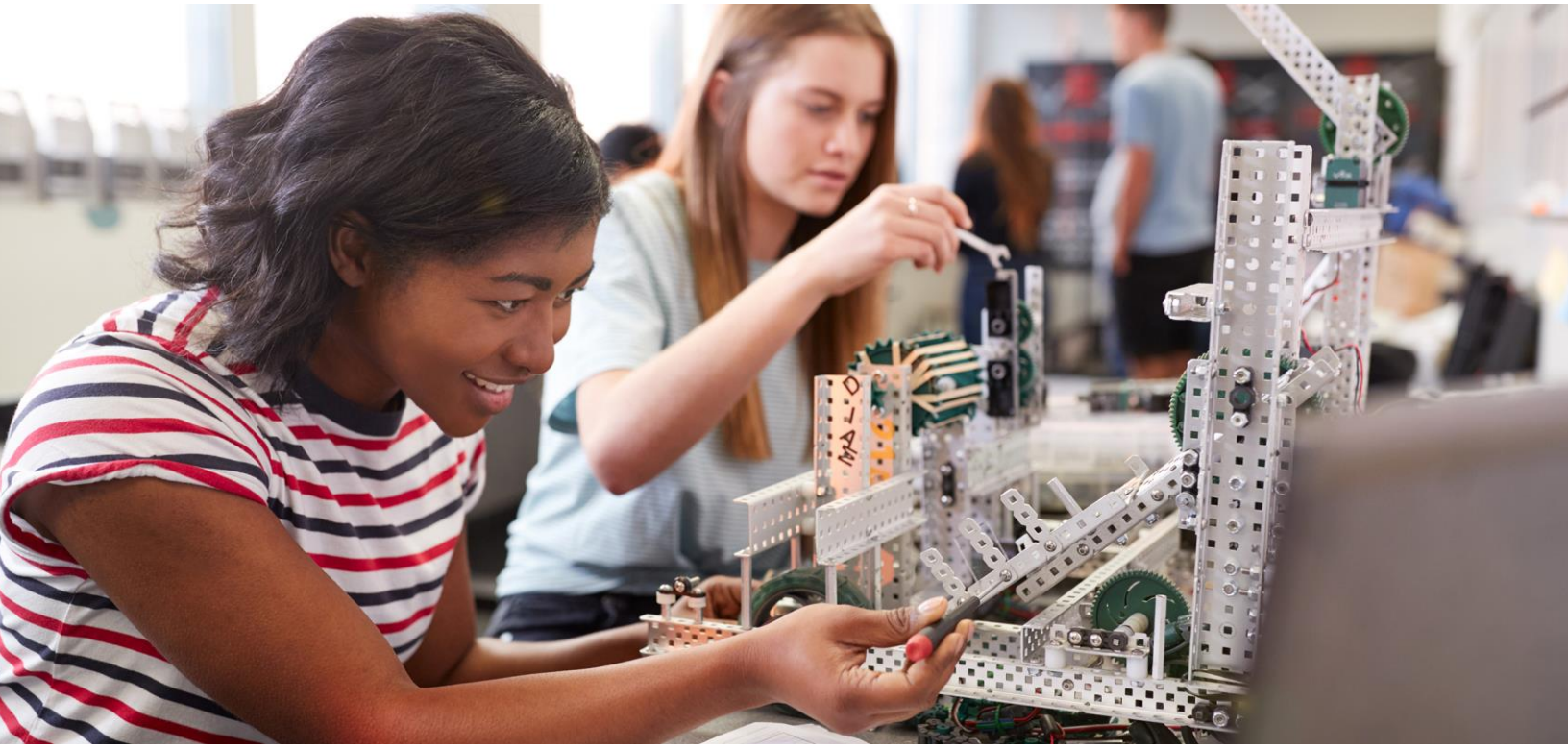


RESEARCH REPORT

Identifying High-Performing Schools for Historically Underserved Students

Exploring a Multistate Model

Theresa Anderson *Erica Blom* *Constance Lindsay*

with Semhar Gebrekristos, Macy Rainer, and Caroline Vilter

January 2020

URBAN INSTITUTE · ELEVATE · THE · DEBATE

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.

Contents

Acknowledgments

This report 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.

We would also like to thank Matthew Chingos for his consultation on this project, Jon Schwabish for his assistance with graphics, and Laura Greenback for her editorial support.

Executive Summary

This report describes the results of the quantitative filtering phase of the Robust and Equitable Measures to Inspire Quality Schools (REMIQS) project. This phase of the project seeks to identify regular public high schools that effectively serve students who have been historically underserved in the education system. It defines effectiveness by recognizing how students achieve along various success pathways and how schools support those successes. To learn more about the project's framework, see Anderson et al. (2019).

The goal for this phase was to combine data across states to develop a multistate model of school quality that would identify the best-performing schools in Kentucky, Massachusetts, and Virginia. We also share lessons from this work for those interested in undertaking or promoting similar efforts to expand the measurement of school quality with cross-state administrative data.

We find that

- a multistate model is possible but is limited by the comparability of data and by fundamental differences in state contexts,
- schools that "add value" based on traditional test score measures are not necessarily equally good at enrolling students in college, and
- generally, high schools that excel along one metric may not excel along all or even several metrics.

Following this project, which was designed to refine the REMIQS methodology, KnowledgeWorks will expand the REMIQS statistical analysis to include two more states. Once high-performing schools have been identified across those five states, KnowledgeWorks will work with researchers to investigate what those schools do to produce high achievement among their graduates. The effort aims to inspire other schools to adopt strategies that are particularly effective in helping students from historically underserved groups succeed after high school.

Identifying High-Performing Schools

Spurred by the Every Student Succeeds Act, districts and states have developed school report cards—usually centered around test scores—to represent school quality. Yet parents, policymakers, and other interested stakeholders need a more thorough understanding of school quality that includes a holistic vision of learning beyond test score achievement.

The REMIQS (Robust and Equitable Measures to Inspire Quality Schools) project seeks to identify regular public high schools that effectively serve students who have been historically underserved in the education system. It defines effectiveness by recognizing how students achieve along various success pathways and how schools support those successes.

A joint effort between the Urban Institute, KnowledgeWorks' Student-Centered Learning Research Collaborative, and multiple states, REMIQS uses state longitudinal data and other public data resources to identify authentic definitions of quality that go beyond test scores and graduation rates. REMIQS approaches school quality differently by

- focusing on long-term outcomes in addition to outputs,
- focusing on schools serving historically underserved populations,
- adjusting for student characteristics and pre–high school academic preparation, and
- combining data across states to create a multistate model.

The project is girded by an examination of schools as sites of social justice that can promote or undermine social equity for groups that have been historically underserved in public education. **Equity refers to an education system where every child develops the capacity to live a productive life. Achieving equity means closing opportunity gaps by helping students who are members of historically underserved groups achieve meaningful benchmarks by the end of high school (outputs) and realize success in achieving their goals after high school (outcomes).** A concept paper defining this work specifies critical concepts in more detail (Anderson et al. 2019).

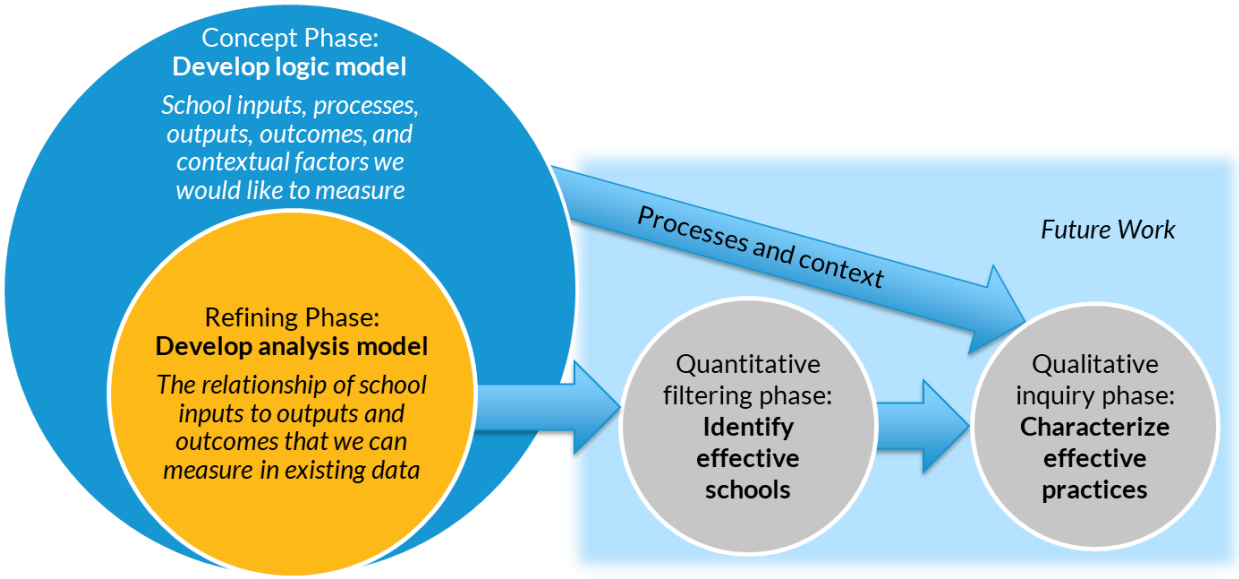
Overview

After the refining phase, which is the subject of this report, the REMIQS project will proceed through two additional phases: quantitative filtering and in-depth qualitative inquiry. Together, the phases will

identify schools that are doing particularly well serving students from historically disadvantaged groups compared with similar students at other schools. They will then investigate what is occurring in "high-flying" schools that contributes to their success. The goal is to inspire other schools to adopt strategies that are particularly effective in helping students succeed after high school. Figure 1 displays the relationships between the logic model from the concept phase (Anderson et al. 2019), the analysis model from this refinement phase, and the future quantitative filtering and qualitative inquiry phases of the project.

FIGURE 1

Phases of REMIQS and Models



This report describes the results of our team's and our partners' efforts to refine the quantitative filtering methodology. The goal was to combine data across three states to develop a multistate model of school quality that is capable of identifying the best-performing schools. We do not name individual schools in this report. Instead, we describe the distribution of schools in the three participating states—Kentucky, Massachusetts, and Virginia—along with the outputs and outcomes we could measure consistently and reliably across the states. The goal in this work was to refine the model, whereas the goals with subsequent phases of REMIQS will be to locate and then study those high-performing schools. To that end, we share lessons from this work for those interested in undertaking or promoting similar efforts to expand the measurement of school quality with cross-state administrative data. The separate appendix contains details about the data and model that may be of interest to a technical audience.[1]

We find that

- a multistate model is possible but is limited by the comparability of data and by fundamental differences in state contexts,
- schools that "add value" based on traditional test score measures are not necessarily equally good at enrolling students in college, and
- generally, high schools that excel along one metric may not excel along all or even several metrics.

Overview of the States, Data, and Samples

The data for this phase come from statewide longitudinal data systems (SLDS) in Kentucky, Massachusetts, and Virginia. These administrative datasets link individual student records across multiple state agencies. These states had accessible, high-quality longitudinal data systems and expressed an interest in making those systems available for this project. Each state contributed deidentified records from students' 8th-grade through 12th-grade years, their postsecondary or college activities, and other indicators of education and well-being, such as adult education and social assistance participation.[2] We supplement these data with publicly available school-level data from the Common Core of Data, accessed via the Urban Institute's Education Data Portal. Data covered the years 2008 through 2017.

Schools included in this study are regular public schools with more than 100 (cumulative) ninth-grade students in the 2009–12 entering classes in which

- at least 50 percent of students were American Indian, Alaskan Native, Black, Hispanic, or Pacific Islander[3] (students from historically underserved groups or populations); or
- at least 50 percent of students were eligible for free and reduced-price lunch (FRPL).

The separate technical appendix discusses the sample selection criteria in more detail. We identify 213 schools that fit these criteria in the three states (table 1).

TABLE 1

Open-Enrollment Public High Schools in the Study Sample

	KY	MA	VA	Total
Total open-enrollment public high schools	186	242	272	701
High schools with over 50% of students eligible for FRPL	86	41	41	169
High schools with over 50% of students from historically underserved groups	5	34	68	108
High schools with over 50% of students eligible for FRPL or from historically underserved groups	86	43	84	213[a]

Source: Statewide longitudinal data systems data in Kentucky, Massachusetts, and Virginia.
Note: Schools with large portions of students eligible for FRPL and schools with large portions of students from historically underserved groups are not mutually exclusive categories.
[a] Indicates final sample.

Logic Model Framework and Analysis Model

In the feasibility study for this project (Anderson et al, 2019), we identified the inputs, processes, outputs, outcomes, and contextual factors that define school functioning and quality and created a *logic model*, which serves as the framework for this study (figure A1). In the refining phase, we sought to collect as much administrative data from states as possible to develop the methodology to measure how much schools affect student outputs and outcomes while controlling for inputs and contextual factors outside their control. This is the *analysis model*, which tried to translate items from the logic model into a statistical approach to identify schools for the quantitative filtering phase. In the quantitative filtering phase, KnowledgeWorks and partners will identify which schools outperform expectations given their inputs and context, indicating that those schools may be doing something unusually effective with students. The qualitative inquiry phase will dig deeper into some of those schools to understand how they serve students and communities.

In the logic model, key inputs include school characteristics, student body composition, resources, teacher characteristics, and other contextual factors outside the direct control of the school within a school year. Many of the inputs and several contextual factors are measurable using administrative data. The final analysis model included the following controls.[4]

Student level:

- eighth-grade test scores in mathematics and English language arts
- race or ethnicity
- gender
- English language learner (ELL) status

- special education (SPED) participation

School level:

- school size
- share of student body eligible for reduced-priced lunch
- share of student body eligible for free lunch
- Black share
- Hispanic share
- Asian share
- ELL share
- SPED share
- student-teacher ratio

District level:

- per student expenditures

Outputs are the results of schooling that high schools can readily affect and that are apparent at the end of a school year or by graduation. Key outputs in the framework include high school engagement indicators; high school content knowledge; indicators of students being on track, being promoted, or completing a credential; college and career readiness; and students' short-term health and wellness. Many traditional measures of school quality focus on outputs, but the framework underlying REMIQS has a broader array of outputs of interest than conventional measures.

Outcomes are measures of well-being that may not become apparent until after a student has left high school. The high school may not have that much control over students' long-term well-being, so these are often weak performance measures. But a good high school can give students tools that will help them be successful in the long term, and we might be able to identify good high schools as those that have a disproportionate number of students realize long-term success. Outcomes of interest in the framework include postsecondary application, attendance, and success; job quality and earnings; deeper learning skills; long-term physical health and wellness; and civic engagement.

Within each output and outcome category, we identified measures in the framework. But few outputs and outcomes are measurable consistently across states. Ultimately, and as described in the

separate technical appendix, we reconciled longitudinal data across the three participating states to report on three measures in the analysis model:[5]

- Outcome: College enrollment, including most for-credit postsecondary programs
- Output: High school attendance in a student's last year of high school
- Output: High school graduation

We also looked at 10th-grade test score growth (relative to 8th grade) as a basis to compare these REMIQS measures with "traditional" measures of school quality.

The only outcome in this list, college enrollment, is our preferred measure because it is the best indicator of student well-being after high school available across states. But college enrollment does not guarantee long-term success (Chingos et al. 2019; Shapiro et al. 2018). Persistence, area of study, completion, employment outcomes, and other elements of well-being matter more. We could characterize persistence and completion for earlier cohorts of students, but there is a trade-off in the timeliness of the information for a given school. We also do not have labor market information, so we do not know if students who are not in college are attached to the workforce. To maintain a consistent analysis across the four student cohorts and the three states, we use college enrollment as our best outcome in the current study.

The logic model (Appendix A) was aspirational, while the analysis model is constrained by data and logistical considerations. We had to exclude control variables that were not available in the administrative data, could not be reconciled across states, or would have been particularly cumbersome to process. The tables in Appendix B summarize the reasons that inputs, outputs, and outcomes in the logic model were included/excluded from the analysis model. (Note that the processes are not part of the analysis model because those will be explored in detail in the qualitative inquiry phase.) As these tables show, some data items that we attempted to analyze raised equity issues (e.g., law enforcement referrals, suspensions, and expulsions), and so we did not include them. In many cases, items were either not available at all in the data or were not consistently available across states.

The biggest gap in meaningful outcomes stems from the unavailable employment/earnings data. These data cannot be released at the individual level in Kentucky per state law and the requests in the other two states could not be fulfilled in time for this analysis.[6] If we had information about employment/earnings, or if future research efforts could obtain those data, it would fill critical gaps in our understanding of wellbeing in the longer term.

Findings

The analysis model indicates how well students from each school do on key outputs and outcomes relative to what one might otherwise expect, given their preparedness at high school entry (based on eighth-grade test scores[7]) and other characteristics. This performance is akin to a growth model because it does not look only at schools that have strong outcomes but accounts for the student population, student academic readiness (through eighth-grade test scores), and other factors. Schools in which students outperform expectations after controlling for all the inputs and conditions are high flying. Some schools fly higher than others. We identify high-flying schools as those in the top 25 percent of all schools across the three states, which are the top 53 of the 213 schools in the sample.

Throughout this discussion, we use the term "expected" and "expectations" frequently. These terms reflect the results of the statistical analysis in the model and are not meant as normative statements. That is, given the characteristics of students and schools in the model, how well would we *statistically* expect students to fare in the short and long term? Importantly, as we will show later when we do separate state-specific models, the "expected" outcome depends on which schools and students are included in the model. If schools do better than the model would expect given the inputs, then there is likely something in the way the school operates that helps students do better than other students with their same characteristics in similar types of schools (in terms of size, funding, and the other school characteristics the model). That is why the model only looks at inputs and outputs/outcomes; the processes in the middle are the subject of future qualitative work that will explain *why* some schools over-perform. However, an important caveat is that if there are other inputs or explanatory factors that are out of the school's control but are not in the analysis model because of data limitations, the results of the statistical analysis may be biased. In that case, some schools might look like they are doing particularly well when it is really because those schools have other advantages not controlled for in the analysis. This is a challenge inherent to most statistical analysis, and it is the reason we have been careful to include as many inputs as possible within the constraints of the data. Due to the design of our model, school-level factors that do not vary over time, such as urbanicity, magnet status, or state context, cannot be included.

Test score growth measures are the "traditional" indicators of school quality (Anderson et al. 2019). We developed a growth measure with 10th-grade standardized test scores to demonstrate how schools would perform on a traditional metric. We then show how schools do on the other result measures, paying the most attention to our only outcome measure, college enrollment. The two output measures—high school attendance and high school graduation—provide additional information about

schools that emerge as high flying by helping students outperform what would otherwise be expected based on individual and institutional characteristics.

Test Score Measures

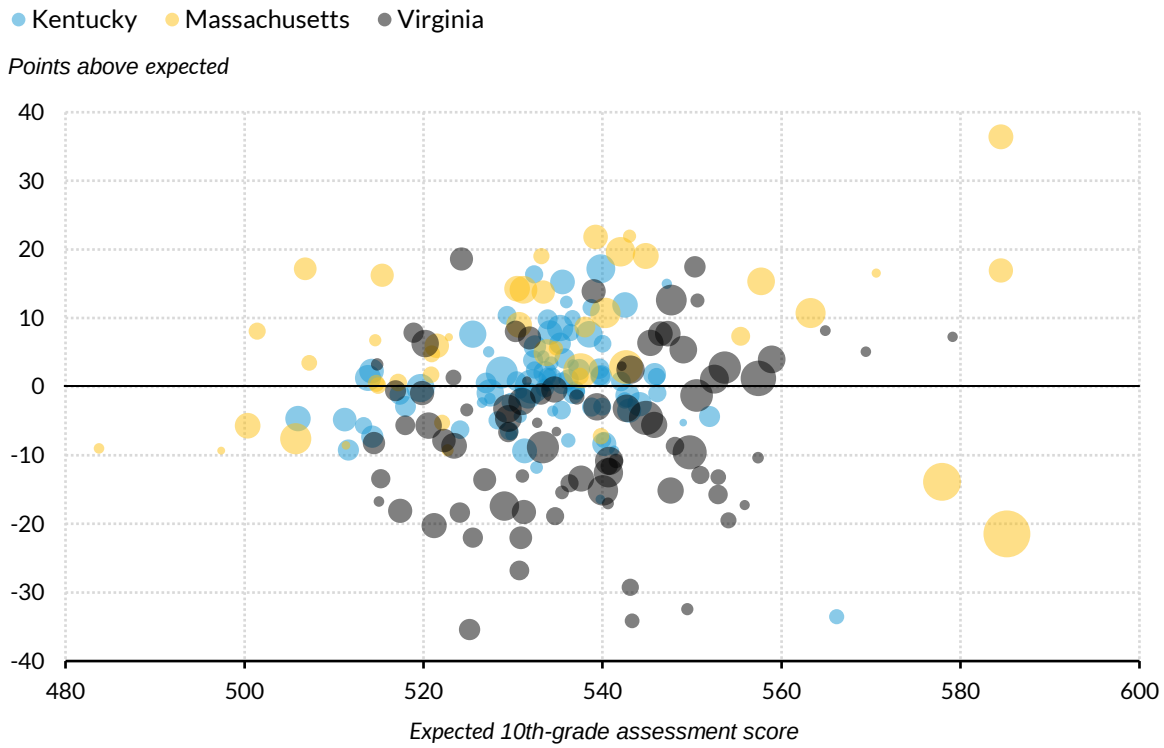
The two test score growth measures are in math and reading. But because performance on subtests is not our focus, we combine them into one test score metric for simplicity.[8] Figure 2 shows how schools perform on 10th-grade test score growth measures across the states. Bubbles further to the right indicate schools that are expected to have higher test score growth than those further to the left. The distance of schools from the horizontal line shows how much these schools exceeded (or fell short of) expectations. The bubble size reflects each school's relative size. In a traditional school quality growth model, the top 25 percent of schools on this metric would be high flyers. Half the high-flying schools based on 10th-grade assessment scores are in Massachusetts, with a little more than a quarter in Kentucky and a little less than a quarter in Virginia.

Schools' growth scores are comparable across states if the model inputs capture differences in student characteristics across states and if the relationships between the inputs and outcomes are the same across states. The data must be genuinely comparable, which we have tried to ensure by, for example, using nationally comparable NAEP scores to align state assessments to a common scale. However, short of having a national assessment that students in all states take, we cannot guarantee that this adjustment is perfect.

FIGURE 2

Points above or below Expected 10th-Grade Assessment Scores, by High School

2009–12



Source: Analysis of Common Core of Data and statewide longitudinal data systems data in Kentucky, Massachusetts, and Virginia.
Note: Bubble size indicates each high school's relative size.

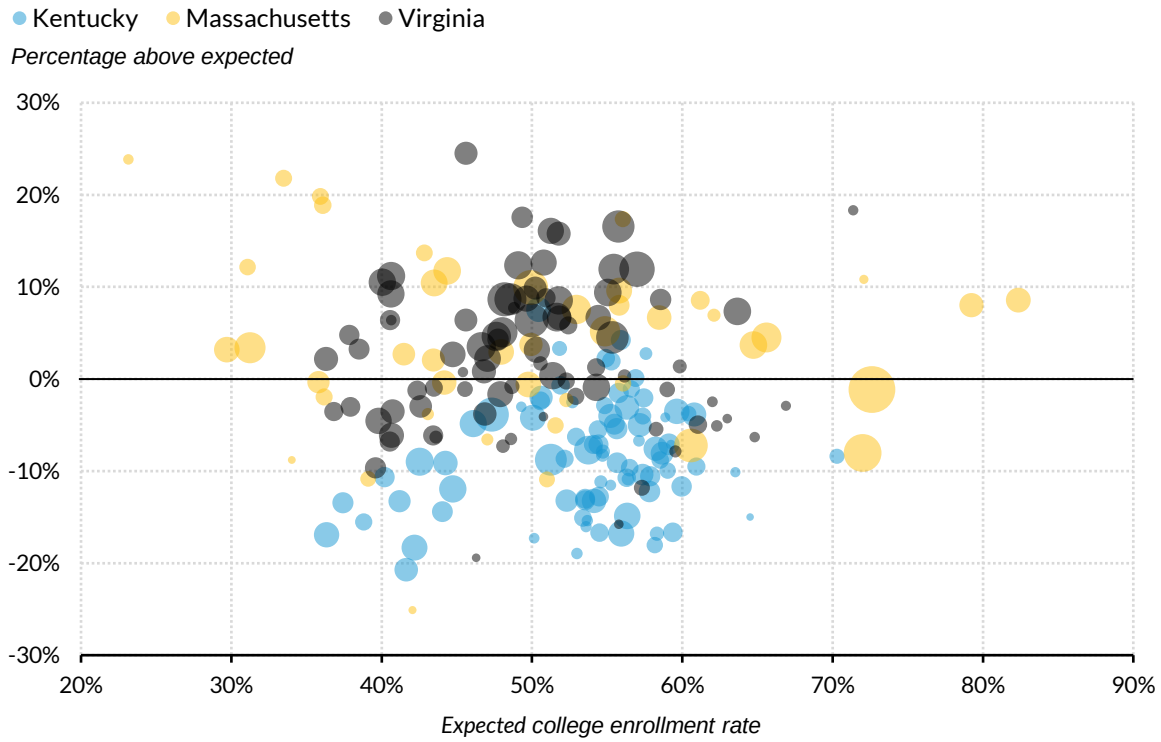
REMIQS Measures

The story changes when we look at high-flying schools based on our primary outcome measure, college enrollment. Figure 3 shows the distribution of schools by how much they promote college enrollment relative to what we would otherwise expect. Some schools in Massachusetts increased college enrollment 30 percentage points above what we would otherwise expect, given student, school, and district characteristics. Schools expected to have students enroll in college at higher rates (those further to the right) are limited in how much they can outperform what is expected, since graduation rates cannot exceed 100 percent. Therefore, it is not surprising that schools further to the right are closer to the horizontal line.

FIGURE 3

Percentage above or below Expected College Enrollment, by 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.

This measure shows important and interesting cross-state differences in how much schools boost college enrollment. Virginia has the highest number of high-flying schools on college enrollment, accounting for 32 of the top 53 schools (60 percent); 20 are in Massachusetts, and one is in Kentucky. Given the model's constraints, it is impossible to determine whether this is caused by true differences in average school quality across states, rather than by differences in students' underlying probability of going to college, which could be driven by state policies, geography, or other factors we do not account for.[9] Certainly, if different states were included in the model, this figure would look different.

A separate issue is that the multistate model constrains all the inputs to affect the outcomes in exactly the same way across states. For example, the relationship between free lunch status and going to college is assumed to the be same in Kentucky, Massachusetts, and Virginia, which may not be the case.

Arguably, the differences in state contexts, which cannot be adequately captured in a cross-state model, suggest running state-specific models. Figure 4 shows how schools would perform on separate state models of college enrollment. When analyzed through separate models, the amount by which high schools exceed expectations is centered at zero in each state (rather than across three states). The state-specific estimates are not comparable across states because separate models allow the roles of FRPL, race or ethnicity, and other inputs to vary by state. For example, White students might generally have better college-going outcomes on average, largely because of structural inequities, in racially diverse states like Massachusetts and Virginia. However, being White might not be a good predictor of college-going in a less racially diverse state like Kentucky. If all states are in the same model, the large portion of White students in Kentucky might make it look like the student population there has a higher rate of expected college enrollment than makes sense for the state context. In the separate models, the way that the independent variable of student race predicts the outcome of college enrollment is allowed to differ, but this means that the estimated relationships are not comparable across states. We show these scatterplots for readers who are more interested in comparisons within (rather than across) states, and we present additional state-specific results in the separate technical appendix.

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.

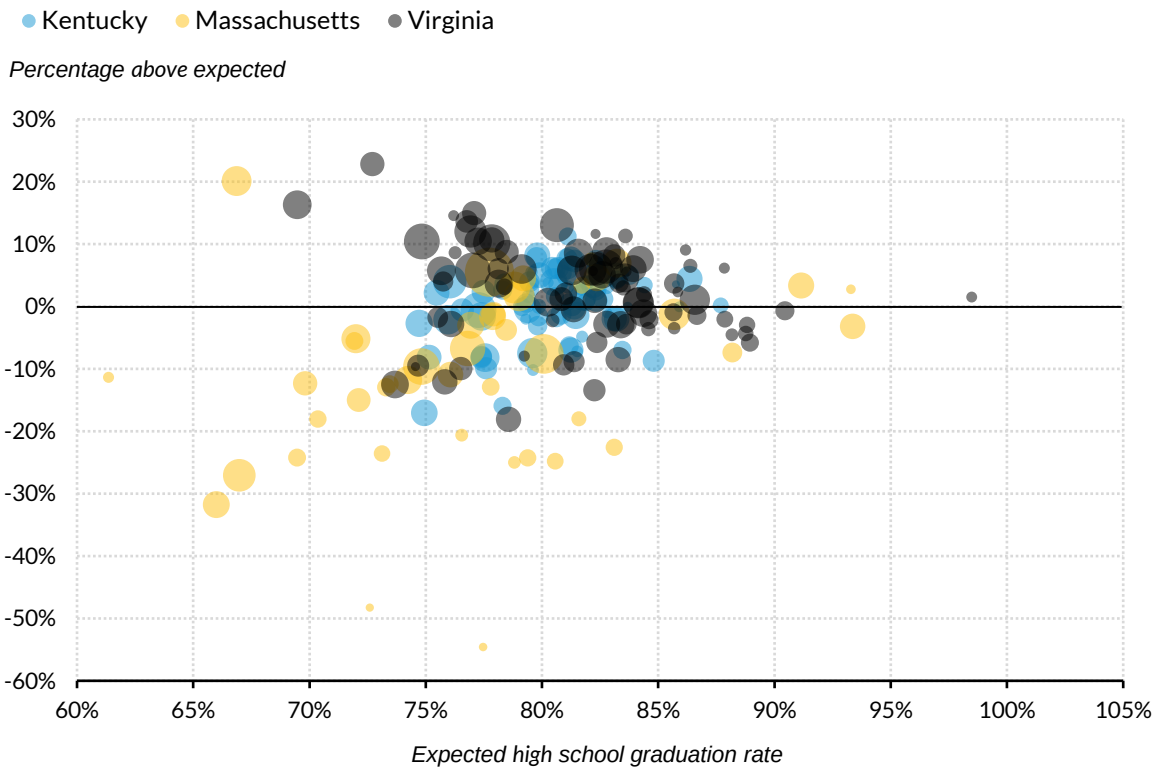
Note: Bubble size indicates each high school's relative size.

Turning back to the multistate model, we can see how schools perform on the output measures. High school graduation is the next-strongest measure after college enrollment because it varies substantially across schools and has consistent data across states. Figure 5 shows how schools perform on helping students graduate from high school at higher rates than might otherwise be expected. States are more evenly distributed above and below the line, though Massachusetts and Virginia have some outlying schools.

FIGURE 5

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

2009–12



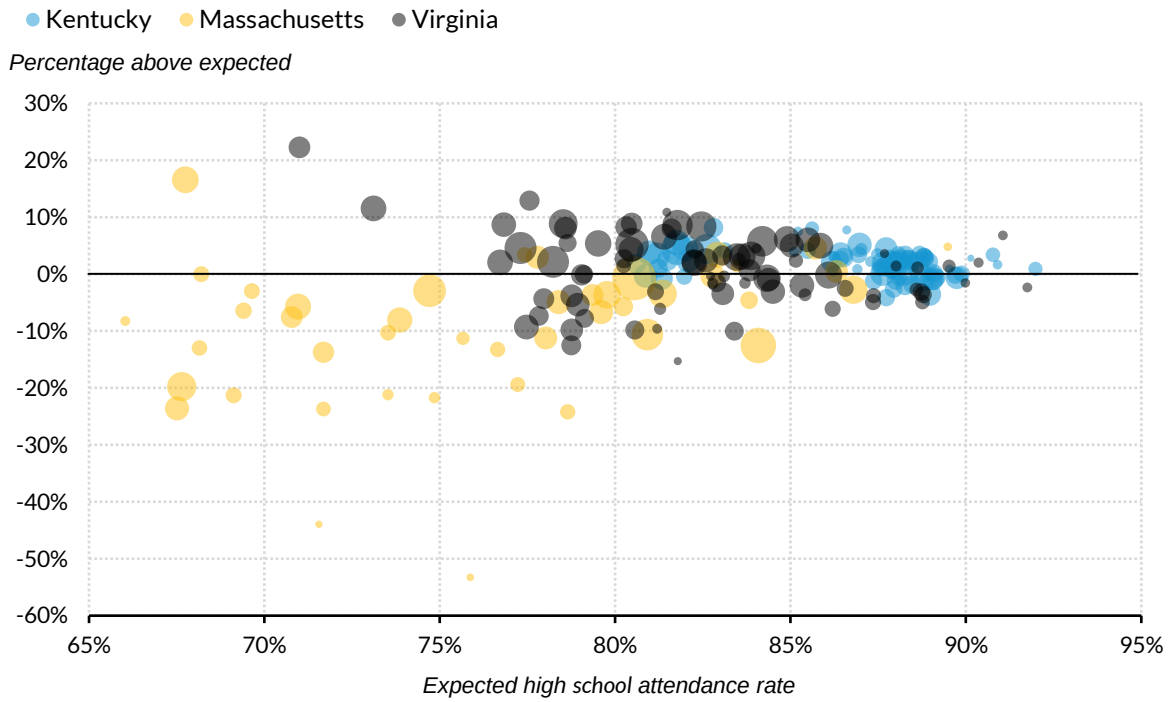
Source: Analysis of Common Core of Data and statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia.
Note: Bubble size indicates each high school's relative size.

Twelfth-grade attendance is also a consistent measure across states, but it is limited because student attendance among those who are still enrolled tends to be high, averaging 94 percent. Our attendance measure also accounts for students who dropped out of high school and who are coded as having 0 attendance, which means this indicator reflects a school's ability both to retain students and to motivate those who are still enrolled to attend each day. Massachusetts has numerous low outliers (figure 6).

FIGURE 6

Percentage above or below Expected High School Attendance Rate, by 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.

Looking across Multiple Measures

The distribution of schools along each measure tells one element of the REMIQS story. The next step is to see how well the measures correlate with each other and how schools stack up across multiple measures. This will allow us to identify schools that rise to the top across multiple measures and understand if each measure is telling us something different about how a school functions.

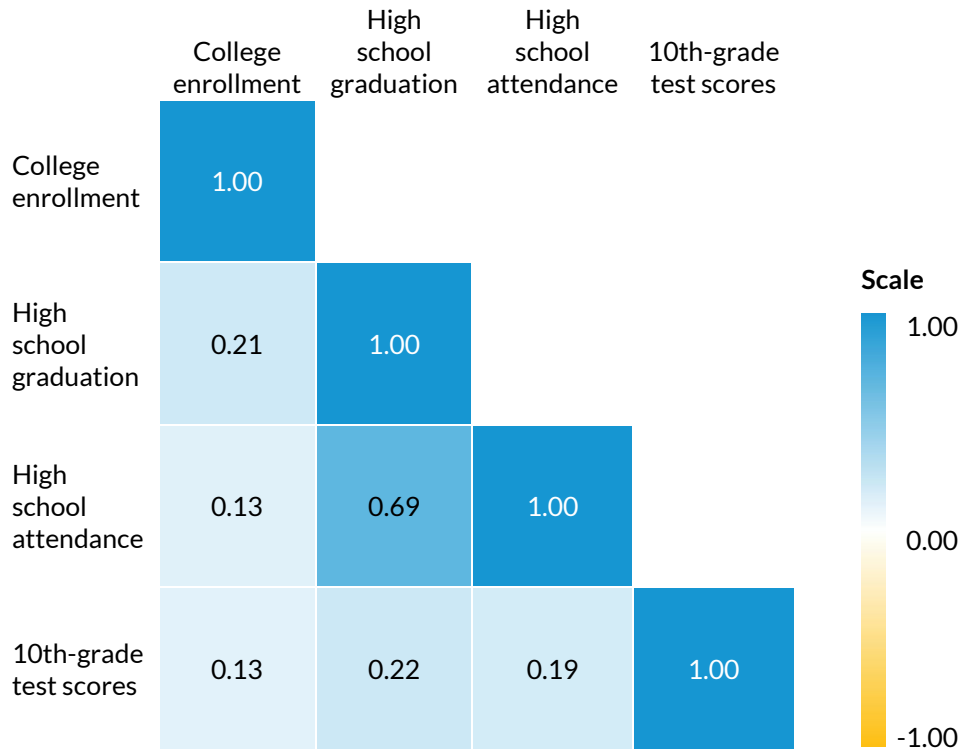
CORRELATION AMONG MEASURES

We begin with correlation, how well a school that performs highly on one measure performs on other measures. These represent improvements high schools achieve above what would otherwise be expected, or the "value-add." School rankings on each measure are not highly correlated.[10] This demonstrates that the measures in this analysis represent different concepts, which makes these measures more robust. But it also means school quality is not a monolith. Schools that see gains on one metric do not necessarily see gains across other metrics. Figure 7 shows the correlations between each

pair of measures, with color coding. High school attendance and high school graduation are the highest correlated, at 0.69. This means that for every standard-deviation increase in attendance a high school promotes for its students, that relates to a 0.69 standard-deviation increase in the amount by which the school improves graduation rates. No measures are negatively correlated, but some have low correlations.

FIGURE 7

Correlation among High School Improvements in Student Results



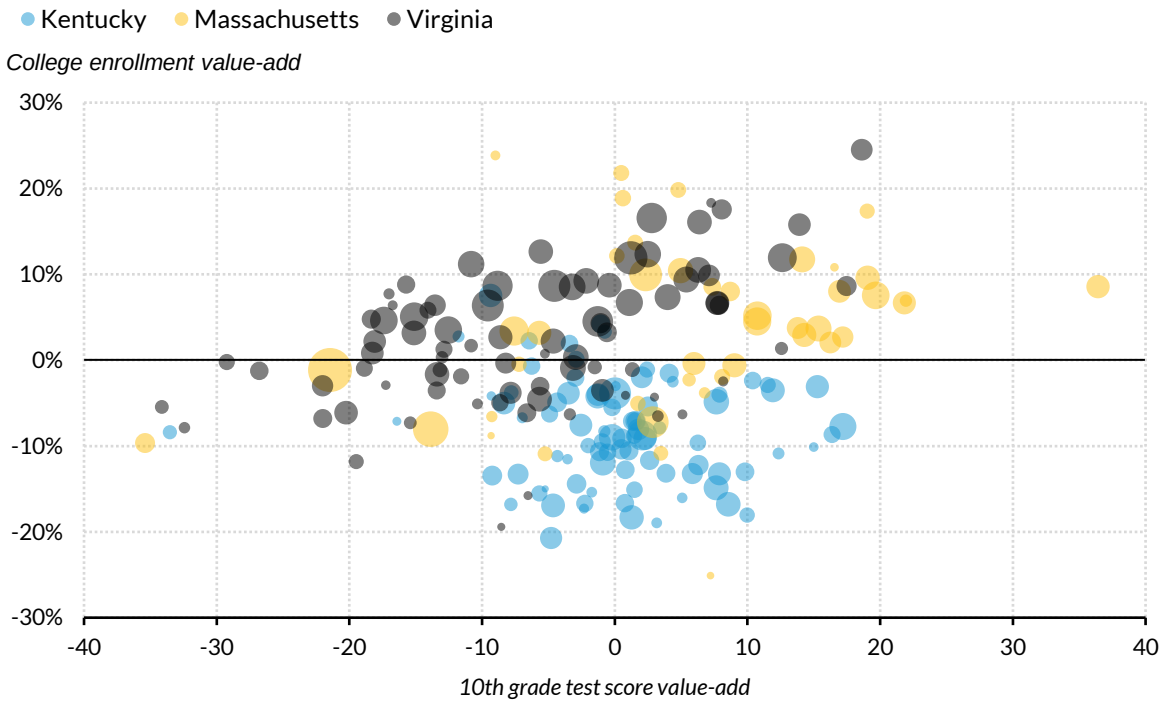
Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia.

The correlation between 10th-grade test scores and college enrollment—our "traditional" quality measure and our best outcome measure—is positive but low, at 0.13 (figure 7). This means that *schools that are good at raising test scores are not necessarily the same schools that are good at preparing students to enroll in college*. To illustrate this point further, figure 8 shows a scatterplot of the correlation between college enrollment value-add and 10th-grade test score value-add. While there is generally a positive relationship between a school raising college enrollment and that same school raising test scores in

Massachusetts and Virginia, as shown by the upward slope of the scatter plots in those states, these two measures are basically unrelated in Kentucky.

FIGURE 8

Correlation between College Enrollment Value-Add and Test Score Value-Add



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

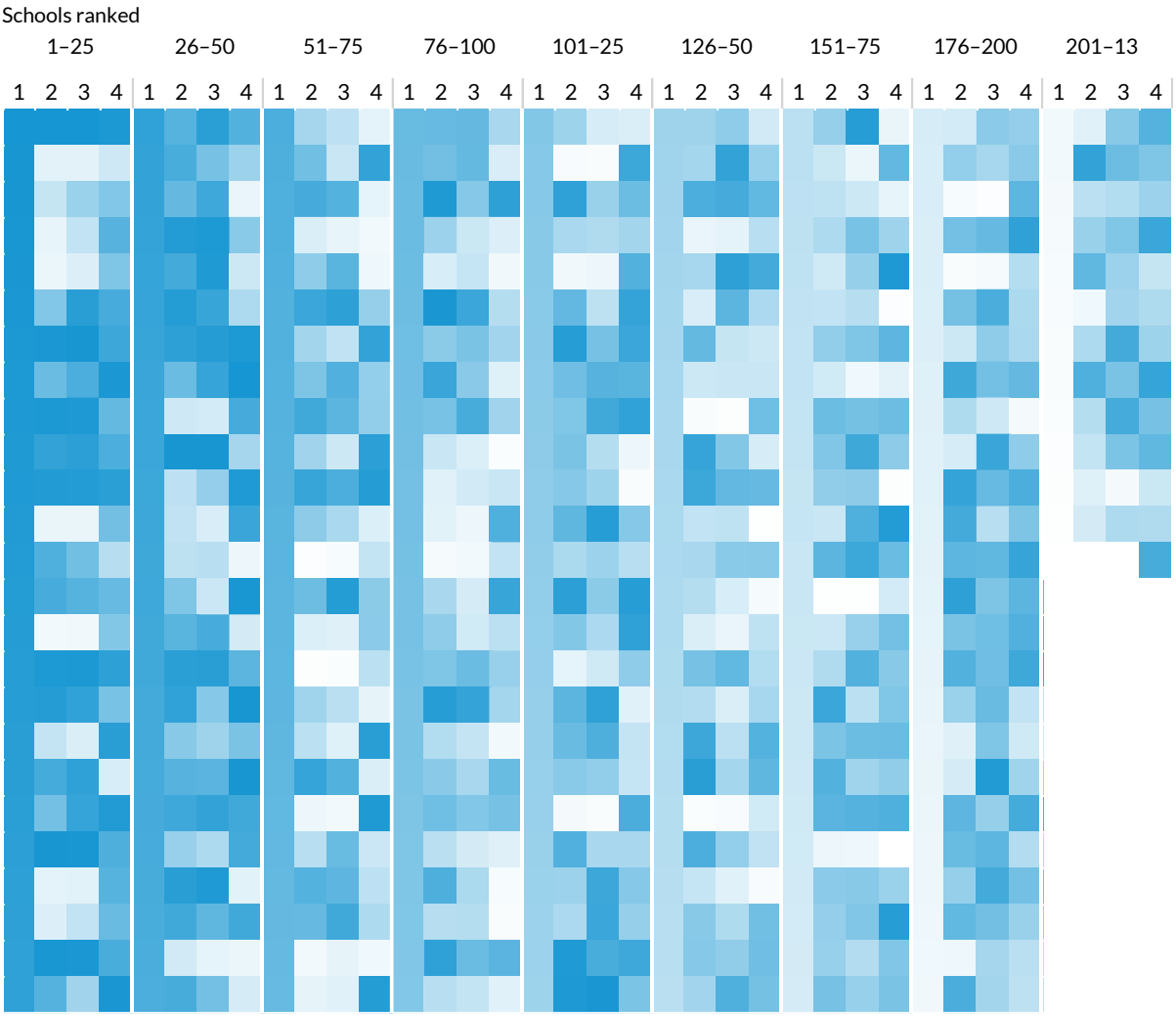
Schools that are good at raising test scores are not necessarily the same schools that are good at preparing students to enroll in college.

Similarly, although the correlations of college attendance gains with high school graduation gains are positive, they are not large. Again, this indicates that helping students finish high school is not enough, at least in terms of ensuring college enrollment. As noted in the introduction and in past work (Anderson et al. 2019), we believe outcomes other than college enrollment matter, but the current study cannot assess employment, wages, military enlistment, or other life outcomes consistently across students and states.

Since college enrollment is our primary measure, it is informative to look at how schools rank on gains in other measures relative to their college enrollment gain ranking. Figure 9 is a heatmap, where schools are sorted by their ranking in college enrollment gains in the left-hand column. The next three columns show shades of blue representing how schools rank on the other measures, with darker blue indicating they are more highly ranked. Because there are so many schools, the four columns are repeated for every subsequent 25 schools. This heatmap shows that schools that rank high on college enrollment do not necessarily rank high on other measures.

FIGURE 9

Heatmap of High School Rankings, Ordered by College Enrollment Rankings



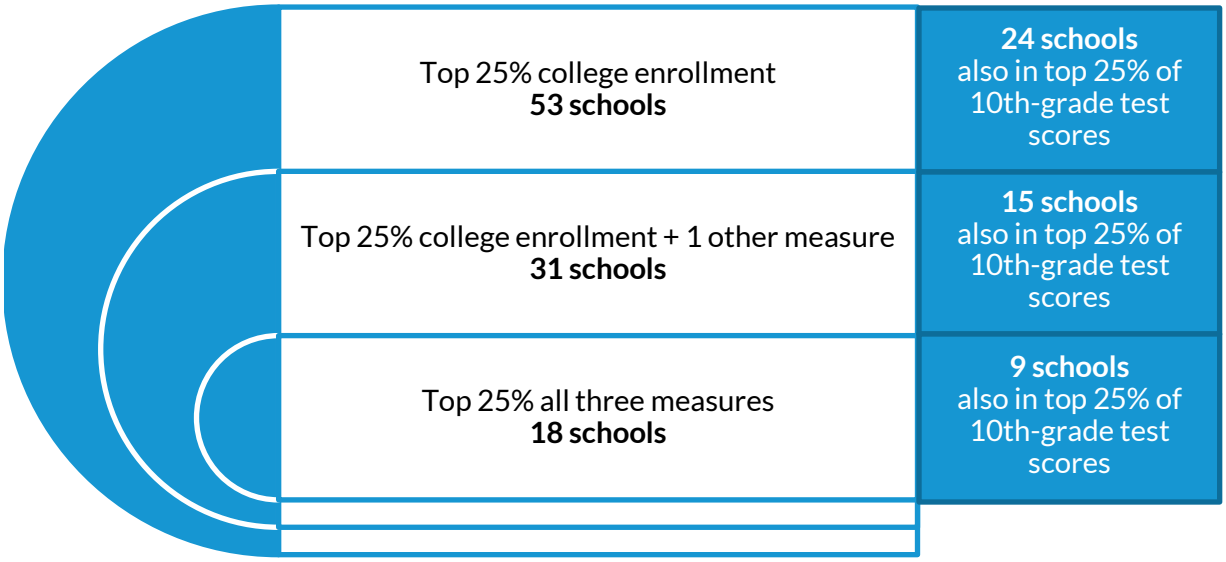
Source: Analysis of statewide longitudinal data system data in Kentucky, Massachusetts, and Virginia.
Notes: 1 = college enrollment; 2 = high school graduation; 3 = high school attendance; 4 = 10th-grade test scores. Darker blue indicates schools that are more highly ranked.

SCHOOLS THAT SCORE HIGH ON MULTIPLE MEASURES

Even though the measures are not highly correlated with each other overall, some schools stand out on multiple measures. Figure 10 shows the number of schools that rank in the top 25 percent for gains in college enrollment and the other REMIQS measures. It also notes which of those schools also score in the top 25 percent of test score rankings.

FIGURE 10

High Schools Scoring in Top 25 Percent on Multiple Measures



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

Of the schools in the top 25 percent across multiple measures, none are in Kentucky because only one Kentucky high schools had a top ranking on our primary measure of college enrollment. Moreover, all of the 18 schools that ranked highly on all three REMIQS measures were in Virginia. Six of these schools also scored high on test score growth.

Discussion and Conclusions

This work contributes several significant conclusions. First, it is possible to construct a multistate model with school quality measures beyond test scores. This model is important for extending the quality conversation beyond test score measures that have inherent drawbacks (Anderson et al. 2019).

Further, the analysis model developed for this project will inform the next stages of REMIQS research: quantitative filtering of schools and an in-depth qualitative investigation of schools that rise

to the top to characterize what sets those schools apart. The integration of multiple data sources—state longitudinal data from eighth grade through college and federal publicly available data—to locate schools where long-term outcomes are most positive and equitable offers the field potentially powerful information about what drives effective schools. This can inform the next stages of the REMIQS project as well as similar efforts to inspire schools to adopt practices that help historically underserved students achieve success after high school.

The rankings produced from this effort may also help state policymakers focus on promising aspects of schools that might not otherwise have risen to the top in existing state rankings. Though the ranking will not be made public, identification of the top-performing schools could spur insights within states about what makes a school effective. Many policymakers are looking for new, meaningful measures of school quality as they implement the state plans developed under the Every Student Succeeds Act.

Drawbacks of Multistate Models

This effort previews a new level of analysis that may be possible with increasingly sophisticated longitudinal student tracking, particularly across states. It highlights the advantages of this approach but reveals lessons about the drawbacks of constructing multistate models.

First, a multistate model forces the research to rely on common variables available across states, limiting the richness and breadth of the analysis. Even variables that looked like they would be comparable across states were not always readily reconciled to inform the same model. Individual states also provided rich data on additional inputs and outcomes, but we could use only what was available across all three contexts. This problem would be magnified if we included more states in the analysis.

Second, if it is not possible to control for all factors that make students, schools, districts, and state contexts different, school rankings could still reflect elements outside the school building. For example, the differences in college enrollment among the three states we present are not likely explained wholly by differences in school quality. Instead, there are probably differences in college access, state policy, employment opportunities, student characteristics, or other factors this model cannot account for, either for lack of data or because the context does not vary for schools across the analysis years.

If the analysis added more states, the model's strength would improve. As a trade-off, the number of comparable inputs and outcomes could diminish. For a small number of diverse states, a more credible approach may be to separate the analysis out by state.

Additional Desirable Variables

Some desirable data items were unavailable for this analysis, many of which statewide longitudinal data systems do not track. Some of the specific measures that would have enhanced this work include the following:

- **An individual-level measure of student poverty.** Given the introduction of the Community Eligibility Provision, FRPL as a school-level measure of poverty is becoming less meaningful.[11] A better measure is the share of students directly certified (Greenberg 2018). It would be better to have this measure at the individual level. An even more nuanced approach would include a nonbinary measure of family resources, such as family or household income. Even richer data would consist of household structure and parental educational attainment, but these are not currently available in statewide longitudinal data systems, to our knowledge.
- **Better college-going and completion data.** A further step in this analysis would be to measure college persistence and graduation to ensure high schools are not merely enrolling students in college but setting them up to succeed once there.[12]
- **Individual wage data.** Although many statewide longitudinal data systems link to the US Department of Labor, these wage data have many shortcomings. In some instances, they are available to researchers only at the aggregate level. In other cases, they are available only for students who remain in state, and they typically report only quarterly earnings and industry, not hours or occupation. Initiatives such as the Western Interstate Commission for Higher Education's multistate SLDS[13] or the Census Bureau's Post-Secondary Employment Outcomes project[14] could help bridge this gap. Even more ideal would be to include household income, rather than individual income, to account for family financial well-being.[15]
- **Comparable state assessments.** The largest factor in our model is the eighth-grade standardized test scores, accounting for about two-thirds of the individual variation in going to college. But each state administers its own exam, and these are not directly comparable. Although we made a concerted effort (using nationally comparable NAEP scores) to place these on an equal scale, differences remain.

- **Other measures of long-term well-being.** Administrative data do not capture such outcomes as socioemotional learning, mental and physical health, life satisfaction, and civic participation. Creative data wranglers might be able to match publicly available voter participation records to student records, but this would require SLDS administrators (who would be tasked with the actual matching) to be willing to take on this task. Enrollment in social safety net programs, an outcome we do not examine here, is also theoretically possible, but those data are not public.

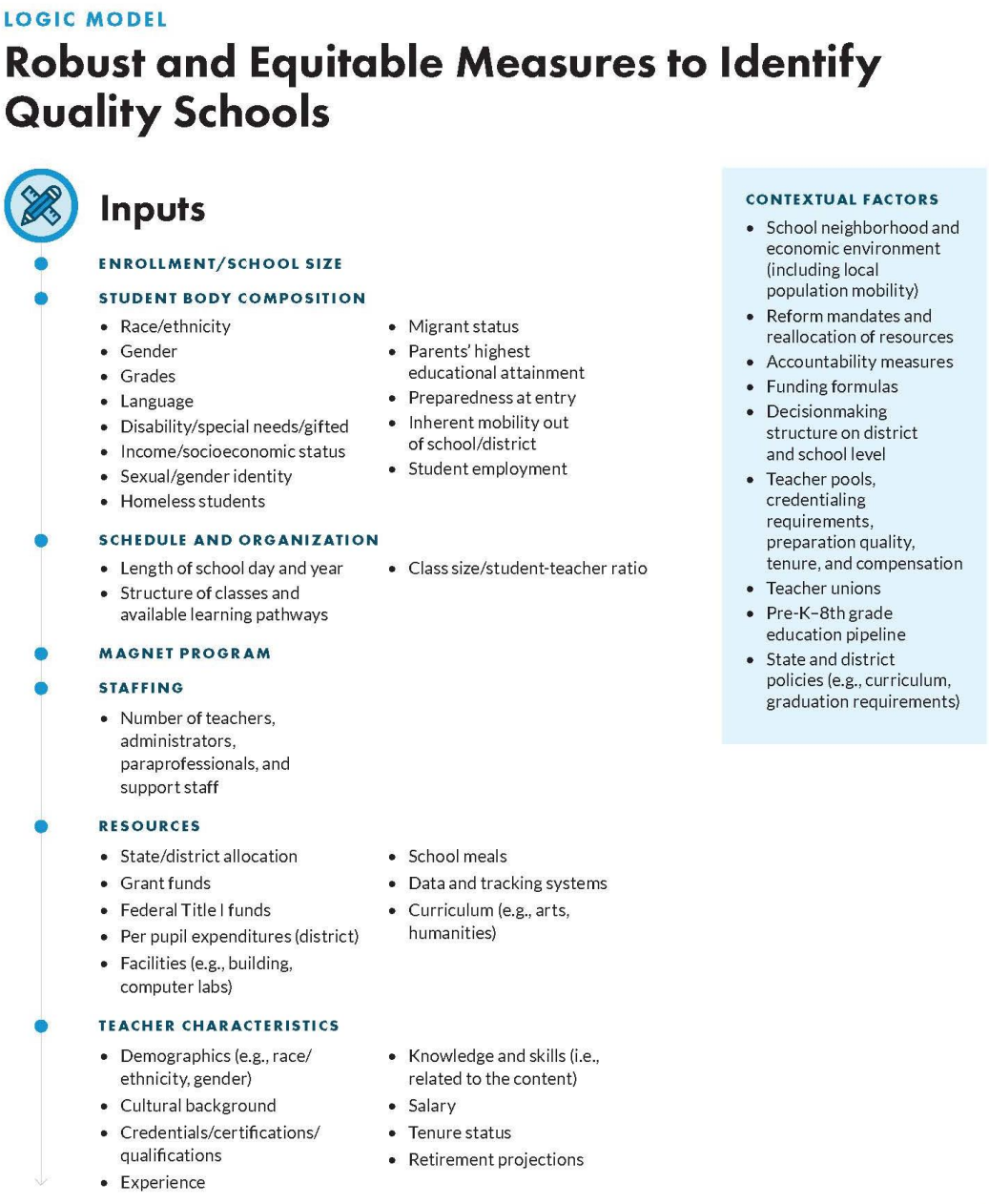
Future Opportunities

In future work, we urge researchers to continue focusing on long-term outcome measures rather than, or at least in addition to, school outputs. Focusing on these measures will become easier as statewide longitudinal data systems mature. More states will likely have systems that are accessible to researchers, the oldest cohorts in the systems will have more long-term data, and other efforts by federal agencies, such as the Census Bureau,[16] may have expanded. These improvements will allow researchers to study the wages of earners who leave the state. The next phase of REMIQS will expand to include five states followed by a two-year in-depth investigation of 10 high-flying schools identified through a five-state quantitative filtering analysis.

Appendix A. Logic Model

FIGURE A1

REMIQS Initial Logic Model from Concept Phase



Implementation/Process Activities

SCHOOL LEADERSHIP

- Decisionmaking processes (collaborative, hierarchical)
- Hiring (if given authority)
- Fund allocation and budget (if given authority)
- Mission
- Disciplinary policy
- Expectations for staff and students
- Ability to keep order
- Communication style
- Consistency
- Use of data/setting a data culture
- Ability to raise additional resources
- Flexibility and adaptivity

TEACHER ENGAGEMENT

- Absences
- Engagement with professional development
- Opportunities for teacher leadership
- Retention/turnover

TEACHING AND LEARNING PRACTICE

- Instructional practice (including customization/ personalization of instruction)
- Intervention models (including strategies for student engagement, formative assessment)
- Progress monitoring (including use of data)
- Time on task
- Expectations and rigor
- Development of student identity as learners

SCHOOL STRUCTURE AND RESOURCES

- General population support services (e.g., guidance counselor, college counselor, employment assistance, emergency funds, disability support)
- ELL or dual-language program(s) offered
- Special education and support for students with disabilities
- Advanced coursework (e.g., AP, IB, dual enrollment/early college)
- Ability tracking
- Career pathway programming for students (e.g., internships, credentials, vocational education, job fairs, job readiness programming)
- College connections (e.g., visits to school by college representatives, college centers)
- Behavior management system (e.g., PBiS systems and fidelity, restorative practices)
- Health and mental health services
- Socioemotional/leadership development interventions
- General population programs/interventions (e.g., reading programs, incentive systems)
- Online learning
- Curriculum/teaching materials (not infrastructural)
- Curriculum development (including for cultural relevance)
- Teacher and staff professional development (including on data usage, collaborative and systemic analysis of student work, formative assessment practices)
- Specials (physical education, electives, extracurricular activities, and enrichment programming)
- Student affiliation or decisionmaking bodies (e.g., GSA, student government)

SCHOOL CULTURE AND CLIMATE

- Teacher-leader relationship and dynamics
- Teacher peer engagement (e.g., peer learning communities, peer observation)
- Teacher investment in school and students
- Mentoring relationships between adults and students
- Consideration of students' social location and how status differences shape student experiences of school
- Student peer relationships (including issues like bullying)
- Student attachment to school/sense of belonging
- Community/family engagement (including formal associations)
- Safety and perceptions of safety

Outputs

HIGH SCHOOL ENGAGEMENT/SCHOOL CLIMATE

- Chronic absenteeism/ average daily attendance
- Suspension and expulsion rates
- Share choosing to reenroll in the same school (in school choice settings)
- Measures of student engagement/enthusiasm/ academic aspirations
- Rates of arrests and law enforcement referrals

HIGH SCHOOL CONTENT KNOWLEDGE

- Student academic proficiency measured by standardized assessments in math and literacy
- Student academic growth measured by standardized assessments in math and literacy
- Average course academic grades
- Share of students enrolling in advanced coursework
- Average science performance
- Average performance on advanced coursework exams
- Average performance on portfolio-alternative assessments
- English language learner redesignation/reclassification

HIGH SCHOOL ON-TRACK/PROMOTION/COMPLETION

- Four-, five-, and six-year high school graduation rates/dropout rates
- Share on track to graduate
- Share overage/undercredited
- Rate of on-time enrollment in prealgebra and algebra

COLLEGE/CAREER READINESS BY GRADUATION

- Rate of completion of a career pathway program while in high school
- Average college admission test scores
- Rate of completion of a college-track curriculum
- Rate of completion of college-level courses/ credits in high school
- Share earning a career readiness certificate by high school completion
- Share earning a military or workforce certification by high school completion
- Share possessing marketable trade skills by high school completion

PHYSICAL HEALTH AND WELLNESS BY GRADUATION

- Rate of teen parenthood
- Rate of drug/substance use/abuse

Outcomes

POSTSECONDARY APPLICATION/ATTENDANCE/ SUCCESS OF ALUMNI

- Postsecondary enrollment rate
- College performance/ completion
- Highest degree attained
- Quality of college attended
- College match

MEDIATORS AND MODERATORS

STUDENT MEDIATORS

Mechanisms for effects

- Student learning
- Nature of relationships (including adult and peer effects)
- Student confidence
- Student aspirations
- Norms around learning
- Development of self- and interpersonal respect

SCHOOL MODERATORS

Conditions or characteristics that may enhance or dampen effects

- Share of students who are English language learners
- Share of students of each major racial/ethnic group
- Share of students with an Individualized Educational Plan
- Share of students who qualify for free and reduced-price lunch
- Average student preparedness at entry
- Average student mobility out of school/district
- Share of female students
- Urban/rural area
- School size
- District size
- Student-teacher ratio
- Local economic indicators
- Local safety and law enforcement indicators
- State, county, metropolitan area, city, district, census tract

Outcomes continued

DEEPER LEARNING SKILLS OF ALUMNI

- Knowledge (academic content, career, citizenship content, practical life knowledge)
- Skills/ability (creativity; confidence; self-regulation, responsibility, goal-setting, reflexivity; social interaction/communication; critical thinking/problem solving; information and technology; resourcefulness)
- Behavior
- Mission motivation to learn and be challenged/academic self-concept
- Appreciation of and ability to engage with diversity/equity

PHYSICAL HEALTH AND WELLNESS OF ALUMNI

- Physical health/fitness

JOB QUALITY AND EARNINGS OF ALUMNI

- Labor force participation
- Employment status
- Military enlistment
- Earnings
- Family income/socioeconomic status
- Job-provided benefits
- Job satisfaction
- Adult poverty
- Government-provided benefits

CIVIC ENGAGEMENT OF ALUMNI

- Voter participation
- Incarceration
- Volunteerism
- Community organization participation and leadership

Appendix B. Variable Inclusion/Exclusion Explanations

TABLE B1

Explanations for Inclusion/Exclusion of Input Variables in Analysis Model

Category	Input	Included?	Explanation
School Character-istics	Enrollment/school size	Yes	
	Magnet program	No	Does not vary within schools across years
	Length of school day and year	No	Does not vary within schools across years
	Structure of classes and available learning pathways	No	No data
	Class size/student-teacher ratio	Yes	Student-teacher ratio included
	Number of teachers, administrators, paraprofessionals, and support staff	Partially	Number of teachers implicit in student-teacher ratio
	State/district allocation	Partially	Included in per-pupil expenditures
	Grant funds	Partially	Included in per-pupil expenditures
	Federal Title I funds	Partially	Included in per-pupil expenditures
	Per pupil expenditures (district)	Yes	District level only
	Facilities (e.g., building, computer labs)	No	No data (also rarely varies within schools across years)
	School meals	No	No data
	Data and tracking systems	No	No data
	Curriculum (e.g., arts, humanities)	No	No data (could possibly determine from transcript data, but not available/complete across states)
Student Character-istics	Race/ethnicity (major groups)	Yes	
	Gender	Yes	Female/male
	Grades	Yes	Used 9th grade cohort
	Language (e.g., ELL status)	Yes	ELL status
	Disability/special needs/gifted (e.g., IEP)	Yes	IEP status
	Income/socioeconomic status (e.g., FRPL)	Yes	Used school-wide free lunch and reduced-price lunch because not available at student level in all states (KY has restriction)
	Sexual/gender identity	No	No data
	Homelessness	No	Not tracked in all states
	Migrant status	No	Not tracked in all states
	Parents' highest educational attainment	No	No data
	Preparedness at entry	Yes	8th grade test scores
	Mobility out of school/district	No	Possibly could compute, but high burden of data processing
	Student employment	No	No data; possibly could determine on-the-books employment if state

Category	Input	Included?	Explanation
			earnings data available (but likely to be very low)
Teacher Character-istics	Demographics (race/ethnicity, gender)	No	No data
	Cultural background	No	No data
	Credentials/certifications/qualifications	No	Difficult to reconcile across states
	Experience	No	No data
	Knowledge & skills (i.e., related to the content)	No	No data
	Salary	No	Not available at individual level or for all years
	Tenure status	No	No data
	Retirement projections	No	No data

TABLE B2

Explanations for Inclusion/Exclusion of Output and Outcome Variables in Analysis Model

Category	Output/Outcome	Included?	Explanation
High School Engagement / School Climate	Absenteeism/attendance	Yes	Attendance
	Suspension & expulsion	No	Could include at school level but we had concerns about equity and consistent measurement/ reporting across states; data not available for all cohorts
	Re-enrollment in the same school (in choice settings)	No	Choice environments vary across states
	Engagement/enthusiasm/academic aspiration	No	No data
	Arrests & law enforcement referrals	No	Could include at school level but we had concerns about equity and consistent measurement/ reporting across states; data not available for all cohorts
High School Content Knowledge	Academic proficiency measured by standardized assessments in math & literacy	No	Included growth measure instead
	Academic growth measured by standardized assessments in math & literacy	Yes	10th grade test scores, normed with NAEP
	Science performance	No	Transcript data not available/ complete from all states
	Course academic grades	No	Transcript data not available/ complete from all states
	Enrollment in advanced coursework	No	Transcript data not available/ complete from all states
	Performance on advanced coursework exams	No	Data not complete over time in all states (MA data incomplete)
	Performance on portfolio-alternative assessments	No	No data
	English language learner redesignation/ reclassification	No	Not applicable to all students
High School On-Track / Promotion / Completion	Graduation/Drop-Out	Yes	Graduation by end of observed period
	On track to graduate	No	Transcript data not available/ complete from all states
	Overage/undercredited	No	Transcript data not available/ complete from all states
	On-time enrollment in prealgebra and algebra	No	Transcript data not available/ complete from all states
College / Career Readiness by Graduation	Completion of a career pathway program while in high school	No	Not available in all states
	College admission test scores	No	Attempted to include, but substantial differences in test-taking rates across states
	Completion of a college-track curriculum	No	Transcript data not available/ complete from all states
	Completion of college-level courses/credits in high school	Partially	Captured in college enrollment
	Earn a career readiness certificate by high school completion	No	Not available in all states

Category	Output/Outcome	Included?	Explanation
	Earn a military or workforce certification by high school completion	No	Not available in all states
	Possess marketable trade skills by high school completion	No	No data
Postsecondary Application / Attendance / Success of Alumni	Postsecondary enrollment	Yes	National Student Clearinghouse in MA and VA; in-state enrollment+high ACT scores in KY (see technical appendix)
	College performance/completion	No	Could measure, but would only reflect older student cohorts; could track persistence as interim measure
	Highest degree attained	No	Could measure, but would only reflect older student cohorts
	Quality of college attended	No	High burden of data processing; could break out 2-year from 4-year enrollment, but concerns about equity in ranking different success pathways
	College match	No	No data
Socio-Emotional / Citizenship / 21 Century Skills of Alumni	Knowledge	No	No data
	Behavior	No	No data
	Skills/ability	No	No data
	Motivation to learn & be challenged/ academic self-concept	No	No data
	Appreciation of & ability to engage with diversity	No	No data
Physical Health & Wellness of Alumni	Physical health/fitness	No	No data
	Teen parenthood	No	No data
	Drug/substance use/abuse	No	No data
Job Quality & Earnings of Alumni	Labor force participation	No	No data
	Employment status	No	Requested data, but restricted in KY and not received from other two states in time for analysis
	Military enlistment	No	No data
	Earnings	No	Requested data, but restricted in KY and not received from other two states in time for analysis
	Job satisfaction	No	No data
Civic Engagement of Alumni	Voter participation	No	Possibly could pull from state voter records, but would require state cooperation and would be burdensome
	Incarceration	No	Possibly could pull from state prison records, but would require state cooperation and would be burdensome
	Volunteerism	No	No data
	Community organization participation & leadership	No	No data

Notes

[1] The separate technical appendix is available at https://www.urban.org/research/publication/identifying-high-performing-schools-historically-underserved-students/technical-appendix .

[2] We sought but were not able to access deidentified, individual-level earnings records from each state.

[3] These are "minorities" as defined in 20 U.S.C. § 1067k (2008). The term "minority" is not ideal. We prefer "students from historically underserved groups" or "students from historically underserved populations."

[4] Included controls needed to have some variation within schools over time, so inherent characteristics of schools (e.g., geography or magnet status) were excluded.

[5] We considered ACT and SAT scores as a possible fourth measure but decided against it because we did not appear to have complete test-taking data for Massachusetts or Virginia.

[6] In Virginia, technical issues made it challenging to obtain earnings data for such a large volume of students; given the lack of earnings data in other states, we opted to forego this effort in the interests of pursuing outcomes that were comparable across states. In Massachusetts, a request for earnings data was still outstanding at the time of this report.

[7] Since states administer different tests, we use the National Assessment of Educational Progress (NAEP) test to adjust the scores to align with the NAEP means and distributions in each state. The technical appendix contains more details about our methodology.

[8] We add together NAEP-adjusted math and reading scores for each high school.

[9] The combined model can consider only factors that vary for schools over time, so even a school's urban, suburban, or rural location cannot be included as a control.

[10] The raw value-add measures have slightly higher correlations than the school rankings but not to a substantial degree.

[11] FRPL has been criticized as an insufficient proxy for student socioeconomic status. See, for example, Harwell and LeBeau (2010).

[12] The "some college, no degree" phenomenon is quite common, as discussed in Shapiro et al. (2014). This may be more likely as high schools promote college enrollment among students who would not likely have enrolled otherwise.

[13] See Western Interstate Commission for Higher Education, "Facilitating Development of a Multistate Longitudinal Data Exchange," press release, September 13, 2010, https://www.wiche.edu/news/release/100913.

[14] "Post-Secondary Employment Outcomes (PESO) (Beta)," US Census Bureau, accessed October 29, 2019, https://lehd.ces.census.gov/data/pseo_beta.html.

[15] Adam Bee, "Understanding the Relationship between Individual Earnings and Household Income," *Random Samplings* (blog), US Census Bureau, November 30, 2017, https://www.census.gov/newsroom/blogs/random-samplings/2017/11/earnings-income.html.

[16] See information about the pilot Post-Secondary Employment Outcomes (PSEO) program, which has long-term outcomes for students in four states as of the date of publication, at "Post-Secondary Employment," US Census Bureau.

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.

Chingos, Matthew, Daniel Kuehn, Tomas Monarrez, Patrick J. Wolf, John F. Witte, and Brian Kisida. 2019. *The Effects of Means-Tested Private School Choice Programs on College Enrollment and Graduation*. Washington, DC: Urban Institute.

Greenberg, Erica. 2018. "New Measures of Student Poverty: Replacing Free and Reduced-Price Lunch Status Based on Household Forms with Direct Certification." Washington, DC: Urban Institute.

Harwell, Michael, and Brandon LeBeau. 2010. "Student Eligibility for a Free Lunch as an SES Measure in Education Research." *Educational Researcher* 39 (2): 120–31.

Shapiro, Doug, Afet Dundar, Fay Huie, Phoebe Khasiala Wakhungu, Ayesha Bhimdiwala, and Sean Eric Wilson. 2018. *Completing College: A National View of Student Completion Rates—Fall 2012 Cohort*. Herndon, VA: National Student Clearinghouse Research Center.

Shapiro, Doug, Afet Dundar, Xin Yuan, Autumn T. Harrell, Justin C. Wild, and Mary B. Ziskin. 2014. *Some College, No Degree: A National View of Students with Some College Enrollment, but No Completion*. Herndon, VA: National Student Clearinghouse Research Center.

About the Authors

Erica Blom is a research associate in the Center on Education Data and Policy at the Urban Institute, where she studies higher education policy. Before joining Urban, she was a principal consultant at Edgeworth Economics, studying such issues as discriminatory practices in the labor market and anticompetitive practices. She also contributed to various pro bono projects, including the evaluation of charter school lotteries and a study of the Public Service Loan Forgiveness program. Blom received a bachelor's degree in mathematics and political science from Queen's University and a master's degree in economics from Western University. She also earned a doctoral degree in economics from Yale University, where her research focused on students' choices in college major. She has published in the Annual Review of Economics, and her dissertation research has contributed to joint work with Benjamin Keys and Brian Cadena on the effects of the business cycle on college major choice.

Theresa Anderson is a senior research associate at the Urban Institute. She is a member of the Building America's Workforce cross-center initiative and is affiliated with the Income and Benefits Policy Center and the Education Data and Policy Center. She conducts in-depth, mixed-methods research on important policy issues, primarily evaluations of workforce, education, and social safety net programs and policies.

Constance A. Lindsay is a research associate at the Urban Institute, where she studies K–12 education policies. She is also an assistant professor at the University of North Carolina at Chapel Hill. Lindsay earned a doctorate in human development and social policy from Northwestern University, where she was an Institute of Education Sciences' predoctoral fellow. Since leaving Northwestern, Lindsay has worked in education policy in various contexts, applying her research training in traditional studies and in creating and evaluating new systems and policies regarding teachers. Lindsay's areas of expertise include teacher quality and diversity, analyzing and closing racial achievement gaps, and adolescent development. Her work has been published in such journals as Educational *Evaluation and Policy Analysis and Social Science Research*. Lindsay received a bachelor's degree in economics from Duke University and master's degree in public policy from Georgetown University. Before doctoral study at Northwestern, she was a Presidential Management Fellow at the US Department of Education.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

URBAN
INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org