NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students

Wes Austin, David Figlio,
Dan Goldhaber, Eric A. Hanushek,
Tara Kilbride, Cory Koedel,
Jaeseok Sean Lee, Jin Lou, Umut Özek,
Eric Parsons, Steven G. Rivkin,
Tim R. Sass, Katharine O. Strunk

WORKING PAPER No. 227-0220-2 • August 2021

Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students

Wes Austin
Georgia State University

David Figlio
Northwestern University
CALDER

Dan Goldhaber
American Institutes for Research/CALDER
University of Washington/CEDR

Eric A. Hanushek
Stanford University
University of Texas at Dallas
CALDER

Tara Kilbride
Michigan State University/EPIC

Cory Koedel
University of Missouri
CALDER

Jaeseok Sean Lee
University of Missouri

Jin Luo
University of Texas at Dallas

Umut Özek
American Institutes for Research/CALDER

Eric Parsons
University of Missouri

Steven G. Rivkin
University of Illinois at Chicago
CALDER

Tim R. Sass
Georgia State University
CALDER

Katharine O. Strunk
Michigan State University/EPIC
CALDER

Contents

Acknowledgments

We gratefully acknowledge financial support from CALDER, which is funded by a consortium of foundations (for more information about CALDER funders, see www.caldercenter.org/about-calder). All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders, data providers and partners, or the institutions to which the author(s) are affiliated, and all errors are our own. This work would not have been possible without the support of our state partners. We make the following acknowledgments and disclaimers regarding the provision of data:

Georgia: The contents of this report were developed using data provided by Georgia's Academic and Workforce Analysis and Research Data System (GA•AWARDS). However, those contents do not necessarily represent the policy of GA•AWARDS or any of its participating organizations, and you should not assume endorsement by GA•AWARDS or any of its participating organizations.

Massachusetts: The authors wish to thank partners at the Massachusetts Department of Elementary and Secondary Education for the provision of data to support this work.

Michigan: This research result used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information, and opinions solely represent the analysis, information and opinions of the author(s) and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. We acknowledge valuable research assistance from Amy Auletto.

Missouri: The authors wish to thank partners at the Missouri Department of Elementary and Secondary Education for the provision of data to support this work.

Oregon: The authors wish to thank partners at the Oregon Department of Education for the provision of data to support this work.

Texas: The authors wish to thank the Texas Schools Project and UT Dallas Education Research Center. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission or the State of Texas.

Washington: The research presented here would not have been possible without the administrative data provided by the Washington Office of Superintendent of Public Instruction through data-sharing agreement 2015DE-030.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students

Wes Austin, David Figlio, Dan Goldhaber, Eric A. Hanushek, Tara Kilbride, Cory Koedel, Jaeseok Sean Lee, Jin Lou, Umut Özek, Eric Parsons, Steven G. Rivkin, Tim R. Sass, Katharine O. Strunk

CALDER Working Paper No. 227-0220-2
August 2021

Abstract

There is empirical evidence of substantial heterogeneity in economic mobility across geographic areas and the efficacy of schools has been suggested as an explanatory factor. Using administrative microdata from seven states covering nearly 3 million students, we explore the potential role of schools in promoting economic mobility by estimating cross-district variation in "academic mobility"—a term we use to describe the extent to which students' ranks in the distribution of academic performance change during their schooling careers. We show that there exists considerable heterogeneity in academic mobility across school districts. However, after aggregating our district-level measures of academic mobility to the commuting-zone level and merging them with geographically matched external estimates of economic mobility, we find little scope for geographic differences in academic mobility to meaningfully account for differences in economic mobility.

1. Introduction

Performance differences between students during K-12 education predict later-life outcomes, including earnings (Cawley, Heckman, and Vytlacil, 2001; Cunha et al., 2006; Murnane, Willett and Levy, 1995; Murnane et al., 2000). Therefore, it is not surprising that differences in schools have been postulated as a driver of the considerable geographic heterogeneity in economic mobility documented in recent research (Chetty, Hendren, Kline, and Saez, 2014; Chetty, Hendren, Jones, and Porter, 2018). But students from high-income households also enter public schools with much stronger academic performance than their low-income peers and these performance differences have a high degree of persistence during K-12 schooling (Goldhaber, Wolff, and Daly, 2020; Sorenson, 2019; Todd and Wolpin, 2007). The role of schools in promoting (or dampening) economic mobility is not obvious.

With an eye toward assessing the potential role of schools in influencing economic mobility, we document and explore variation across school districts in how much initially low-achieving students—who are disproportionately from lower-income households (Jang and Reardon, 2019; Reardon, 2011)—gain in the performance distribution during K-12 schooling. To do this we introduce the concept of "academic mobility." Academic mobility is a measure that captures the extent to which a student's academic performance early in the schooling career maps to his or her own longer-term educational outcomes through high school. Differences in academic mobility across school districts can be used to gain insight into the scope for districts to narrow *intragenerational* gaps in educational outcomes, and ultimately, *intergenerational* gaps in adult outcomes such as earnings.

We estimate academic mobility for school districts using administrative microdata from seven states covering almost 3 million students and tools developed by Chetty, Hendren, Kline, and Saez (CHKS, 2014) and Chetty, Hendren, Jones, and Porter (CHJP, 2018).[1] Students' initial

[1] In turn, these studies build on a large prior literature on economic mobility—for reviews see Black and Devereux (2011) and Solon (1999).

performance levels are assessed using test scores in the third grade (the earliest grade we have universal data on test performance in public schools) and we use four long-term outcomes to assess their academic mobility as they progress through the K-12 education system: eighth-grade test performance, high-school test performance, on-time high school graduation, and high school graduation within one year of on-time. Our district-level academic mobility metrics indicate the degree to which students' late-grade placements in the outcome distributions are determined based on early-grade performance and allow us to assess how distributional "stickiness" varies across districts.

Our access to individual-level microdata allows us to decompose total academic mobility in each district into two components, which we call "absolute mobility" and "relative mobility." A useful context for illustrating these concepts is to consider the improvement of initially low-achieving students in the larger performance distribution (e.g., statewide) as they progress through school. A district with high *absolute mobility* is one in which initially low-achieving students gain in the performance distribution over time as part of a rising tide that lifts all boats—e.g., a district that has large positive effects that are uniform throughout the distribution of initial achievement would exhibit high absolute mobility. Whereas *relative mobility* captures changes in the performance of initially low-achieving students compared to their higher-achieving peers locally within the same district. A district with high relative mobility narrows internal achievement gaps over time. Both types of mobility affect the total gains of initially low-achieving students in the larger performance distribution. Disentangling them is an important first step toward understanding what drives observed variation across districts in academic mobility.[2]

[2] Both a student's absolute position in the performance distribution and a student's relative position within a class, school, or district are important outcomes of interest. A student's absolute position is important given causal evidence on the link between test scores and later life outcomes (Goldhaber and Ozek, 2019). There is also increasing evidence that a student's relative rank has independent effects on student behaviors and outcomes, as social comparisons help to shape ability beliefs. See, for instance, Cicala et al. (forthcoming), Denning et al. (2020), Elsner and Insphording (2017a, 2017b), Elsner et al. (2019), and Murphy and Weinhardt (2020).

In what follows, we first summarize academic mobility in our sample states and how patterns of mobility vary by student race/ethnicity, socioeconomic status, and the urbanicity of the school attended in the third grade. We then document heterogeneity across districts in academic mobility. We find clear evidence of cross-district heterogeneity that is statistically significant and economically meaningful. Variation in total academic mobility across districts is driven primarily by differences in absolute mobility.

Our district-level mobility metrics should not be interpreted as causal. They are place-based descriptive measures and may be influenced by school and/or non-school factors associated with the districts in which students are enrolled. That said, the cross-district variance in academic mobility that we document—and decompose into its components—is useful for thinking about the ways in which districts are most likely to help low-achieving students improve. Specifically, the fact that differences in relative mobility across school districts are much smaller than differences in absolute mobility is inconsistent with the idea of widespread differences across districts in the ability to narrow achievement gaps.[3] This could change with changes to education policies, but in the context of current practice, we show that there is much more variability across districts in absolute mobility than in relative mobility.

To explore the link between variation in academic mobility and economic mobility, we link our district-level academic mobility metrics to the intergenerational economic mobility estimates of CHKS. CHKS report economic mobility at the commuting zone level; correspondingly, we aggregate our metrics to the same level in our sample states. Despite our finding of substantial variation in academic mobility across individual districts, we conclude that academic mobility across commuting zones is not an important driver of the variation in intergenerational economic mobility documented by CHKS. The reason is that the vast majority of the cross-district variance in academic mobility occurs within commuting zones (about 70-85

[3] It may be that some districts are particularly effective in this way, as has been argued in several small-scale studies (Leithwood, 2010; Rorrer, Skrla, and Scheurich, 2008), but our results cast doubt on the notion that there are large differences across school districts in terms of narrowing internal achievement gaps.

percent depending on the outcome we use to measure academic mobility). The small cross-commuting-zone variance share that remains cannot account for a meaningful fraction of observed differences across commuting zones in intergenerational economic mobility. Our findings in this regard are consistent with findings from Rothstein (2019), who shows that differences in the relationship between parental income and children's human capital across commuting zones—a metric akin to academic mobility, although Rothstein's data and methods differ substantially from our own—can account for only a small fraction of the cross-commuting-zone variance in economic mobility documented by CHKS.

2. Data and Measurement of Academic Mobility

2.1 Data

We use state administrative microdata from public schools in seven states—Georgia, Massachusetts, Michigan, Missouri, Oregon, Texas, and Washington. We assemble cohorts of all students who have standardized test scores in math and English language arts (ELA) in the third grade—the initial statewide testing grade in most states—and follow them through high school. Academic mobility is assessed as cohorts progress through school.

Table 1 reports descriptive information for the third-grade cohorts in each state, as well as for K-12 students in the entire U.S. for comparison. We track academic mobility for two to four year-cohorts of students in our sample states between 2005-06 and 2008-09 (hereafter, including in Table 1, we identify school years by the spring year; e.g., 2006 for "2005-06"). The earliest year cohort is 2006 because this is the first year of consistent testing in grades 3-8 in most states, and the latest year cohort is 2009 because this is the oldest cohort for whom we can track graduation outcomes (within one year of on-time) using our data panels.

Our analysis includes about 2.9 million students in total and the sample states exhibit substantial heterogeneity in terms of their student populations. For example, the percent of Black and Hispanic students across states range from 3.0-38.1 and 4.0-47.7, respectively. There is also considerable variation across states in the shares of students receiving free or reduced-price lunch (FRL), identified for an Individualized Education Program (IEP), and who are geographically mobile.[4] Finally, the structure of the education system differs across states in terms of the shares of schools located in urban, suburban, and rural areas; and the numbers districts and schools, both in absolute and per-capita terms. While our sample is not designed to be representative of the United States as a whole, the seven states we examine are diverse along many dimensions and provide substantively different evaluation contexts.

Under the No Child Left Behind Act (NCLB) and Every Student Succeeds Act (ESSA), all students are required to be tested in math and ELA/reading in grades 3-8, and at least once in grades 10-12. While each state administers its own tests, our analysis of academic mobility between grades 3 and 8 is fairly uniform across states due to the federal testing requirements.[5] At the high school level, however, the flexibility of testing requirements means that the grades in which test outcomes are observed, and in which subjects, varies across states. To assess academic mobility based on high-school achievement, in each state we identify the exam with the highest coverage rate administered in a common grade. These tests are shown in Table 2.[6] With the exception of Michigan, which has a universal ACT/SAT policy, the common-grade requirement is such that the subject of the selected test is ELA-based. This is because the English curriculum in high schools is more rigidly structured than in other subjects. Table 2 shows that the focal high school tests are administered mostly in the tenth and eleventh grades (the one exception is Georgia, where the test is administered in ninth grade), have very high coverage rates, and are overwhelmingly taken in a common grade. In Oregon there is no test that is given

---

[4] Geographic mobility is defined by students who are enrolled in more than one school during the year in which they took the 3rd grade test. States differ in terms of the frequency of collecting school enrollment information, which may account for some of the heterogeneity across states in this variable. Note that the FRL data used for these cohorts pre-date the option of schools to use the Community Eligibility Provision, which leads to measurement error in FRL data (Koedel and Parsons, 2021).

[5] Some students take an algebra-I end-of-course (EOC) test instead of the statewide grade-8 math test in grade-8. For these students, we use their grade-7 test performance to predict what the grade-8 test would have been had they taken the statewide test. More details about the prediction model are available upon request.

[6] The requirement of a common grade limits concerns about the confounding effect of test-timing on our cross-district measures of academic mobility, which has come up most often with respect to studies of Algebra-I end-of-course exam performance (Clotfelter, Ladd, and Vigdor, 2015; Domina et al., 2015; Parsons et al., 2015).

overwhelmingly in the same grade in high school, so we omit Oregon from the high-school-achievement portion of our analysis.

In addition to assessing test-based academic mobility, we also assess mobility in terms of the likelihood of high school graduation. We consider both on-time and graduation within one year of on-time.

2.2 Measuring Academic Mobility

2.2.1 Rationale and Basic Framework

Our methodological approach follows closely on the framework developed by CHKS and CHJP to study intergenerational economic mobility. Focusing first on our test-based mobility metrics, they are constructed based on percentile rankings in the test distribution at different points in the schooling career. Like CHKS and CHJP, we have sufficiently rich data to describe the joint distribution of early- and late-career student performance nonparametrically in the form of 100x100 percentile matrices for each outcome and state. However, a key insight from CHKS, which permits a more parsimonious presentation, is that the rank-rank relationship between intergenerational economic outcomes is functionally linear. This also turns out to be true in our application, allowing us to summarize academic mobility with just the slope and intercept parameters from a linear regression. Documentation of the linearity of the rank-rank relationships between early- and late-career student outcomes in our data is provided in Appendix A.[7]

Given the linear relationships, the mapping between students' late-career and early-career (grade-3) outcomes can be summarized by equation (1):

$$O_i = \alpha + \beta R_i + \varepsilon_i \qquad (1)$$

[7] As shown in the appendix, the achievement-based rank-rank relationships are clearly linear. For the relationships involving high school graduation, which is a binary outcome, we use binned scatter plots where the first point is the average high school graduation rate for students who enter the panel in the 1st percentile, and so on. CHJP use this method to explore several dichotomous outcome variables in their study. As in CHJP, the rank-rank relationship is linear throughout most of the initial placement distribution (roughly the upper 80 percent) for our binary graduation outcomes. The nonlinearity at lower percentile values is attributable to strong floor effects in graduation combined with the binary outcome.

In the equation, O_i is a late-career rank outcome for student *i* and R_i is student *i*'s initial rank (assessed in the third grade). We use the average rank on third grade math and ELA tests to set R_i for each student. The two late-career rank outcomes are defined as: (1) the average rank on math and English Language Arts (ELA) tests in the eighth grade, and (2) the rank on the high school test using the test indicated in Table 2 for each state.

Figure 1 illustrates two stylized (extreme) mobility scenarios within the percentiles-to-percentiles framework. The first graph in the figure shows a case where $\alpha = 0$ and $\beta = 1$. This is a scenario with no academic mobility, as the average outcome rank is the same as the entry rank at every percentile in the distribution. At the other extreme, the second graph where $\alpha = 50$ and $\beta = 0$ indicates perfect academic mobility; here, the average outcome rank is at the median regardless of the student's entry percentile.

Figure 1 illustrates the interdependence of α and β when the rank-rank relationship is estimated on the entire population, which in our context is the population of public school students in a state. Because the estimated regression line for an entire state must pass through the mean of the data and the model regresses percentiles on percentiles, then by construction it must pass through (50, 50). As a result, the mobility relationship is fully captured by the slope coefficient, β, which also defines the y-intercept, α, given by $\alpha = 50 - 50\beta$.

When we disaggregate the data below the state level—i.e., for subpopulations of students within a state, or for school districts—the parameters α and β become separately identifiable and provide unique information about absolute and relative mobility, respectively. This is because the rank-rank lines need not pass through the point (50, 50) for each subpopulation. To illustrate this point, consider the following modified versions of equation (1) that permit subgroup analyses:

$$O_{is} = \alpha_s + \beta_s R_{is} + \varepsilon_{is} \tag{2}$$

$$O_{id} = \alpha_d + \beta_d R_{id} + \varepsilon_{id} \tag{3}$$

In equation (2), the subscript *s* indicates group membership for student *i*. We define groups *s* by race/ethnicity, FRL eligibility, and the urbanicity of the school attended in the third grade (urban, suburban, or rural). In equation (3), the subscript *d* identifies students who attend district *d* in the third grade.

As long as the dependent and independent variables in equations (2) and (3) continue to be defined by the full statewide distributions, the intercepts and slopes for the groups indexed by *s* and *d* are separately identified and provide unique information about the nature of academic mobility. The intercept terms, α_s and α_d, capture absolute mobility in the statewide distribution at the bottom of the panel entry rankings, while β_s and β_d capture relative mobility.

Total academic mobility at initial percentile *p*, inclusive of absolute and relative mobility, can be expressed for district *d* as follows:

$$\bar{O}_{pd} = \alpha_d + \beta_d p \tag{4}$$

Similarly, $\bar{O}_{ps}$ gives the student-subgroup-level analog. Following CHKS, we focus on the mobility of students at the 25th percentile of the initial performance distribution to produce measures of total academic mobility for initially low-achieving students throughout our analysis, denoted by $\bar{O}_{25}$. From equation (4), $\bar{O}_{25}$ for students in district *d* is estimated by $\hat{\alpha}_d + \hat{\beta}_d * 25$.

Figure 2 plots linear mobility functions for two hypothetical districts corresponding to equation (3). In each graph, the slopes of the solid and dashed lines are held constant (i.e., neither β_{Solid} nor β_{dashed} change across panels), with the solid line exhibiting more relative mobility because it is less steep ($\beta_{solid} < \beta_{dashed}$). Hence, two students from the solid district who enter the panel with a given performance gap will have a smaller later outcome gap, on average, than two students from the dashed district who enter the panel with the same performance gap. This can be seen visually in Figure 2 by the fact that for a fixed entry-percentile gap (shown on the X-axis), the outcome-percentile gap (shown on the Y-axis) is smaller for students represented by the solid line. In each graph the intercept is larger for the solid district ($\alpha_{solid} > \alpha_{dashed}$)—i.e., the lowest-performing students from the solid district perform better on the outcome measure than the lowest-performing students from the dashed district.

Figure 2 makes clear that changes to α have uniform implications throughout the initial performance distribution, which is why we refer to it as capturing "absolute mobility." In contrast, changes to β have different implications for students depnding on their position in the initial performance distribution, thus our interpretation of β as a measure of "relative mobility." CHKS point out that it is straightforward to interpret higher absolute mobility as a positive attribute, but the same is not true of relative mobility. This point is exemplified by the comparison in the third panel of Figure 2. In this comparison, initially low-achieving students perform similarly in both districts and the higher relative mobility of the solid line reflects the underperformance of initial high achievers, which is not desirable. Many researchers and education systems use the district achievement gap as a measure of performance, but comparisons between relative and absolute mobility highlight potential limitations of this measure. For example, a rightward shift of the achievement distribution for one district could potentially leave the achievement gap unchanged or slightly larger, while a leftward shift of the achievement distribution that is more pronounced at higher achievement percentiles could reduce the achievement gap, but still leave initially low-achievers worse off in comparison to peers in other districts.

Finally, we turn to the application of this framework to our analysis of graduation outcomes. The interpretation of the intercept and slope parameters described thus far applies to their estimation on outcomes that are percentile-ranked, but graduation outcomes are binary indicators equal to one if student *i* graduated and zero otherwise. Noting this difference, the academic mobility parameters are conceptually similar in their interpretation in the graduation models (CHKS, 2014). For example, $\bar{O}_{pd}$ for on-time graduation indicates the likelihood of graduation for a student in the 25th percentile of the third grade performance distribution from district *d*. This likelihood can be compared to the likelihood of graduation for a student in the 25th percentile in district *c*, *c≠d*, to compare mobility across districts as measured by graduation.

2.2.2 Measurement Error and Geographic Student Mobility

There are two main complications associated with the measurement of academic mobility as described in the preceding section. First, the initial third grade percentile ranks are measured with error because they are based on noisy test scores (Boyd et al., 2013; Lockwood and McCaffrey, 2014). The measurement error will attenuate our estimates of β and correspondingly inflate our estimates of α.

One way we reduce test measurement error in the initial ranks is by using the average of the ranks in math and ELA in third grade, rather than using a single test. This reduces the influence of measurement error, but does not fully address the problem. Therefore, we also leverage data on measurement error in the testing instruments to further adjust our estimates. The state tests we use to set the initial ranks include publisher-provided reliability ratios, which give estimates of the measurement-error variance in student scores. We use the test-specific reliability ratios to construct composite reliability ratios for students' initial, averaged percentiles on the two tests following Wang and Stanley (1970).[8] We then estimate standard errors-in-variables versions of the above-described regressions to correct β for attenuation bias.[9,10]

---

[8] Wang and Stanley (1970) give general formulas for reliability ratios based on multiple assessments. In our case, define r_m and r_e as the reliability ratios for the third grade math and ELA standardized tests individually, and $\theta_{m,e}$ as the correlation of performance on the two tests. The reliability of performance as measured by the average performance across the two tests is given by: $r_c = \frac{0.25r_m + 0.25r_e + 0.50\theta_{m,e}}{0.50 + 0.50\theta_{m,e}}$. Note that the reliability ratios for each state test in each subject, taken by different cohorts in the sample, vary slightly from year-to-year—in our calculations we use one reliability ratio for each subject for all students in the formula, which we calculate using the average subject-specific ratio across all cohorts.

[9] The individual test reliability ratios are quite high on the assessments in our sample states, on the order of about 0.90, and the combined reliability ratios are higher. The applicability of test-reliability ratios when the data undergo a monotonic transformation (in our case, from scale scores to percentile ranks) depends on several factors and is the subject of some debate in the literature (May and Nicewander, 1994; de Gruijter, 1997). If anything, research suggests that our use of the reported reliability ratios may overstate reliability of the percentile ranks. Note, however, that May and Nicewander (1994), who raise concerns about percentile-transformed data, only find that reliability ratios translate poorly in cases where tests are especially easy or difficult.

[10] Another way to address the measurement error problem is to use more tests to set the initial ranks. This approach builds on analogous approaches used by CHKS and Solon (1992) to reduce the influence of measurement error in

A second complication is that we do not observe later-grade outcomes for all students in the intial entry cohorts because we only only observe outcomes for students who remain in public schools in their home states.[11] Table 3 shows that state exiters are negatively selected by showing that the average entry percentiles (and imputed outcome percentiles, which we discuss below) of students whose later-grade outcomes are unobserved are almost always below those of students with observed outcomes in our sample states.[12] These results are consistent with outside evidence on negative selection into student mobility—e.g., see Grigg (2012) and Mehana and Reynolds (2004).

There are two types of biasing concerns raised by attriton from our samples as the cohorts progress through school. The first is reference bias and applies to our analyses of the eighth grade and high school test percentiles, which are normed against the population of test takers. Because state leavers are negatively selected their departure from the data, if left unaddressed, will result in reference bias in the outcome percentiles. It will lead us to understate academic mobility among state stayers even if there is no unobserved selection associated with state exits (note that the reference bias issue is not relevant to our analysis of graduation outcomes because these outcomes are not normed in the distribution). The second concern is related to unobserved heterogeneity. Specifically, our findings will be biased by sample composition changes if state exiters are different from state stayers in unobserved ways conditional on their initial ranks. This concern applies to both our test-based and graduation-based mobility metrics and has the potential to be especially problematic for subgroup analyses. For example, consider the case

annual earnings in their investigations of economic mobility. In our case, we can use tests from later grades, but this comes with additional complications—most notably, as we use tests from later grades to set the initial ranks, the initial rank period gets closer in time to the outcome rank period, limiting our ability to study academic mobility. Nonetheless, in results omitted for brevity we explore this approach by bringing in 4th grade math and ELA assessments to set the initial ranks. The results are similar to what we find using the errors-in-variables correction, although we acknowledge there is some ambiguity owing to the aforementioned compression of time between ranking periods.

11 For the test outcomes we also lose students who do not take the tests, although the tests we use are meant to be given to all students to minimize sample attrition for this reason.

12 There is one exception in Texas. Note that with an underlying continuous distribution of scores, the mean of each rank distribution should be exactly 50. The mean in several states deviates (very) slightly from 50 because of lumpiness in the underlying test-score distributions, which produces lumpiness of percentiles that can fall above or below the median.

where state exiters are negatively selected conditional on their initial ranks and district *A* has a higher proportion of exiters than district *B*. In this scenario, the differential attrition between districts will cause a compositional difference in their comparison and lead to an overstatement of outcome variance between them.

We address both of these concerns by including students with missing outcomes in our analysis via imputation. Our imputation procedure uses all available test information prior to the missing outcome, up to the seventh grade, to impute test percentiles in eighth grade and high school, and both high school graduation outcomes.[13] The imputed values allow us to preserve the full entry-cohort distributions in each state, mitigating the concern about reference bias.[14] Regarding bias due to unobserved selection, although we cannot directly estimate the effect of unobserved selection, we address it by building five hypothetical selection scenarios into the imputation framework. The baseline selection scenario, which is the scenario we maintain throughout our primary analysis, is that students with missing outcomes are negatively selected on unobservables to the same degree as district movers within the same state. In other words, we assume that state and district changes are equally indicative of unobserved student circumstances. We produce imputed values for state exiters that embody this condition by relying on observed outcomes for district movers within each state to parameterize a "mobility selection parameter."

Using this scenario as an anchor, we then consider the sensitivity of our findings to four scenarios where the degree of selection among students with missing outcomes is re-parameterized in the imputation procedure. The re-paramaterizations of selection relative to baseline are as follows: (1) 25 percent more negative, (2) 10 percent more negative, (3) 10 percent less negative, and (4) 25 percent less negative. With the selection-adjusted imputed

[13] We additionally make an *ad hoc* correction to the variance of the imputed values to avoid complications due to shrinkage. Appendix C describes our imputation procedure in detail.

[14] Noting that state exiters and state entrants are both negatively selected, an alternative but less comprehensive strategy to combat reference bias is to replace state exiters with state entrants in the outcome distributions. We have pursued this strategy as well and our results are qualitatively similar throughout (results omitted for brevity); we prefer our imputation-based approach because it is more comprehensive and tractable.

values in hand, we can re-estimate our academic mobility models to determine the sensitivity of our findings to different assumptions about the direction and magnitude of unobserved selection into missing outcomes, above-and-beyond selection into district mobility within a state. Full details regarding our imputation procedure are provided in Appendix C.

This sensitivity analysis shows that none of our findings are substantively sensitive to the different unobserved selection conditions we test. This is the combined result several aspects of outcome missingness in our data: (1) even in the most extreme unobserved selection scenario, and noting that we already capture *observed* selection via early-grade performance, the degree of parameterized negative selection into exit is modest (based on within-state district movers), (2) although the likelihood of outcome missingness is not evenly distributed across student subgroups or districts, the divergence across subgroups and districts is also not extreme, and (3) most students in our sample states are not missing outcomes (Table 3), which limits the scope of impact on our findings.

Finally, we turn to the issue of geogrpahic student mobility within our sample states. We set students' district placements based on the third grade, which means that our estimates of cross-district variability take on an interpretation akin to "intent-to-treat" parameters. They broadly reflect the set of schooling experiences associated with students' third-grade districts. Although most students remain in the same district during grades 3-12, many change districts as well. Some of these changes are structural (e.g., a district that ends after eighth-grade) or opportunistic (e.g., our data cover a period of growth in the charter sector in many states), although moves surely occur for many other reasons as well.[15] Disentangling the reasons for student mobility across districts, and the implications, is a substantial undertaking and a natural extension of this work, but for the time-being we focus on understanding differences in student academic mobility across districts as defined by the district attended in the third grade.

[15] An important structural factor is the size of districts within a state, but even in a small-district state like Missouri, about two-thirds of students remain in the same district for grades 3-12.

3. Findings

We conduct the analysis described above for each state separately. The results are broadly consistent across states and accordingly, we consolidate the findings in the main text by reporting simple-average values across states. State-by-state breakouts are reported in Appendix B.

3.1 Broad Patterns of Academic Mobility at the State Level

Figure 3 reports averages of the state-level estimates of β and $\bar{O}_{25}$ from equation (1) (recall that α is redundant in the statewide models). Consistent with evidence that early measures of achievement are highly predictive of later outcomes, a student's position in the test distribution in the third grade is highly predictive of eighth grade and high school test rankings—the average estimates of β for the eighth grade and high school tests in Figure 3 are 0.84 and 0.82, respectively. Put plainly, where a student starts in the distribution when tested in the third grade is highly predictive of where she ends up in the distribution in eighth grade and high school.

The estimates of β for the graduation outcomes are much lower—0.35 for on-time graduation and 0.27 for lagged graduation, on average—reflecting a much weaker gradient between initial percentile ranks and the likelihood of high school graduation. The weaker gradient is visually apparent in the scatterlplots in Appendix A and is driven by the fact that graduation rates are high throughout most of the entry-rank distribution. Put another way, because high school graduation is a fairly indiscriminate outcome, students' early-career performance ranks are weaker predictors of success.

The $\bar{O}_{25}$ values for the test outcomes are similar to each other, at 29.7 and 30.4 for the eighth-grade and high-school tests, respectively. These numbers indicate that the average 25th percentile entrant scored at the 29.7th percentile of the combined eighth-grade math and ELA tests, and at the 30.4th percentile of the high school test. The fact that the high-school $\bar{O}_{25}$ values are higher than the 8th-grade values is consistent with academic mobility being a cumulative process during K-12 schooling, although the difference in the estimates is modest.[16] The graduation-based $\bar{O}_{25}$ values, which capture on-time and delayed graduation likelihoods for the average 25th percentile student, are 75.8 and 80.6, respectively, on average across the sample states.

Appendix B shows that the graduation-based mobility metrics exhibit more cross-state variability than the test-based metrics. A key reason is the distributions of test score ranks are forced into alignment across states by the percentiles conversion, but the distributions of graduation outcomes are quite different across states because there are large differences in statewide graduation rates. Given that most students graduate, the statewide graduation rate differences are particularly impactful for students in the lower end of the performance distribution. States with higher graduation rates overall have higher graduation-based $\bar{O}_{25}$ values.

This highlights is an important source of ambiguity in interpreting the mobility findings across states with respect to graduation. One interpretation of a high $\bar{O}_{25}$ value is that it reflects a state's success in pushing initially low-achieving students through high school. But an alternative interpretation is that a high graduation rate for initially low-performing students reflects low standards for receiving a high school diploma (Costrell, 1994). Unfortunately, our data are ill-suited to distinguish between these interpretations at the state level. When we get to the district-level analysis below, one finding is that districts' test-based and graduation-based mobility metrics are positively correlated within states ($\rho \approx 0.25$-0.30), which provides some support for the more optimistic interpretation of high graduation-based mobility, at least measured at the district level.

[16] Because our models are corrected for measurement error in the initial percentiles as described above, the finding that the test-based $\bar{O}_{25}$ values are above the 25th percentile does not reflect mean reversion induced by measurement error.

3.2 Academic Mobility for Sub-Groups Within States

In Figures 4, 5 and 6 we report results from versions of Equation (2) where we define student groups (*s*) by third-grade racial/ethnic designations, FRL designations, and school urbanicity (urban, suburban, rural). The entering and outcome percentile ranks are not group-specific, but rather remain normed against the full state distribution. We continue to report simple-average values across the sample states in the figures, with full state-by-state results available in Appendix B.

Recall that once we split the sample into subgroups within states, α_s and β_s are separately identified and contribute unique information. $\bar{O}_{25s}$ continues to serve as the summary measure of total academic mobility. The charts in row 1 of each figure show estimates of α_s, β_s, and $\bar{O}_{25s}$ for the test-score outcomes and the charts in row 2 show the same parameter estimates for graduation.

We begin with Figure 4, which shows results for the splits by race/ethnicity. We compare Asian, Black, Hispanic, and White students.[17] The gaps in $\bar{O}_{25s}$ in the third column of charts show that initially low-performing Asian students have the highest academic mobility. There is a significant gap between Asian students and all other racial/ethnic groups, and this is true for both test-based and graduation-based mobility. In Appendix B we show that this pattern holds with great consistency on a state-by-state basis.

The academic mobility differences between the other racial/ethnic groups are less stark but still clearly present. Within these groups, White students are the most academically mobile, followed by Hispanic and then Black students. The Black-White and Hispanic-White $\bar{O}_{25s}$ gaps in terms of eighth-grade achievement are 3.6 and 0.9 pecentage points, respectively. For on-time graduation these same gaps are 4.1 and 1.9 percentage points.

[17] There is also an "other race/ethnicity" category in the data to capture all other students, but it is a small group and omitted from our focal comparisons.

The Black-White mobility gaps in Figure 4 are consistent with evidence on the widening of the Black-White achievement gap in North Carolina (Clotfelter, Ladd, & Vigdor, 2009) and nationally (McDonough, 2015; Todd & Wolpin, 2007). In contrast, Clotfelter, Ladd, and Vigdor (2009) find that the Hispanic-White achievement gap narrows in North Carolina during grades 3-8, but this result is not replicated in our data. Our findings for the Hispanic-White gap align more closely with evidence from Reardon and Galindo (2009), who find that the Hispanic-White achievement gap is fairly flat from grades 1-5 using a nationally representative sample, and Todd and Wolpin (2007), who find that it remains flat or widens modestly.[18]

Another takeaway from Figure 4 is that variation between racial/ethnic groups in absolute mobility (α_s) drives the variation in total mobility ($\overline{O}_{25s}$). This is reflected in the visual correspondence between the heterogeneity in α_s and $\overline{O}_{25s}$ in the first and third columns of charts. One way to illustrate the importance of absolute versus relative academic mobility is to decompose the total change in $\overline{O}_{25}$ between groups into the portions that reflect differences in α and β. For example, consider the average gap in $\overline{O}_{25}$ between the highest-mobility group—Asian students—and the lowest-mobility group—Black students—in Figure 4. From row 1, column 3 of the figure, and focusing on the eighth-grade test, the $\overline{O}_{25}$ gap is 11.9 percentile points. Of this total gap, 10.6 points are accounted for by the gap in α between Asian and Black students, and only about 1.3 points are accounted for by the gap in β (which is roughly 0.05 in the chart, multiplied by a factor of 25 to map to $\overline{O}_{25}$).[19] The value of α is overstated relative to β by focusing at a point in the distribution below the 50th percentile in this comparison; still, even evaluated at $\overline{O}_{50}$, α is the dominant explanatory factor over the total mobility gap.

[18] A more nuanced explanation of Reardon and Galindo's (2009) findings is as follows: point estimates imply a modest shrinking of the gap in math and a modest increase in reading. Although we do not perform formal tests, based on their reported standard errors it seems likely that their confidence intervals would include our estimates if the analytic approaches were otherwise aligned.

[19] The numbers from this calculation are depicted in the charts, but difficult to infer precisely visually. The raw data are available in Appendix B.

Next, Figure 5 shows analogous splits by third-grade FRL status. Compared to FRL students, non-FRL students have much higher $\bar{O}_{25}$ values. In terms of test scores, the gaps are 4.5-6.7 percentage points bewteen FRL and non-FRL students who start at the 25th percentile in the third grade. The on-time and lagged graduation gaps between these groups are 12.5 and 11.0 percentage points, respectively.

The last subgroup comparison is by the urbanicity of the school attended in the third grade, shown in Figure 6. Here there is much less heterogeneity across groups than in the preceding figures. The most notable variation in Figure 6 is in terms of graduation rates and observed most easily in the bottom-right chart, which shows graduation rates for initially low-performing students who attend urban schools are significantly below those of their suburban and rural counterparts (who have similar graduation rates to each other).

Appendix Tables B2, B3, and B4 provide the state-by-state results that underlie Figures 4, 5, and 6. For the comparisons along all three dimensions, there is broad consistency across states in the results. The biggest exception is that the urbanicity gaps exhibit some cross-state variability, although even then the comparisons are qualitatively consistent in the sample states.

3.3 District-Level Variation in Mobility

Figure 7 documents within-state, cross-district heterogeneity in α, β, and $\bar{O}_{25}$ as estimated by equations (3) and (4). Specifically, the figure reports averages of the estimated standard deviations of these district-level parameters for each outcome in each state. These estimates capture the extent to which absolute, relative, and total academic mobility vary across school districts.

The raw variances of $\hat{\alpha}_d$, $\hat{\beta}_d$, and $\hat{\bar{O}}_{25d}$ will overstate the true variances because the raw variances include sampling variance. We net out the sampling variance using a randomized inference procedure in which we randomly assign students to districts, then estimate "null distributions" of $\hat{\alpha}_d$, $\hat{\beta}_d$, and $\hat{\bar{O}}_{25d}$ that entirely reflect sampling variance. We repeat this procedure 300 times and use the average variance across the 300 null distributions as an estimate of the sampling variance to be subtracted from each raw-variance estimate.[20] To illustrate, define $\sigma^2_{\hat{\alpha}}$ as the unadjusted variance of $\hat{\alpha}_d$ as estimated using the real data in a given state, and $\bar{\sigma}^2_{\hat{\alpha},null}$ as the average value of the null-distribution variance of $\hat{\alpha}_d$ with random student assignments to districts. The standard deviation of the parateter of interest, α_d, net of sampling variance, can be estimated as:

$$\sigma_\alpha = \sqrt{\sigma^2_{\hat{\alpha}} - \bar{\sigma}^2_{\hat{\alpha},null}} \tag{5}$$

A similar procedure is applied to obtain error-variance-corrected estimates of σ_β and σ_{O25}.

The null distributions from the randomized inference procedure also allow for direct tests of statistical significance of the variances of these parameters. We say that the variance of a given parameter across districts is statistically signficant in a state if the variance estimate using the observed data falls outside of the 95-percent confidence interval of the null-distribution values.

A second, smaller issue is with regard to interpreting cross-parameter variance differences. This issue arises because α, β, and $\bar{O}_{25}$ are not measured in the same units. We are especially interested in comparing the variance magnitudes of α_d and β_d, as this speaks to the importance of absolute versus relative academic mobility in driving variation in total academic mobility. Accordingly, for comparability purposes, we multiple β_d by 50 and calculate the variance of $50*\beta_d$ to reflect the effect of variability in β_d assessed at the center of the initial rank distribution. This allows for an appropriate comparison of variance magnitudes between α_d and β_d given that the variance of α_d is indepenent of initial rank.

The average standard deviations across sample states after the error-variance correction are shown in Figure 7, and in the case of β_d, after multiplying the estimates by 50. The variances

[20] We replicate this procedure just 200 times in Texas due to complications associated with computing demands (and noting that the Texas database is especially large).

across districts for all parameters, all outcomes, and in all states are statistically signficant. The results indicate that one standard deviation in the distribution of total academic mobility, assessed at the 25th percentile of the entry distribution, corresponds to a change in student rank on the eighth-grade test of about 4.8 percentile points, on average. The standard deviation with respect to the high-school test is the similar, at 4.9 percentage points. For on-time and delayed graduation rates, the analogous values are are 5.5 and 4.8 percentage points, respectively.

Adding context from Figure 3, these estimates imply that a third-grade entrant at the 25th percentile who attends a district with mobility that is one-standard-deviation above average on the high school test would be expected to score at the 35.3rd percentile in the state distribution, compared to the 30.4th percentile at the average district. In terms of on-time graduation, a similar comparison at the 25th percentile of the entry distribution would yeild a graduation likelihood at the high-mobility district of 81.3 percent, versus 75.8 percent at the average district.

The results in Figure 7 also make clear that the variance of α_d consistently exceeds the magnitude-aligned variance of β_d. For the eighth-grade and high-school test outcomes, respectively, the average standard deviation of α_d is 72 and 97 percent larger than the average standard deviation of β_d; for on-time and lagged graduation it is 56 and 65 percent larger. Appendix Table B5 shows that α_d has a larger standard deviation than β_d in all states and for all outcomes.

These results make clear that districts exhibit greater variability in their academic-mobility intercepts than their slopes. This result provides insight into how districts potentially impact initially low-achieving students. Specifically, the much larger variance of α_d suggests the potential for generally high-performing districts to positively impact initial low-achievers, while the lower variance of β_d suggests less scope for districts to improve outcomes for low-performers by targeting instruction effectively at certain points in the distribution. It is useful to be mindful in this interpretation that our estimates are not causal, but the variance ratio of α_d to

β_d is large and suggestive about the scope for impact of school districts along different dimensions, absent reforms to current practice.

Finally, noting that the cross-outcome comparisons in this section are made independently, interpretation of these results is aided by understanding how the measures of academic mobility across outcomes are related within districts. Table 4 presents correlation matrices that document these relationships, adjusted for estimation error following Kraft (2017). The table shows that the district-level mobility metrics are positively correlated across outcomes. After adjusting for estimation error, the correlations within outcome mode are very high—for test-score outcomes the estimated correlation is 0.88, and for graduation outcomes it is 1.0. The correlations between test-score-based and graduation-based metrics range from 0.28-0.33.[21]

4. Extension: District-Level Correlates of Academic Mobility

In this section we explore links between academic mobility and district characteristics by regressing our estimates of $\overline{O}_{25d}$ on the percentage of students who are (a) Black, (b) Hispanic, (c) FRL eligible, (d) participants in an individualized education plan (IEP), and (e) geographically mobile. We also use a Theil index to measure school segregation within districts following CHKS.[22] All of these metrics are constructed for school districts using data from cohort students in the third-grade year. A final district measure we use to explain academic mobility is district-level value added to student test scores in math and ELA from grades 4-8. Our value-added estimates capture district contributions to student test score growth in both subjects conditional on student characteristics. We estimate value added using data from the same time periods during which we follow the analytic cohorts in each state, but jackknife the measures around the analytic cohorts to remove any mechanical correlation between our

[21] Correlations that are not adjusted for estimation error are reported in Appendix Table B8—they are lower, but substantively similar.

[22] The Theil index measures the degree of racial/ethnic segregation in a district and ranges from 0 (where all schools within a district have the same racial/ethnic composition as the district as a whole) to 1 (where racial/ethnic groups are entirely segregated between schools within a district). Districts with only one school are dropped from our analysis of district segregation as the Theil index is undefined.

academic-mobility and value-added metrics. Appendix D provides estimation details for our value-added models.

Figure 8 shows the predictors of $\bar{O}_{25d}$. We report average coefficients across the seven states from district-level univariate and multivariate regressions where the dependent variable is $\bar{O}_{25d}$ and the independent variables are the district characteristics described above. The independent variables are standardized in each state to have a mean of zero and variance of one. The coefficient averages thus reflect the predicted change in academic mobility associated with a one-standard-deviation move in the district distribution of the independent variable, on average across the states.[23] The detailed state-by-state regression output summarized by Figure 8 is provided, along with information about statistical significance, in Appendix Table B6.

The preceding analysis offers some predictions about the directions of the coefficients, particularly in the univariate regressions, for which the results should map closely to the results in Figures 4 and 5 for racial/ethnic- and FRL-share differences across districts. Indeed, the first chart in Figure 8 shows that higher underrepresented minority shares and higher FRL shares correspond with lower academic mobility. More broadly, all indicators of student disadvantage in the univariate regressions—student shares by URM, FRL, and geographic mobility—are negatively related to academic mobility on average across the states, as is the school segregation index. The other clear result from the univariate regressions is that district value-added is positively associated with academic mobility. The value-added associations are somewhat stronger in the regressions of test-based academic mobility, which is not surprising, but positive for the graduation-based mobility metrics as well.[24]

[23] A one-standard-deviation change with respect to the value-added measure is based on the raw data. Given that the value-added measures are shrunken using the approach of Lefgren and Sims (2012), a one-standard-deviation change in the raw data corresponds to more than a one-standard-deviation change in the true (unobserved) distribution of value added (Chetty, Friedman, and Rockoff, 2014a).

[24] The average coefficients on value added in the graduation-based $\bar{O}_{25}$ models are buoyed by particularly large estimates in Michigan. The coefficients in all states are positively signed, and many are statistically significant, but the estimates in Michigan are larger than others. See Appendix Table B6.

Turning to the multivariate regressions, district value added and the share of FRL students continue to predict academic mobility as in the univariate regressions. The associations between value added and academic mobility are essentially unchanged in the multivariate regressions, which follows from the two-step procedure we use to construct the value-added measures following Parsons, Koedel, and Tan (2019), as described in Appendix D. The predictive influence of the FRL share increases in the multivariate regressions, highlighting this variable as the primary variable among the indicators of student disadvantage. The associations with the other disadvantage metrics attenuate substantially, and even flip signs in some cases, in the multivariate regressions.

5. Connecting Intragenerational Academic Mobility to Intergenerational Economic Mobility

CHKS document geographic heterogeneity in intergenerational economic mobility and suggest that differences in schools may contribute to this heterogeneity. In this section we explore this possibility using our measures of intragenerational academic mobility. Given well-documented relationships between family income and student achievement (Jang and Reardon, 2019; Reardon, 2011), and student achievement (and achievement-inducing interventions) and earnings (Chetty et al., 2011; Chetty, Friedman, and Rockoff, 2014b; Lazear, 2003; Murnane et al., 2011), it is reasonable to hypothesize that all else equal, areas with higher academic mobility will have higher economic mobility.

Connecting our academic-mobility metrics to CHKS' economic-mobility metrics is is not straightforward; indeed, several issues limit our ability to draw strong inference about how the two metrics relate. One issue is that the time frames for the two metrics are mismatched. Linking the estimates can still be justified if we assume that at least some aspects of place that contribute to the different types of mobility are fixed, but some divergence is expected due to the time inconsistency (note that due to data limitations, it is not possible for us to go back further in time to better align our measures to the CHKS measures). A second issue is research design. In particular, a statistical association between the two measures—if such an association could be established—would not necessarily indicate a causal relationship (also see Rothstein, 2019). Neither our academic mobility metrics, nor the CHKS economic mobility metrics, are causally identified. However, these issues are largely moot because of a third issue we turn to next—namely, there is insufficient scope for variation in academic mobility to explain variation in economic mobility, at least at the level of aggregation at which economic mobility is measured by CHKS.

CHKS estimate economic mobility at the commuting-zone (CZ) level. A CZ is a relatively large geographic unit of anaylsis intended to reflect the local economy where people live and work. Our district-level analysis is much more fine-grained—a typical CZ contains many school districts. A quick empirical investigation reveals that most of the cross-district variation in academic mobility occurs within, and not across, commuting zones. Specifically, within each state, and focusing on CZs for which at least 50 percent of the population resides within a sample state (CZs are not constrained by state boundaries), we run district-level regressions where the dependent variable is $\overline{O}_{25d}$ and the independent variables consist of a vector of indicator variables for the CZs. We interpret error-corrected R-squared estimates from these regressions as measuring the cross-CZ variance in academic mobility.[25] This is the only variance that we can feasibly connect to the CHKS commuting-zone measures; i.e., we must throw out all within-CZ variance in academic mobility.

Figure 9 shows the average cross-CZ variance shares of district-level academic mobility in the sample states for each late-grade outcome. For the test-based measures, the cross-CZ variance shares are between 0.14 and 0.16. They are somewhat higher for the graduation outcomes, but still top out at 0.30. The implication is that most of the variation we measure in

[25] For the error correction we use output from the randomized inference procedure described above to estimate the share of the variance in the district O25 estimates that reflects true variance (net of estimation error), then divide the raw R-squared values by this ratio, as in Aaronson, Barrow and Sander (2007). This adjustment re-scales the R-squared to be over the range of explainable variance in the dependent variable. Without this correction, the cross-CZ variance share will be understated by the R-squared value because the estimation-error variance passes through to the error term.

academic mobility across districts cannot be connected to the variation across CZs reported by CHKS. This result is not surprising as education research consistently shows the greatest impacts of interventions at narrower localities—i.e., individual differences between teachers are larger than differences between schools, which are larger than differences between districts, etc.[26] Still, the realized magnitudes of the between-CZ variance shares are quite low, limiting our ability to meaningfully link variation in the two types of mobility.

We flesh this out further with back-of-the-envelope calculations in Appendix E. Our calculations are based on (a) the distributions of academic mobility across districts, (b) the cross-CZ variance shares, and (c) estimates from the literature mapping changes in test performance and graduation outcomes to earnings. We show that depending on the late-grade outcome we consider, a one-standard-deviation move in the academic mobility distribution across CZs, converted to income gains, maps to a move in the cross-CZ economic mobility distribution estimated by CHKS of just 0.00-0.08 standard deviations. These small numbers indicate the limited potential for variation in academic mobility to explain variation in economic mobility across commuting zones, even assuming away other problems related to research design and causal inference. Simple tests confirm that values in the range of 0.00-0.08 are well below the thresholds at which statistical relationships can be detected using CZ-level data in our sample states. Our findings in this regard are consistent with related evidence from Rothstein (2019). He uses entirely different data and methods but reaches the substantively similar conclusion that differences in the distribution of human capital accumulation can, at best, explain only a small fraction of the measured variance in economic mobility across commuting zones.

While our findings corroborate Rothstein (2019) by showing that schools do not (meaningfully) explain cross-CZ variance in economic mobility, this should not be taken to imply that schools do not matter. Indeed, most of the variance we estimate in academic mobility

[26] This result is also in line with recent, related place-based work by Schoefer and Ziv (2021), who show that most of the measured variance in productivity across cities is driven by plant-level productivity differences. In the education context, Laliberte (2021) finds that differences in schools are an important driver of place-based effects on students' educational attainment using narrowly-defined geographic areas.

occurs within commuting zones. Laliberte (2021) conducts a much narrower geographic analysis of neighborhoods and uses causal tools to show that access to school quality accounts for 50-70 percent of the benefits of moving to a better neighborhood as measured by gains in students' long-term educational attainment.

6. Conclusion

We introduce the concept of "academic mobility" and use it to study variation across school districts in the distributional stickiness of students' education trajectories. Using administrative data from seven states covering nearly 3 million students, we estimate statistically significant and economically meaningful differences in academic mobility across school districts. Initially low-performing students who attend districts that are one standard deviation higher in the academic mobility distribution gain about five percentile points on tests in the eighth grade and high school relative to their peers who attend districts with average mobility. They are also about five percentage points more likely to graduate from high school.

Our analysis of academic mobility across student groups divided by race-ethnicity, eligibility for free and reduced-price lunch, and school urbanicity produces patterns that are largely as expected based on existing research, which has typically been conducted over far shorter timespans. For example, initially low-performing Asian and White students have higher academic mobility than Black and Hispanic students, and higher-income students have higher mobility than lower-income students. Still, some results stand out, such as the very large and consistent upward mobility advantage among Asian students relative to other racial/ethnic groups. It is also notable that initially low-performing students who attend schools in rural areas do not have less upward mobility than their suburban peers, on average, which is at odds with the prevailing theme of the "rural schools problem" in education research (Burton, Brown, and Johnson, 2013).

Our decomposition of total academic mobility into its components reveals that differences across districts in absolute mobility are the primary driver of cross-district variance in total academic mobility. This suggests low-performing students experience the largest performance gains when attending districts where students generally excel. It also casts doubt on the narrative that districts vary substantially in the degree to which they narrow achievement gaps internally, at least given current educational practice and absent reforms that make it a greater priority to improve outcomes among initially low-achieving students.

Finally, we use our estimates of academic mobility to gain insight into the scope for differences in school quality to explain geographic variation in economic mobility across commuting zones, as estimated by CHKS. We find that variation in academic mobility cannot explain a meaningful fraction of the variance in economic mobility across commuting zones, corroborating related findings from Rothstein (2019). That said, this does not rule out more substantive impacts of schools within narrower geographic areas, as we find that most of the cross-district variance in academic mobility (70-85 percent) occurs within commuting zones (also see Laliberte, 2021).

It bears repeating that our academic mobility metrics do not carry a causal interpretation. We do not know if our estimates reflect the true impacts of the local areas we define by school districts, or something else (e.g., the selection of families). Moreover, if we overcome this hurdle and can recover causal estimates of these areas—an objective we intend to pursue in future research—it will still be difficult to assess what it is about them that drives the findings (inclusive of factors inside and outside of schools). These problems are endemic to the burgeoning field of place-based research (Chetty, Hendren, and Katz, 2020; Harding et al., 2021; Kaestner, 2020). Noting this important caveat about causality, our findings illuminate broad patterns in academic mobility and suggest directions for future research aimed at improving the ability of the education system to produce more equal outcomes for students.

References

Aaronson, Daniel, Lisa Barrow and William Sander. 2007. Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25(1), 95-135.

Black, S.E., & Devereux, P.J. (2011). "Recent developments in intergenerational mobility," in *Handbook of Labor Economics Vol. 4* (eds. O. Ashenfelter and D. Card), 1487-1541. Amsterdam: Elsevier.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Measuring Test Measurement Error: A General Approach. *Journal of Educational and Behavioral Statistics* 38(6), 629-663.

Burton, M., Brown, K., and Johnson, A. (2013). Storylines about rural teachers in the United States: A narrative analysis of the literature. *Journal of Research in Rural Education* 28(12), 1-12.

Castex, G., & Kogan Dechter, E. (2014). The changing roles of education and ability in wage determination. *Journal of Labor Economics* 32(4), 685-710.

Cawley, J., Heckman, J., & Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics* 8(4), 419-442.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. *Quarterly Journal of Economics* 126(4), 1593-1660.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014a). Measuring the impacts of teacher I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593-2632.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014b). Measuring the impacts of teacher II: Teacher value-added and student outcomes in adulthood. *American Economic Review* 104(9), 2633-2679.

Chetty, R., Hendren, N., Jones, M.R., & Porter, S.R. (2018). Race and economic opportunity in the United States: An intergenerational perspective. National Bureau of Economic Research Working Paper No. 24441.

Chetty, R., Hendren, N., and Katz, L.F. (2020). Response to "The moving to opportunity experiment: What do heterogeneous estimates of the effect of moving imply about causes?" *Econ Journal Watch* 17(2), 299-304.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics* 129(4), 1553-1623.

Cicala, S., Fryer, R. G., and Spenkuch, J. L. (forthcoming). Comparative advantage in social interactions. *Journal of the European Economic Association*.

Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy* 122(2), 282-318.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2015). The aftermath of accelerated algebra: Evidence from district policy initiatives. *Journal of Human Resources* 50(1), 159-188.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2009). The Academic Achievement Gap in Grades 3 To 8. *Review of Economics and Statistics* 91(2), 398-419.

Costrell, R.M. (1994). A simple model of educational standards. *American Economic Review* 84(4), 956-971.

Cunha, F., Heckman, J.J., Lochner, L., & Masterov, D.V. (2006). "Interpreting the evidence on life cycle skill formation," in *Handbook of the Economics of Education Vol. 1* (eds. E. Hanushek & F. Welch), 697-812. Amsterdam: Elsevier.

de Gruijter, D.N.M. (1997). On Information of Percentile Ranks. *Journal of Educational Measurement* 34(2), 177-78.

Denning, J. T., Murphy, R., and Weinhardt, F. (2020). Class rank and long-run outcomes. National Bureau of Economic Research working paper no. 27468.

Domina, T., McEachin, A., Penner, A., & Penner, E. (2015). Aiming high and falling short: California's eighth-grade algebra-for-all effort. *Educational Evaluation and Policy Analysis* 37(3), 275-295.

Elsner, B., and Isphording, I. E. (2017a). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics* 35(3), 787-828

Elsner, B., and Isphording, I. E. (2017b). Rank, sex, drugs and crime. *Journal of Human Resources* 53(2), 356-381

Elsner, B., Isphording, I.E., and Zoelitz, U. (2019). Achievement Rank Affects Performance and Major Choices in College. University of Zurich working paper.

Ferrer, A., & Riddell, W. (2008). Education, credentials, and immigrant earnings. *The Canadian Journal of Economics* 41(1), 186-216.

Goldhaber, D., & Ozek, U. (2019). How much should we reply on student test achievement as a measure of success? *Educational Researcher* 48(7), 479-483.

Goldhaber, D., Wolff, M., and Daly, T. (2020). Assessing the accuracy of elementary test scores as predictors of students' high school outcomes. CALDER Working Paper No. 235-0520.

Grigg, J. (2012). School enrollment changes and student achievement growth: A case study in educational disruption and continuity. *Sociology of Education* 85(4), 388-404.

Harding, D.J., Sanbonmatsu, L., Duncan, G.J., Gennetian, L.A., Katz, L.F., Kessler, R.C., Kling, J.R., Sciandra, M., and Ludwig, J. (2021). Evaluating contradictory experimental and non-experimental estimates of neighborhood effects on economic outcomes for adults. NBER Working Paper No. 28454. Cambridge, MA: National Bureau of Economic Research.

Jang, H., & Reardon, S.F. (2019). States as sites of educational (in)equality: State contexts and the socioeconomic achievement gradient. *AERA Open* 5(3), 1-22.

Kaestner, R. (2020). The moving to opportunity experiment: What do heterogeneous estimates of the effect of moving imply about causes? *Econ Journal Watch* 17(2), 282-298.

Koedel, C., and Parsons, E. (2021). The effect of the community eligibility provision on the ability of free and reduced-price meal data to identify disadvantaged students. *Educational Evaluation and Policy Analysis* 43(1), 3-31.

Kraft, M.A. (2017). Teacher effects on complex cognitive skills and social-emotional competencies." *Journal of Human Resources*, vol. 54, no. 1, 2017, pp. 1-36.

Laliberte, J.W. (2021). Long-term contextual effects in education: Schools and neighborhoods. *American Economic Journal: Economic Policy* 13(2), 336-377.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review* 10(3), 179–214.

Lefgren, L., & Sims, D.P. (2012). Using subject test scores efficiently to predict teacher value-added. *Educational Evaluation and Policy Analysis* 34(1), 109-121.

Leithwood, K. (2010). Characteristics of school districts that are exceptionally effective in closing the achievement gap. *Leadership Policy Schools* 9, 245–291.

Lockwood, J.R. and McCaffrey, D.F. (2014). Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects. *Journal of Educational and Behavioral Statistics* 39(1), 22-52.

May, K., & Nicewander, W.A. (1994). Reliability and Information Functions for Percentile Ranks. *Journal of Educational Measurement* 31(4), 313-25.

McDonough, I.K. (2015). Dynamics of the black-white gap in academic achievement. *Economics of Education Review* 47, 17-33.

Mehana, M., & Reynolds, A.J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review* 26(1), 93-119.

Mulligan (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy* 107, S184–S224.

Murnane, R.J., Willett, J.B., Duhaldeborde, Y., & Tyler, J.H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management* 19(4), 547-568.

Murnane, R.J., Willet, J.B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics* 77(2), 251-266.

Murphy, R., and Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *Review of Economic Studies* 87, 2777–2826.

Parsons, E., Koedel, C., Podgursky, M., Ehlert, M., Xiang, P.B. (2015). Incorporating End-of-Course Exam Timing into Educational Performance Evaluations. *Journal of Research on Educational Effectiveness* 8(1): 130-147

Parsons, E., Koedel, C., & Tan, L. (2019). Accounting for Student Disadvantage in Value-Added Models. *Journal of Educational and Behavioral Statistics* 44(2), 144-179.

Reardon, S.F. (2011). "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations," in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (eds. G.J. Duncan & R.J. Murnane), 91-115. New York, NY: Russell Sage Foundation.

Reardon, S.F., & Galindo, C. (2009). The Hispanic-White Achievement Gap in Math and Reading in the Elementary Grades. *American Educational Research Journal* 46(3), 853-891.

Rorrer, A.K., Skrla, L., & Scheurich, J.J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly* 44(3), 307-358.

Rothstein. J. (2019). Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income. *Journal of Labor Economics* 37(S1), S85-S123.

Schoefer, B., and Ziv, O. (2021). Productivity, place, and plants: Revisiting the measurement. Unpublished manuscript. UC Berkeley Department of Economics.

Solon, G., (1999). "Intergenerational mobility in the labor market," in *Handbook of Labor Economics Vol. 3* (eds. O. Ashenfelter and D. Card), 1761-1800. Amsterdam: Elsevier.

Solon, G. (1992). Intergenerational income mobility in the United States. *American Economic Review* 82(3), 393-408.

Sorenson, L.C. (2019). "Big data" in educational administration: An application for predicting school dropout risk. *Educational Administration Quarterly* 55(3), 404-446.

Todd, P.E., & Wolpin, K.I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital* 1(1), 91-136.

Wang, M. & Stanley, J. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research* 40, 663-705.

Figures and Tables

Figure 1. Hypothetical illustrations of the linear rank-rank relationship. No mobility (left) versus perfect mobility (right).

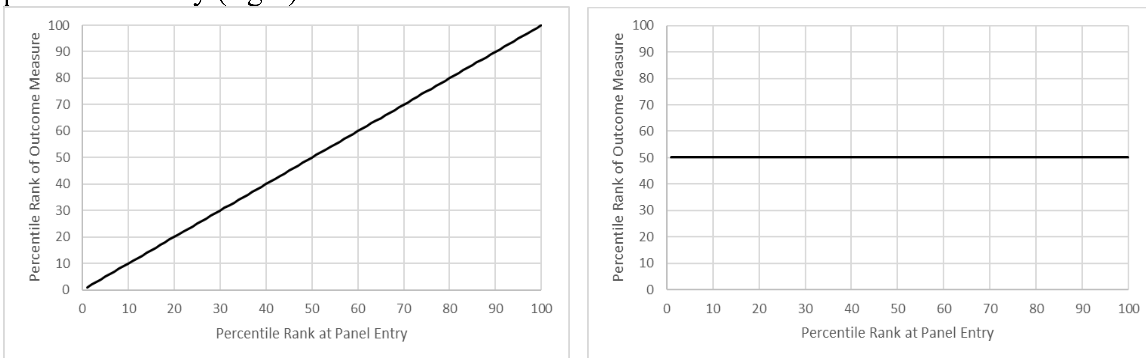


Figure 2. Comparison of two hypothetical student groups, one with higher relative mobility (solid lines) and one with lower relative mobility (dashed lines), with differing gaps in absolute mobility.

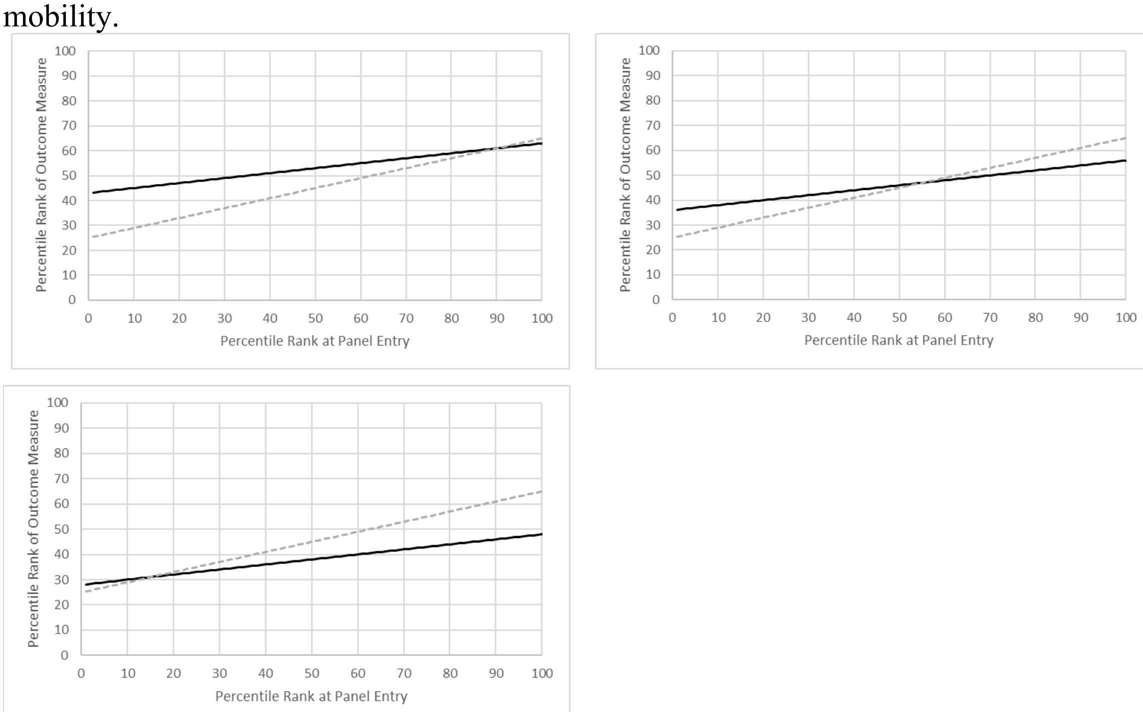
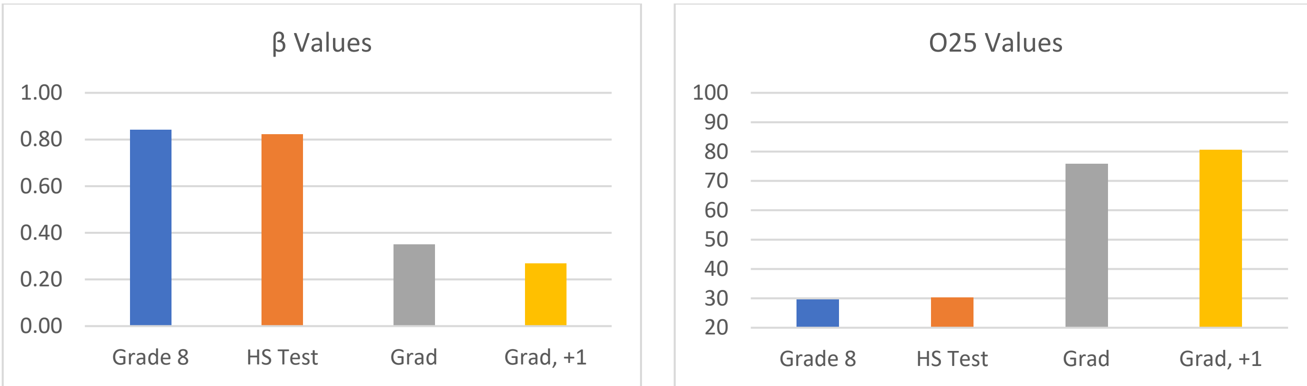


Figure 3. Simple averages of state-level estimates of β and $\overline{O}_{25}$ for each outcome.



Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. α is redundant when all statewide data are used, as described in the text. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. Full results broken out by each state individually are reported in Appendix B.

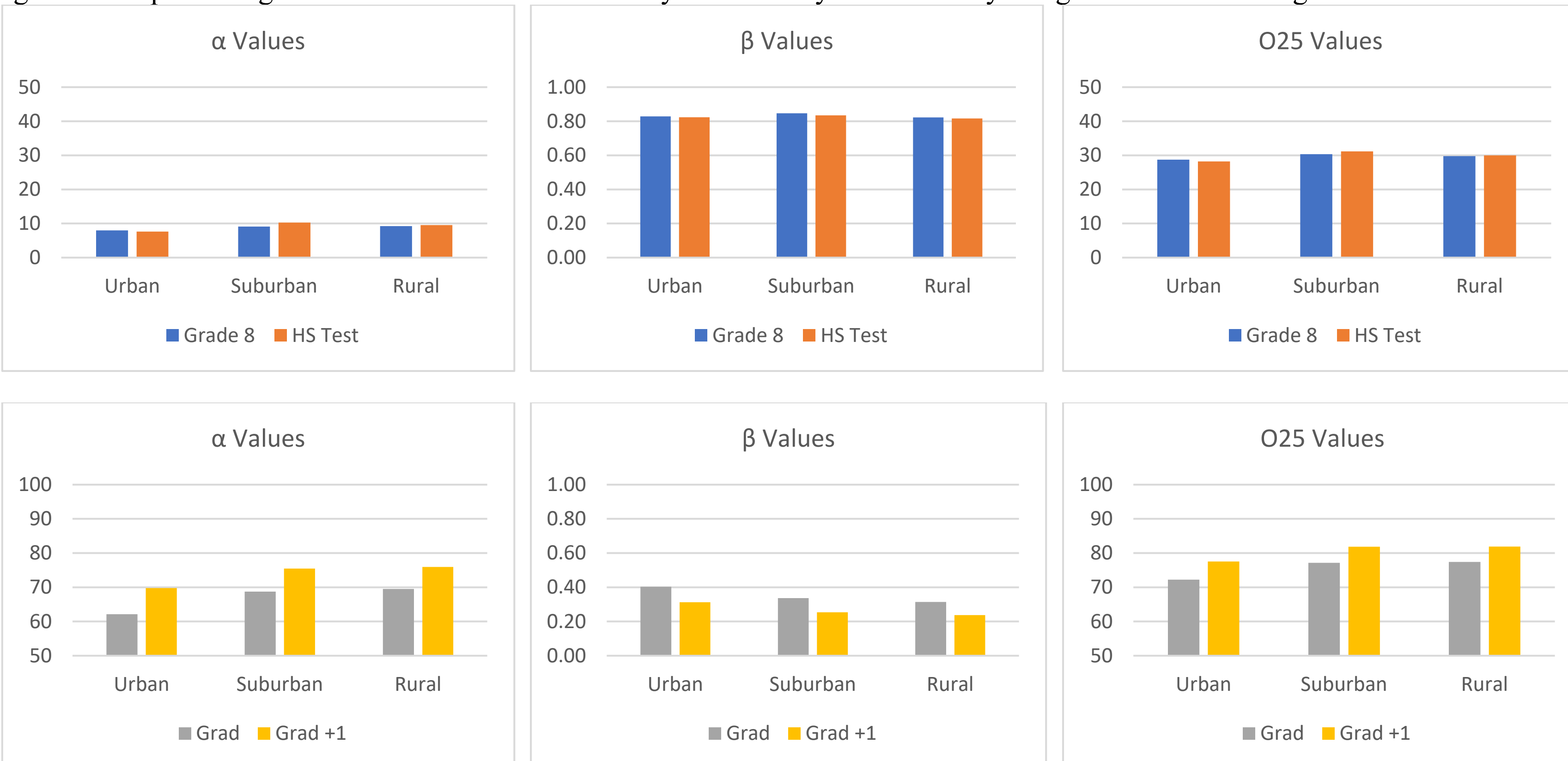
Figure 4. Simple averages of statewide academic mobility measures by race/ethnicity.

Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. Full results broken out by each state individually are reported in Appendix B.

Figure 5. Simple averages of statewide academic mobility measures by FRL status.

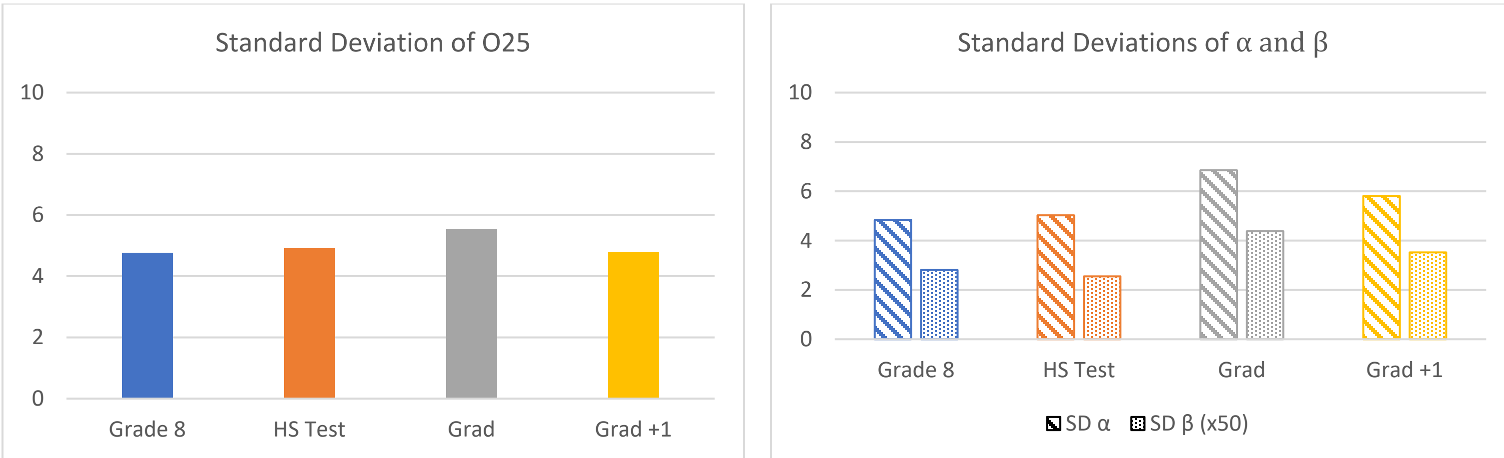
Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. Full results broken out by each state individually are reported in Appendix B.

Figure 6. Simple averages of statewide academic mobility measures by the urbanicity designation of the third-grade school.



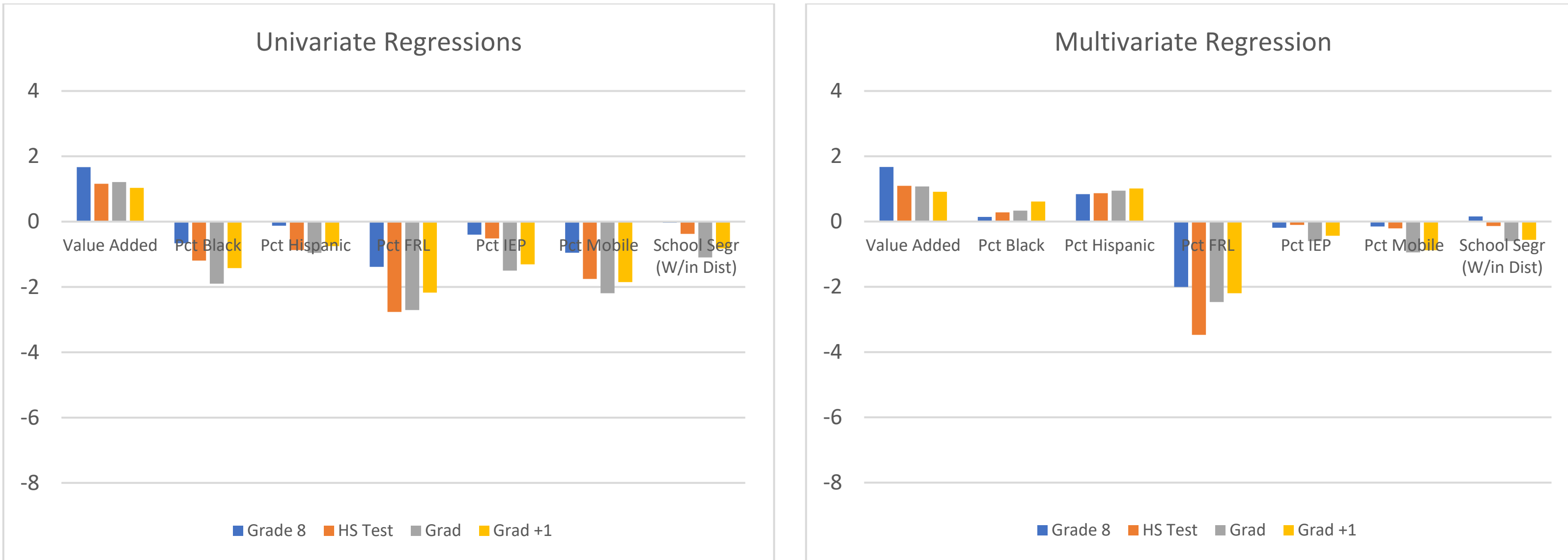
Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. Full results broken out by each state individually are reported in Appendix B.

Figure 7. Simple averages of the within-state, cross-district standard deviations of $\bar{O}_{25d}$, α_d, and β_d.



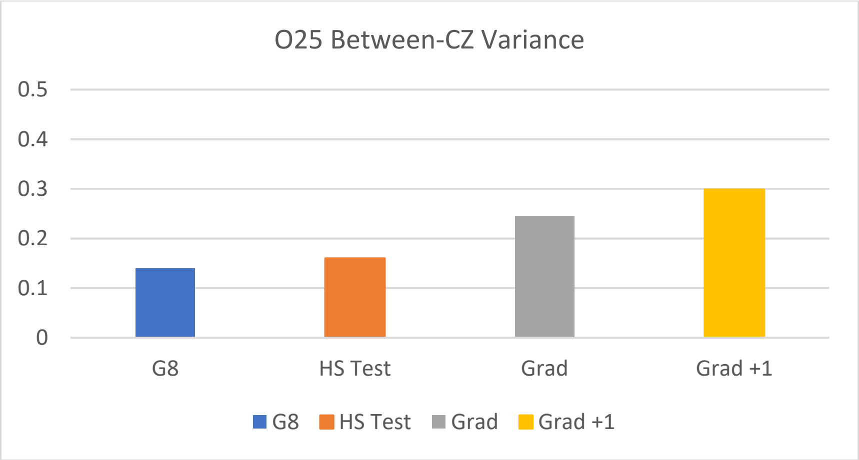
Notes: The standard deviations of β_d are multiplied by 50 to align the magnitudes of variances of α_d and β_d for comparative purposes in this figure. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. Full results broken out by each state individually are reported in Appendix B.

Figure 8. Average coefficients on district-level predictors of $\bar{O}_{25d}$. Univariate and multivariate regression results.



Notes: All predictors are standardized to have a variance of one so the average coefficients can be interpreted as showing associations with one-standard-deviation moves in each predictor, on average across states. For the value-added measures, the standard deviations are in raw values; noting that these are shrunken estimates, a one-standard-deviation move is equal to more than one standard deviation in the true distribution (Chetty, Friedman, and Rockoff, 2014). State-by-state regression output underlying these graphs, including information about the statistical significance of the relationships in each state, is reported in Appendix B.

Figure 9. Average cross-commuting-zone variance shares of $\overline{O}_{25d}$.



Notes: These values are the average (error-corrected) R-squared values from a regression of $\overline{O}_{25d}$ on a vector of CZ indicators in each state. One minus these values gives the average within-CZ variance shares.

Table 1. Definition of the analytic sample and descriptive statistics at panel entry for each state.

	Cohort Years	N (entry cohorts)	Pct. Black	Pct. Hispanic	Pct. FRL	Pct. IEP	Pct. Mobile	Pct. Urban	Pct. Suburban	# of Districts	# of Schools	Private Schl Enrl % (2008)
Georgia	2007-2009	376,427	38.08	12.72	56.02	12.47	8.12	8.81	39.66	182	1255	8.71
Massachusetts	2007-2008	139,337	7.83	13.94	31.65	17.30	2.32	20.11	68.19	304	1,116	13.60
Michigan	2006-2009	453,946	19.03	5.72	40.99	10.92	12.19	20.96	44.39	755	2,039	8.59
Missouri	2006-2009	264,612	18.17	4.00	46.34	15.16	6.62	18.79	30.87	548	1,200	12.05
Oregon	2006-2008	123,833	3.03	16.83	47.59	15.37	4.03	30.69	25.60	208	1,086	10.49
Texas	2006-2009	1,309,114	13.54	47.68	57.84	5.86	6.68	42.27	27.90	1,173	4,338	5.96
Washington	2006-2008	218,051	5.70	15.80	42.26	11.44	1.04	26.12	45.30	296	1,254	9.17
Entire U.S.	2008	--	17.04	21.13	42.95	12.35	--	29.03	35.10	--	--	--

Table Notes: "Cohort Years" refers to the years of panel entry for the cohorts included in the analytic sample, i.e. the years in which the students were in the third grade. The spring year is used to indicate the academic year (e.g., 2009 = 2008-09 school year). Students who took both the Math and ELA third-grade state tests are included in the core sample. For Washington and Massachusetts, in earlier years of data, enrollment surveys were not conducted frequently, which likely contributes to the low reported mobility rates in those two states. In more recent data, the mobility rates in Massachusetts and Washington are around 5 and 8-9 percent, respectively. Note that the numbers of schools and districts indicate the numbers of unique schools and districts included in the analysis in each state. Data for the "Entire U.S." are reported in the bottom row of the table for context and taken from the 2008 common core of data and are for students in public K-12 elementary and secondary grades. Note that we do not report a mobility percentage because a comparable variable is not available in the common core of data.

Table 2. High school exams by state.

	HS Exam	Grade Typically Taken	Pct. Of Cohort Students Taking the Exam On-Grade	Pct. Of Cohort Students Taking the Exam Within 1 Year of On-Grade
Georgia	American Lit EOC	9	97.7	2.0
Massachusetts	MCAS ELA	10	99.5	0.2
Michigan	ACT/SAT	11	99.3	0.7
Missouri	English II EOC	10	93.1	3.8
Oregon	--	--	--	--
Texas	Reading/English II EOC	10	94.1	5.7
Washington	HSPE ELA, SBAC ELA	10, 11	98.3	1.4

Notes: In Washington a test change led to the change in the grade in which the third grade cohorts took their high school exit exams (from grade 10 to 11), as shown in the Table. Michigan transitioned from the ACT to the SAT in the 2016-17 school year. The first two analysis cohorts took the ACT in 11th grade, the second two cohorts took the SAT in 11th grade. In Oregon there is no single high school test given to more than 90 percent of students in a fixed grade to support our analysis of mobility using HS academic achievement.

Table 3. Documentation of sample attrition in each state and for each late-grade outcome.

		Original Cohort Members					
		Panel Entry	Observed with Outcome		Observed without Outcome		
		N	N	Avg. Outcome Pctl. or Grad Rate	N	Avg. Entry Pctl.	Avg. Imputed Outcome Pctl. or Grad Rate
Grade 8 – Combined Math and ELA	Georgia	376,427	308,624	49.58	67,803	41.21	41.55
	Massachusetts	139,337	124,606	49.41	14,731	46.51	47.99
	Michigan	453,946	395,263	49.41	58,683	39.44	41.83
	Missouri	262,366	227,459	50.69	34,907	47.68	46.63
	Oregon	123,833	105,674	50.44	18,159	45.70	44.07
	Texas	1,280,996	1,094,987	48.73	186,009	49.29	53.68
	Washington	218,051	185,501	49.98	32,550	45.24	45.26
High School Exam	Georgia	376,427	310,207	50.43	66,220	44.96	45.27
	Massachusetts	139,337	114,374	49.31	24,963	46.12	47.23
	Michigan	453,946	346,705	50.40	107,241	39.38	39.45
	Missouri	262,366	205,634	51.23	56,732	42.73	40.53
	Oregon	--	--	--	--	--	--
	Texas	1,280,996	1,095,603	50.57	185,393	41.11	44.19
	Washington	218,051	172,229	51.02	45,822	42.71	42.69
Graduation (On-Time)	Georgia	376,427	314,346	80.29	62,081	43.75	69.83
	Massachusetts	139,337	114,413	93.92	24,924	46.13	90.23
	Michigan	453,946	392,186	84.97	61,760	45.13	77.79
	Missouri	262,366	210,423	91.08	51,943	46.10	85.99
	Oregon	123,833	101,692	80.99	22,141	47.43	70.51
	Texas	1,280,996	1,129,684	84.27	151,312	41.59	76.79
	Washington	218,051	176,505	82.66	41,546	43.38	70.33
Graduation (Within One Year of On Time)	Georgia	376,427	314,346	83.76	62,081	43.75	76.21
	Massachusetts	139,337	114,413	94.18	24,924	46.13	90.61
	Michigan	453,946	392,186	87.86	61,760	45.13	81.72
	Missouri	262,366	210,423	93.59	51,943	46.10	89.81
	Oregon	123,833	101,692	82.62	22,141	47.43	72.39
	Texas	1,280,996	1,129,684	87.73	151,312	41.59	82.82
	Washington	218,051	176,505	86.66	41,546	43.38	76.29

Notes: Sample sizes and entry percentiles are based on the average of the grade 3 math and reading percentiles (i.e., percentiles at entry). For the test outcomes, the mean of each rank distribution should be 50 but in several states it deviates (very) slightly from 50 because of lumpiness in the underlying test-score distributions. For graduation outcomes, we report the percent of students who graduate among stayers because percentiles are not informative. *In Michigan, lagged graduation outcomes are not available for the 2009 cohort and the numbers reported in the bottom panel omit that cohort.*

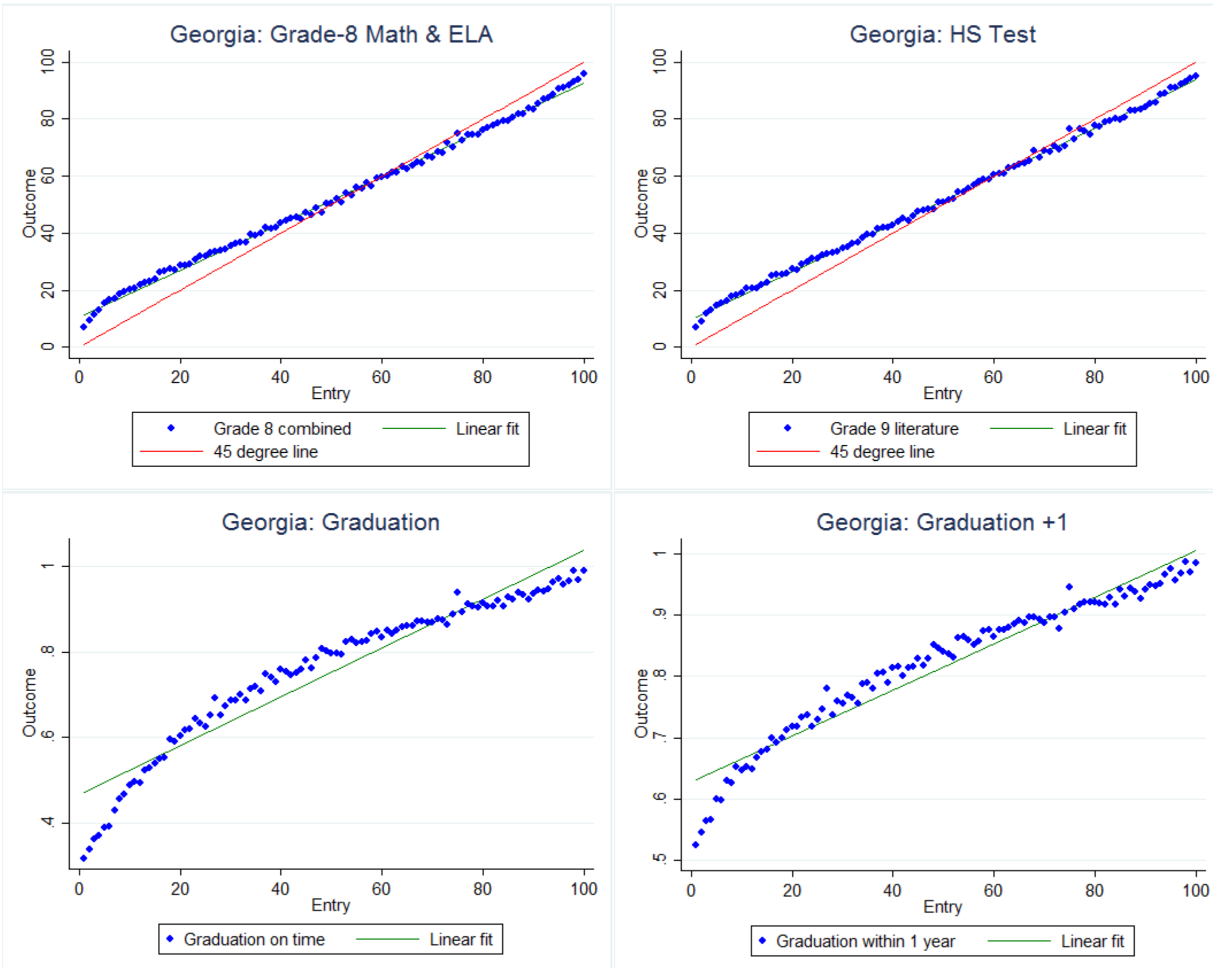
Table 4. Average correlations of district-level mobility metrics across outcomes for the sample states, with estimation-error correction.

	8^{th} Grade Test	High School Test	On-Time Grad	Grad W/in One Year of On Time
8^{th} Grade Test	1.00	-	-	-
High School Test	0.88	1.00	-	-
On-Time Grad	0.33	0.31	1.00	-
Grad W/in One Year of On Time	0.32	0.28	1.00	1.00

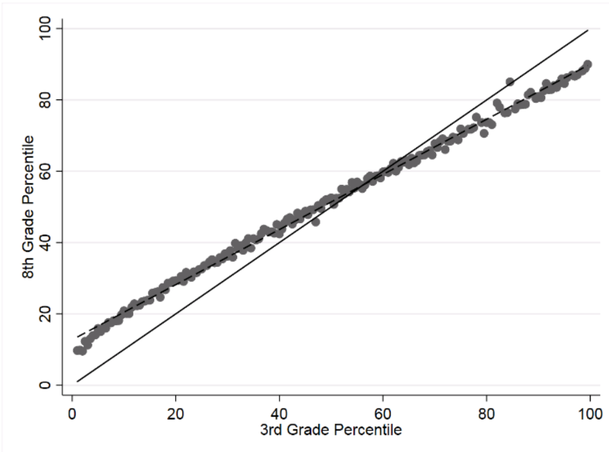
Notes: The uncorrected analogs to these numbers are available in Appendix B.

Appendix A: Linear rank-rank regression output for selected states

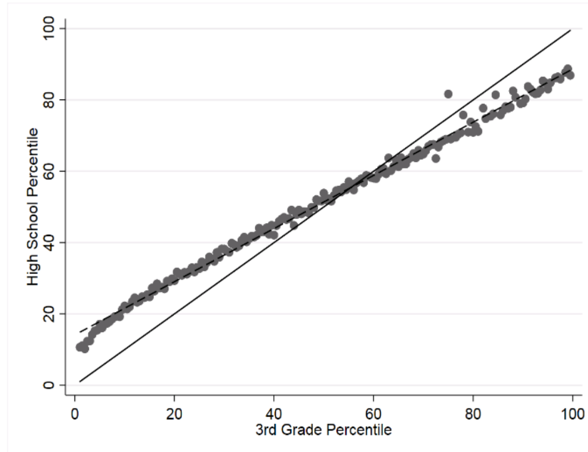
Appendix Figure A1. Binned scatter plots of percentiles on percentiles for each outcome to assess the linearity of the rank-rank relationships.



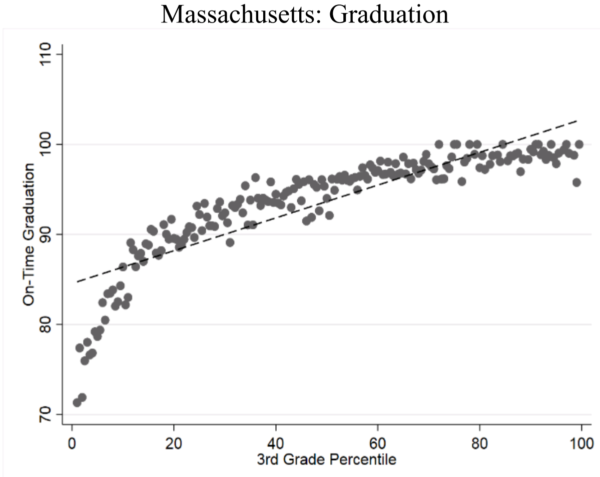
Massachusetts: Grade-8 Math & ELA



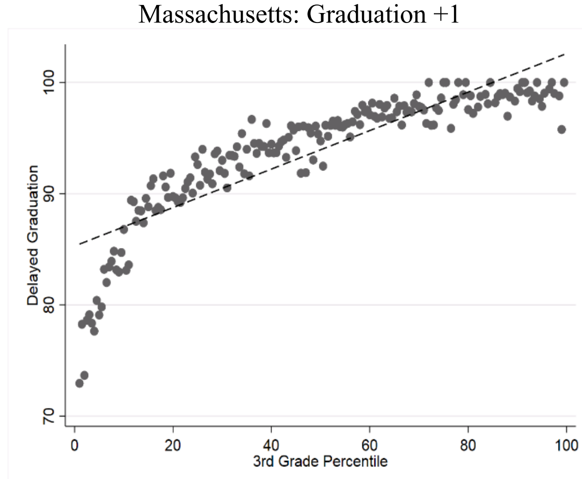
Massachusetts: HS Test



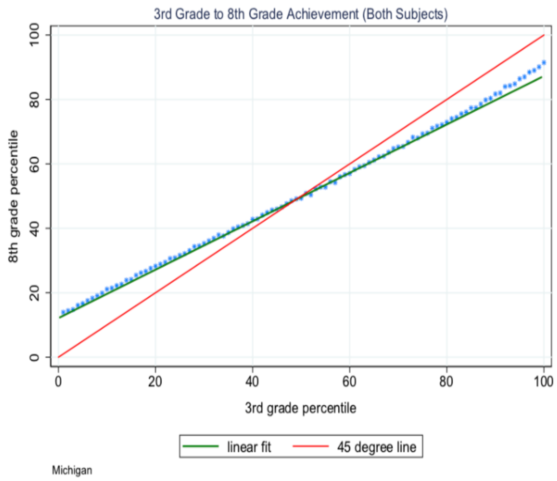
Massachusetts: Graduation



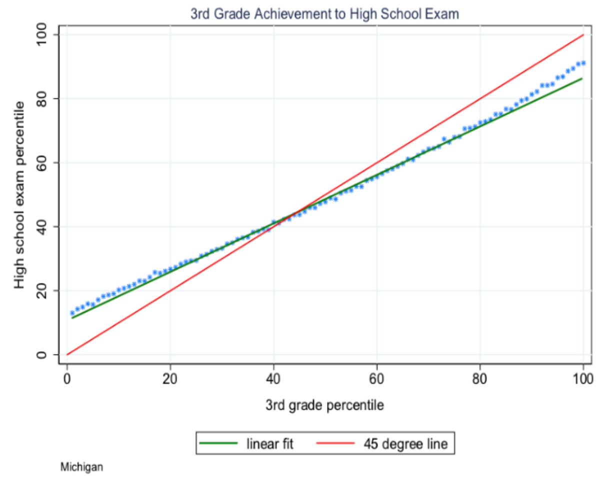
Massachusetts: Graduation +1



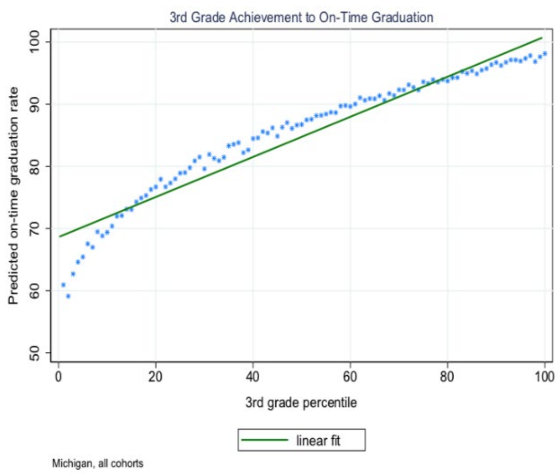
Michigan: Grade-8 Math & ELA



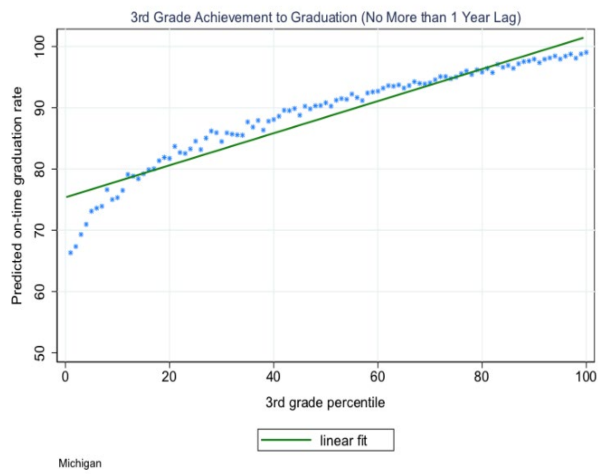
Michigan: HS Test



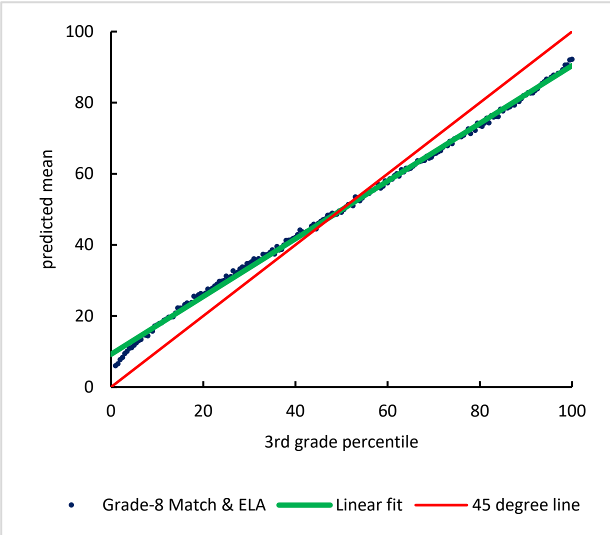
Michigan: Graduation



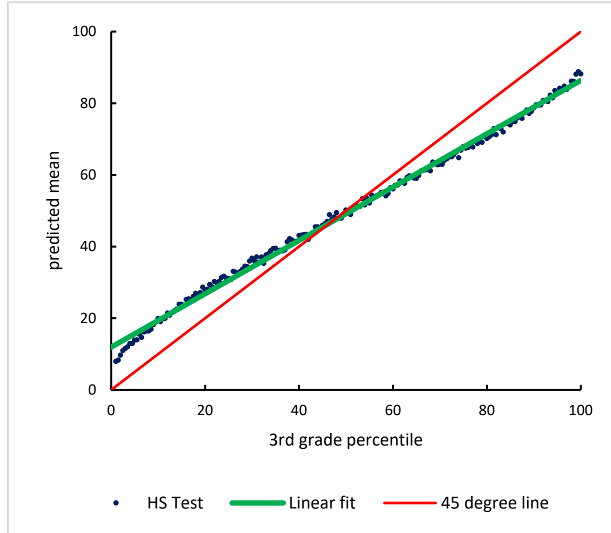
Michigan: Graduation +1



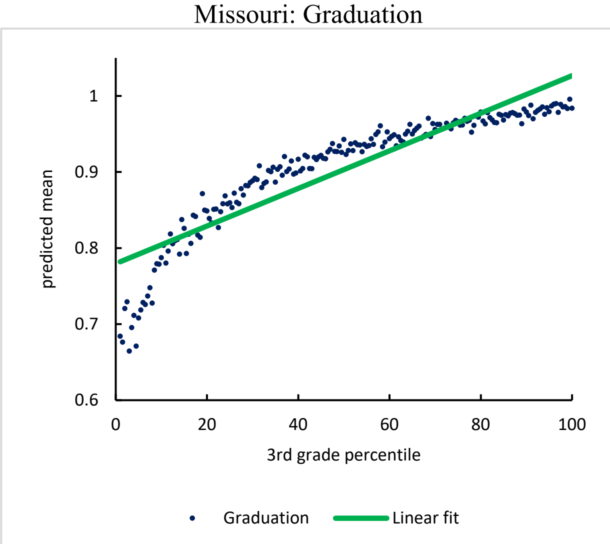
Missouri: Grade-8 Math & ELA



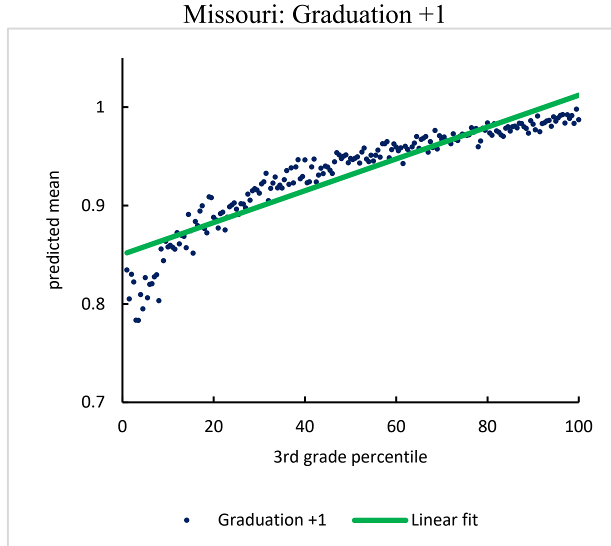
Missouri: HS Test



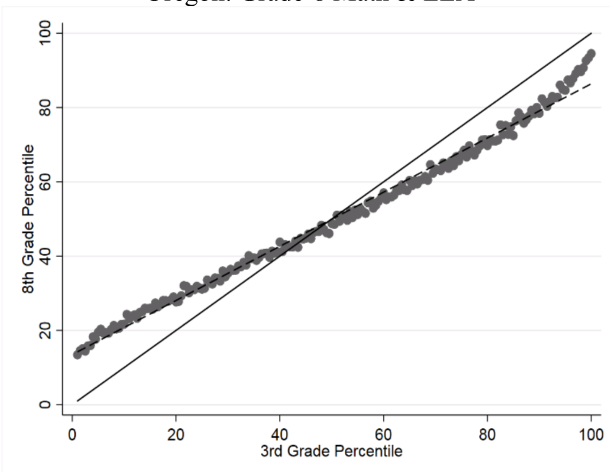
Missouri: Graduation



Missouri: Graduation +1



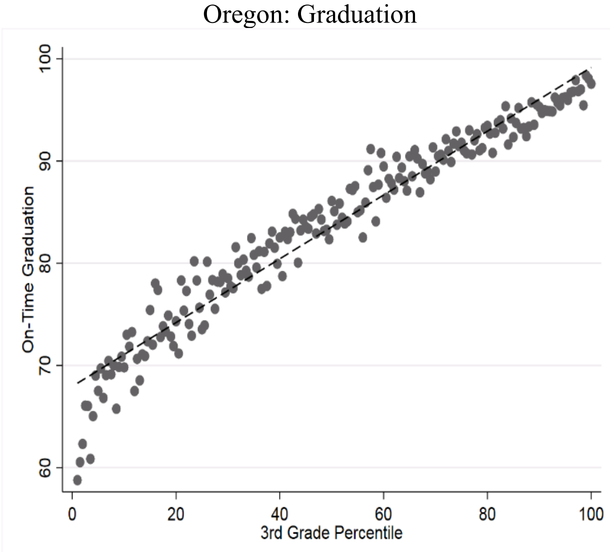
Oregon: Grade-8 Math & ELA



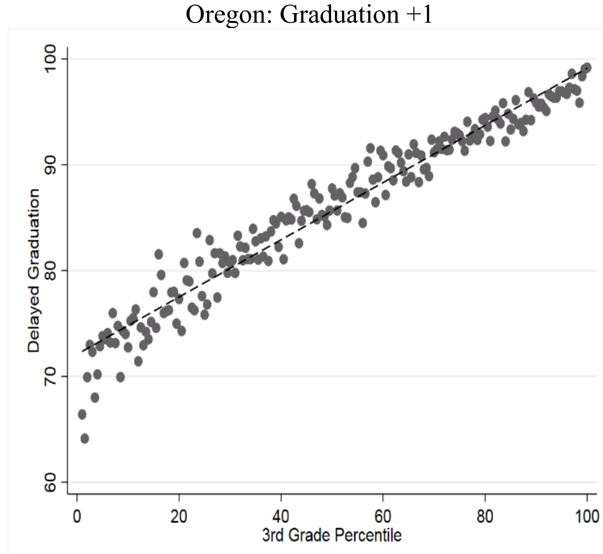
Oregon: HS Test

(omitted)

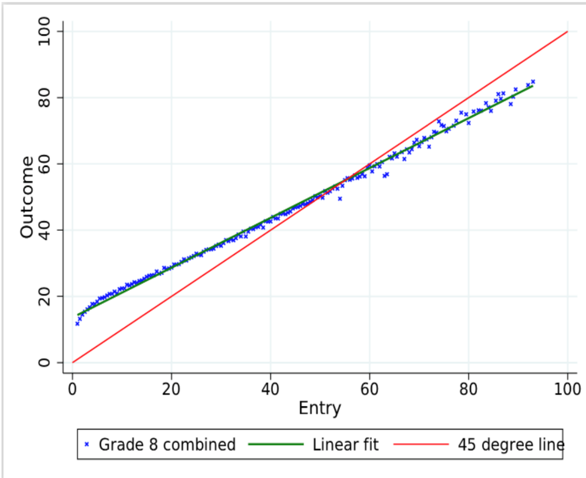
Oregon: Graduation



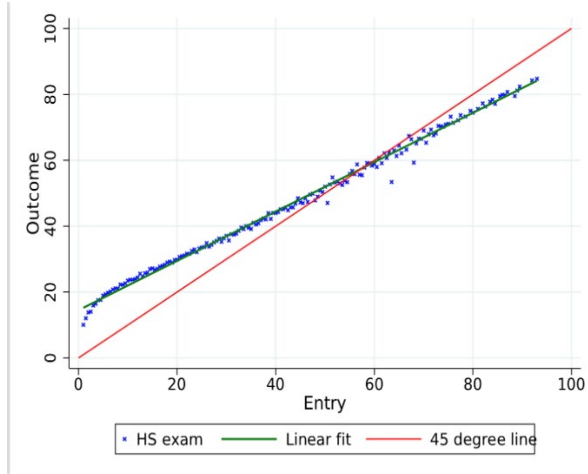
Oregon: Graduation +1



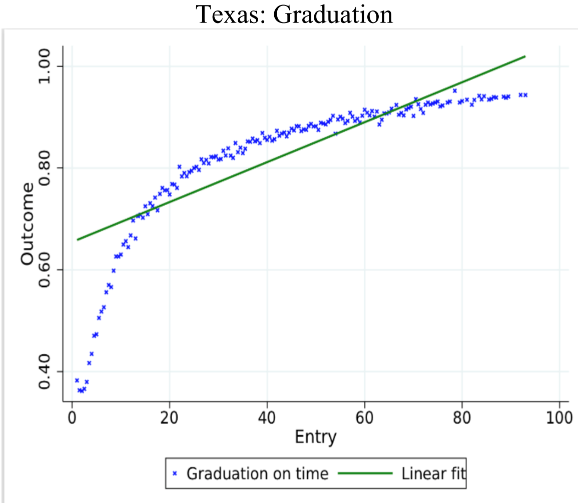
Texas: Grade-8 Math & ELA



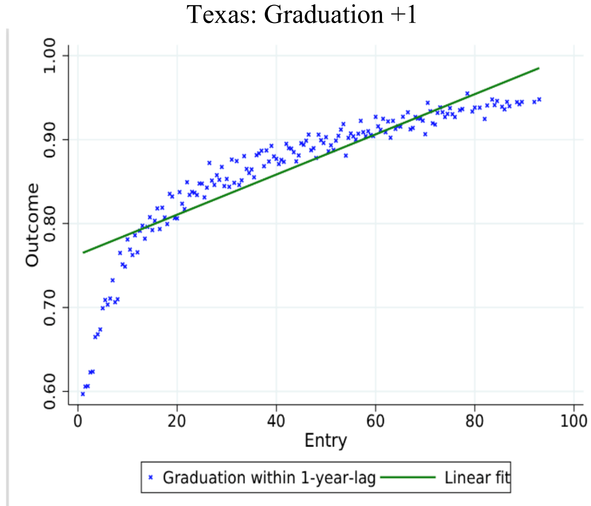
Texas: HS Test



Texas: Graduation

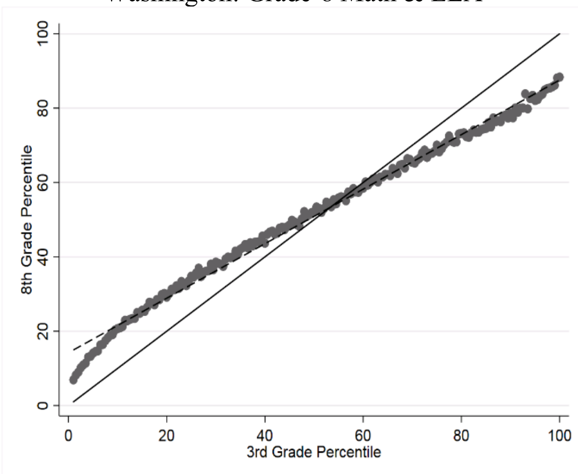


Texas: Graduation +1

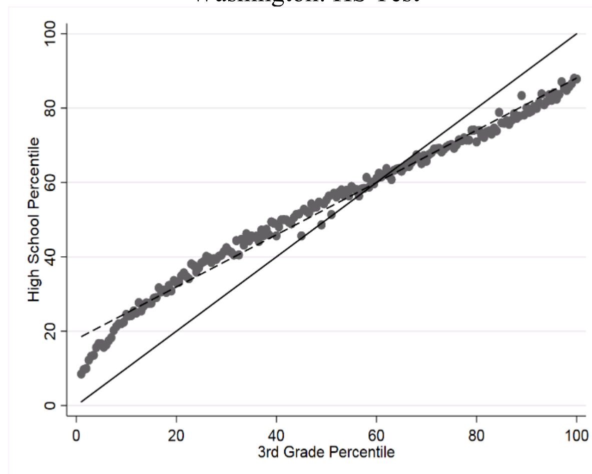


Note: Vertical and horizontal axes are scaled from 0-100 in percentiles.

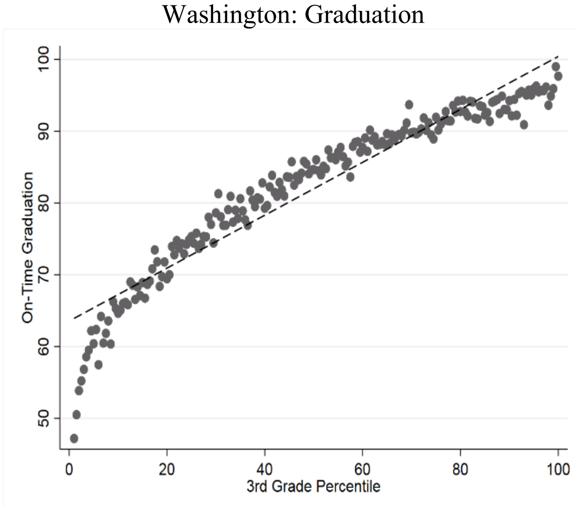
Washington: Grade-8 Math & ELA



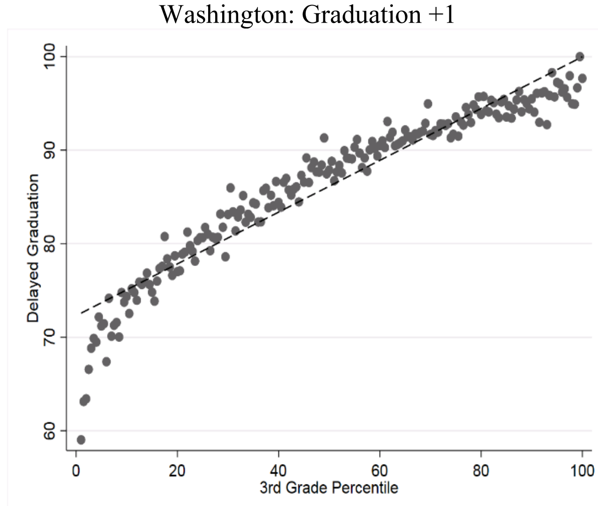
Washington: HS Test



Washington: Graduation



Washington: Graduation +1



Appendix B: Detailed state-by-state results & supplementary tables

Appendix Table B1. State-by-state numeric results corresponding to Figure 3.

	Grade-8 Test		HS Test		Grad		Grad +1	
	β	O25	β	O25	β	O25	β	O25
All (Avg)	***0.84***	***29.66***	***0.82***	***30.44***	***0.35***	***75.76***	***0.27***	***80.59***
GA	0.86	29.92	0.86	29.78	0.52	66.03	0.39	73.12
MA	0.84	29.56	0.83	29.65	0.19	88.73	0.19	89.23
MI	0.84	28.48	0.80	29.82	0.35	76.02	0.30	80.22
MO	0.87	28.18	0.82	29.97	0.25	83.61	0.17	88.58
OR	0.81	29.86			0.33	71.03	0.29	73.63
TX	0.85	31.69	0.80	33.16	0.43	73.77	0.26	81.39
WA	0.82	29.94	0.83	29.68	0.39	71.15	0.29	77.94

Notes: The notes to Figure 3 apply.

Appendix Table B2. State-by-state numeric results corresponding to Figure 4.

	Grade-8 Test			HS Test			Grad			Grad +1		
Student Group: Asian												
	α	β	O25	α	β	O25	α	β	O25	α	β	O25
All (Avg)	18.43	0.82	39.00	19.76	0.81	39.91	79.58	0.19	84.93	84.44	0.16	88.36
GA	19.93	0.80	40.02	16.90	0.82	37.45	66.63	0 .36	75.74	74.53	0.27	81.33
MA	18.99	0.82	39.50	20.85	0.80	40.81	91.43	0.10	93.84	92.09	0.09	94.30
MI	16.59	0.85	37.75	14.48	0.89	36.73	83.20	0.18	87.64	87.09	0.14	90.53
MO	15.61	0.87	37.40	18.81	0.81	39.03	86.42	0.14	90.05	90.65	0.10	93.07
OR	14.09	0.84	34.99				75.88	0.23	81.68	78.51	0.20	83.52
TX	28.08	0.77	47.24	32.82	0.71	50.52	81.68	0.18	86.26	89.17	0.08	91.21
WA	15.75	0.81	36.07	14.68	0.81	34.93	71.82	0.30	79.32	79.06	0.22	84.59
Student Group: Black												
	α	β	O25	α	β	O25	α	β	O25	α	β	O25
All (Avg)	7.81	0.77	27.06	8.45	0.76	27.36	62.94	0.39	72.73	71.10	0.29	78.32
GA	8.12	0.79	28.09	7.56	0.81	27.69	51.83	0.57	65.95	63.26	0.41	73.55
MA	9.82	0.75	28.49	10.33	0.71	28.18	80.70	0.21	85.89	81.92	0.19	86.70
MI	8.15	0.71	25.88	5.09	0.71	22.73	62.60	0.39	72.35	68.35	0.34	76.86
MO	5.62	0.78	25.09	8.12	0.75	26.82	70.11	0.32	78.18	79.01	0.21	84.29
OR	5.40	0.76	24.49				57.60	0.32	65.65	63.53	0.26	70.14
TX	10.25	0.80	30.21	12.53	0.75	31.32	61.20	0.51	73.94	74.91	0.29	82.17
WA	7.29	0.79	27.15	7.05	0.81	27.41	56.51	0.42	67.13	66.74	0.31	74.53
Student Group: Hispanic												
	α	β	O25	α	β	O25	α	β	O25	α	β	O25
All (Avg)	10.25	0.78	29.76	9.91	0.77	29.25	65.85	0.36	74.86	73.17	0.27	79.90
GA	11.28	0.82	31.67	9.75	0.82	30.42	51.57	0.53	64.71	62.62	0.38	72.24
MA	9.10	0.74	27.59	8.96	0.71	26.77	76.16	0.27	82.98	77.19	0.26	83.69
MI	8.83	0.80	28.75	6.35	0.82	26.96	66.66	0.32	74.69	72.04	0.28	78.94
MO	9.30	0.82	29.85	12.21	0.77	31.37	76.09	0.25	82.42	83.49	0.16	87.57
OR	11.58	0.73	29.77				67.25	0.27	74.00	70.50	0.24	76.43
TX	11.47	0.78	31.04	13.41	0.73	31.70	61.13	0.50	73.64	74.39	0.30	81.80
WA	10.16	0.78	29.62	8.79	0.78	28.30	62.07	0.38	71.61	71.93	0.27	78.66
Student Group: White												
	α	β	O25	α	β	O25	α	β	O25	α	β	O25
All (Avg)	9.93	0.83	30.69	11.24	0.82	31.66	68.79	0.32	76.82	74.79	0.25	81.08
GA	10.29	0.85	31.53	11.34	0.84	32.42	53.52	0.50	66.06	62.40	0.39	72.25
MA	9.06	0.84	29.95	10.09	0.81	30.38	88.40	0.13	91.72	88.86	0.13	92.04
MI	9.16	0.82	29.69	7.89	0.85	29.10	70.58	0.31	78.30	75.86	0.26	82.34
MO	7.75	0.86	29.24	10.80	0.81	30.94	81.14	0.20	86.21	87.41	0.13	90.63
OR	10.18	0.80	30.24				61.48	0.35	70.21	64.91	0.31	72.72
TX	13.03	0.84	33.97	17.23	0.78	36.69	64.22	0.38	73.60	72.73	0.26	79.26
WA	10.02	0.81	30.20	10.08	0.81	30.45	62.16	0.38	71.61	71.35	0.28	78.31

Notes: The notes to Figure 4 apply.

Appendix Table B3. State-by-state numeric results corresponding to Figure 5.

	Grade-8 Test			HS Test			Grad			Grad +1		
Student Group: FRL												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	9.03	0.76	28.11	8.61	0.75	27.47	62.32	0.35	71.15	69.94	0.26	76.48
GA	8.78	0.79	28.61	7.96	0.79	27.86	48.99	0.53	62.45	60.24	0.39	70.02
MA	8.92	0.73	27.09	9.04	0.69	26.40	77.13	0.23	82.80	78.11	0.22	83.49
MI	8.31	0.75	26.98	5.55	0.78	24.99	60.89	0.34	69.37	66.69	0.30	74.17
MO	6.47	0.81	26.79	8.46	0.76	27.52	73.00	0.26	79.62	81.12	0.17	85.37
OR	10.54	0.72	28.56				60.12	0.25	66.30	63.83	0.21	69.09
TX	11.06	0.77	30.34	12.74	0.72	30.65	59.14	0.50	71.55	72.63	0.29	79.81
WA	9.13	0.77	28.42	7.89	0.78	27.37	56.96	0.36	65.99	66.97	0.26	73.44

Student Group: non-FRL												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	12.36	0.82	32.59	14.28	0.80	34.18	77.99	0.23	83.64	83.39	0.16	87.45
GA	13.28	0.82	33.96	15.22	0.81	35.47	67.46	0.35	76.27	75.61	0.25	81.95
MA	11.39	0.82	31.90	12.89	0.79	32.71	92.94	0.08	94.87	93.35	0.07	95.16
MI	10.28	0.82	30.86	8.88	0.85	30.15	79.91	0.21	85.14	84.36	0.17	88.53
MO	9.29	0.86	30.71	14.50	0.78	34.02	86.76	0.15	90.43	91.91	0.08	94.04
OR	12.48	0.80	32.50				73.42	0.24	79.47	76.50	0.21	81.72
TX	15.10	0.83	35.80	20.48	0.76	39.41	73.39	0.28	80.40	81.97	0.16	86.06
WA	12.59	0.79	32.41	13.68	0.79	33.34	72.08	0.27	78.93	80.00	0.19	84.70

Notes: The notes to Figure 5 apply.

Appendix Table B4. State-by-state numeric results corresponding to Figure 6.

	Grade-8 Test			HS Test			Grad			Grad +1		
Student Group: Urban												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	7.97	0.83	28.73	7.65	0.82	28.22	62.14	0.40	72.24	69.76	0.31	77.53
GA	5.84	0.81	26.34	6.29	0.83	26.97	47.00	0.56	61.20	57.59	0.44	68.50
MA	8.40	0.81	28.53	9.00	0.76	28.12	77.07	0.26	83.62	78.19	0.25	84.38
MI	7.02	0.84	27.99	2.69	0.89	24.91	62.62	0.41	72.90	68.40	0.36	77.31
MO	4.83	0.85	26.17	7.38	0.81	27.54	68.86	0.33	77.11	77.49	0.23	83.17
OR	10.06	0.82	30.66				61.45	0.34	70.00	65.68	0.30	73.08
TX	10.26	0.84	31.29	12.66	0.80	32.62	60.36	0.48	72.31	73.58	0.28	80.67
WA	9.38	0.83	30.12	7.90	0.85	29.15	57.60	0.44	68.51	67.40	0.33	75.61
Student Group: Suburban												
	α	*β*	*O25*	*α*	*β*	*O25*	*A*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	9.11	0.85	30.32	10.30	0.84	31.18	68.75	0.34	77.17	75.49	0.25	81.87
GA	9.10	0.87	30.91	9.39	0.87	31.24	52.35	0.53	65.72	62.36	0.41	72.64
MA	8.91	0.85	30.07	9.76	0.83	30.38	86.81	0.16	90.75	87.36	0.15	91.13
MI	8.21	0.83	29.04	6.69	0.86	28.25	70.82	0.32	78.82	76.29	0.26	82.91
MO	6.29	0.89	28.56	11.55	0.82	32.12	78.43	0.24	84.42	85.35	0.16	89.27
OR	10.08	0.82	30.65				63.78	0.34	72.34	67.26	0.30	74.85
TX	11.41	0.85	32.68	14.47	0.80	34.48	66.23	0.39	75.96	77.77	0.22	83.38
WA	9.76	0.82	30.31	9.94	0.83	30.58	62.81	0.38	72.21	72.05	0.28	78.93
Student Group: Rural												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	9.22	0.82	29.77	9.55	0.82	29.99	69.52	0.31	77.43	75.94	0.24	81.90
GA	9.04	0.84	30.23	8.28	0.85	29.64	55.57	0.49	67.84	65.98	0.35	74.89
MA	9.13	0.83	29.78	9.89	0.81	30.15	88.37	0.13	91.60	88.89	0.12	91.96
MI	9.53	0.80	29.46	7.42	0.83	28.28	70.30	0.31	78.07	75.38	0.27	82.03
MO	7.68	0.85	28.99	10.01	0.80	30.00	81.42	0.20	86.50	87.83	0.13	90.98
OR	9.72	0.78	29.10				63.46	0.31	71.26	66.54	0.28	73.54
TX	10.23	0.85	31.48	13.15	0.79	32.88	64.26	0.40	74.33	74.59	0.25	80.90
WA	9.21	0.81	29.36	8.53	0.82	29.00	63.28	0.36	72.39	72.40	0.26	78.97

Notes: The notes to Figure 6 apply.

Appendix Table B5. State-by-state numeric results corresponding to Figure 7.

	Grade-8 Test			HS Test			Grad			Grad +1		
Standard Deviations												
	α	*β (x50)*	*O25*	*α*	*β (x50)*	*O25*	*α*	*β (x50)*	*O25*	*α*	*β (x50)*	*O25*
All (Avg)	4.84	2.81	4.77	5.03	2.55	4.90	6.85	4.38	5.54	5.81	3.52	4.78
GA	3.41	1.75	3.51	3.87	2.05	4.00	6.62	4.45	6.98	5.93	3.50	6.19
MA	6.50	2.75	6.03	6.53	2.65	6.06	4.86	2.85	3.6	4.62	2.70	3.43
MI	4.25	2.85	4.28	4.87	3.90	5.02	7.92	4.90	6.53	7.23	4.45	5.79
MO	4.14	3.15	4.27	4.37	3.20	4.23	4.88	2.85	3.69	3.05	1.40	2.51
OR	6.77	3.20	6.21				8.99	5.15	6.96	8.64	4.80	6.71
TX	5.52	3.00	5.11	5.79	3.00	5.46	7.42	6.00	5.08	5.63	4.00	4.23
WA	3.28	2.95	3.95	4.72	3.05	4.65	7.26	4.45	5.91	5.54	3.80	4.61

Notes: The notes to Figure 7 apply.

Appendix Table B6. State-by-state numeric results corresponding to Figure 8. Coefficient values on standardized variables.

	Univariate, Grade-8 Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.67	-0.66	-0.12	-1.39	-0.40	-0.95	-0.03
GA	1.29*	-2.04*	1.01*	-2.21*	0.83*	-1.32*	-0.64*
MA	1.79*	-0.74*	-1.92*	-2.52*	-0.54	-2.69*	-0.43
MI	2.35*	-0.41	-0.46	-1.38*	-2.05*	-1.30*	-0.03
MO	1.88*	-1.11*	-0.30	-1.52*	0.47	-1.75*	-0.57*
OR	2.32*	0.15	1.58*	0.68	-0.73	1.31*	1.07*
TX	0.88*	-0.29	-0.46*	-1.62*	-0.65*	0.04	-0.15
WA	1.17*	-0.22	-0.32	-1.13*	-0.13	-0.96*	0.53*

	Multivariate, Grade-8 Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.67	0.14	0.84	-2.00	-0.19	-0.15	0.15
GA	1.58*	-0.50*	0.56*	-1.50*	0.12*	-0.14*	-0.51*
MA	1.79*	1.11*	1.19*	-4.48*	0.22	-1.08*	1.37*
MI	2.29*	0.74	0.05	-1.68*	-0.99*	-0.17	0.33*
MO	1.66*	-0.13	0.54*	-1.21*	-0.03	-0.47	-0.22*
OR	2.17*	0.33	2.05*	-1.10*	-0.10	1.11*	0.21
TX	1.07*	-0.32	0.57*	-2.27*	-0.57*	0.32*	-0.72
WA	1.14*	-0.24	0.94*	-1.79*	0.02	-0.62*	0.62*

	Univariate, HS Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.16	-1.19	-0.87	-2.76	-0.52	-1.76	-0.37
GA	0.47*	-2.06*	1.02*	-2.76*	0.49*	-0.96*	-0.40*
MA	1.46*	-1.05*	-2.54*	-3.31*	-0.97*	-3.07*	-0.93*
MI	2.68*	-2.12*	-0.99*	-2.98*	-3.43*	-2.48*	-0.45*
MO	1.35*	-1.13*	-0.68*	-2.66*	1.31*	-2.68*	-0.64*
OR							
TX	0.22	-0.59*	-0.88*	-2.48*	-0.27	-0.24	-0.04
WA	0.79*	-0.21	-1.14*	-2.39*	-0.23	-1.10*	0.21

	Multivariate, HS Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.09	0.28	0.87	-3.47	-0.10	-0.21	-0.14
GA	0.67*	-1.05*	0.25*	-2.14*	-0.27*	0.43*	-0.49*
MA	1.39*	1.87*	1.93*	-6.29*	0.11	-0.86*	1.10*
MI	2.25*	-0.04	0.11	-2.85*	-0.55	0.17	0.16
MO	1.00*	0.73*	0.65*	-2.77*	0.43	-0.71*	-0.24*
OR							
TX	0.48*	-0.32	0.71*	-3.17*	-0.29*	0.12	-1.56*
WA	0.77*	0.49*	1.55*	-3.60*	-0.05	-0.41	0.21

	Univariate, Grad						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.21	-1.90	-0.96	-2.71	-1.50	-2.19	-1.10
GA	0.19*	-2.09*	0.06*	-1.98*	1.48*	-2.07*	-1.00*
MA	0.64	-1.83*	-2.54*	-1.97*	0.11	0.01	-2.36*
MI	4.76*	-3.19*	-2.52*	-5.51*	-12.62*	-6.26*	-0.52*
MO	0.98*	-2.81*	-1.91*	-3.13*	1.42*	-4.31*	-1.69*
OR	0.69	-0.73	1.30*	-1.58*	-1.23*	-1.59*	-0.37
TX	0.06	-0.78*	-0.31	-2.02*	1.05*	-1.12*	-0.80*
WA	1.18*	-1.87*	-0.79*	-2.75*	-0.72*	-0.01	-0.93*

	Multivariate, Grad						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.07	0.33	0.95	-2.47	-0.60	-0.95	-0.60
GA	0.67*	0.60*	0.43*	-1.43*	0.49*	-0.97*	-0.08*
MA	0.44	-0.58	-1.62	1.20	1.13*	-0.14	-2.27*
MI	3.48*	2.46	-0.44	-4.01*	-7.40*	-3.13*	0.63
MO	0.50*	-0.75*	-0.16	-1.73*	-0.05	-1.07*	-0.72*
OR	0.82*	0.53	4.15*	-3.88*	0.33	-0.80	-0.97
TX	0.34	0.64*	1.80*	-3.04*	1.25*	-0.87*	-0.42
WA	1.28*	-0.55	2.47*	-4.38*	0.05	0.36	-0.37

	Univariate, Grad+1						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	1.03	-1.42	-0.75	-2.17	-1.31	-1.85	-0.82
GA	0.15*	-1.60*	0.00	-1.55*	1.29*	-1.94*	-0.91*
MA	0.64	-1.71*	-2.41*	-1.88*	0.19	-0.04	-2.17*
MI	3.88*	-2.26*	-2.35*	-4.57*	-10.59*	-5.10*	-0.33*
MO	1.04*	-2.16*	-1.55*	-2.37*	1.15*	-3.44*	-1.36*
OR	0.47	-0.41	1.13*	-1.72*	-1.03*	-1.56*	-0.16
TX	0.22	-0.25	0.50*	-0.77*	0.33	-0.94*	-0.20
WA	0.84*	-1.59*	-0.55	-2.36*	-0.51	0.07	-0.61*

	Multivariate, Grad+1						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
All (Avg)†	0.91	0.61	1.01	-2.19	-0.44	-0.88	-0.56
GA	0.61*	1.52*	0.70*	-1.44*	0.53*	-1.20*	-0.11*
MA	0.46	-0.54	-1.59	1.07	1.16*	-0.16	-2.05*
MI	2.86*	2.38	-0.68	-3.33*	-6.24*	-2.92*	0.58
MO	0.63*	-0.42*	-0.11	-1.24*	0.05	-0.96*	-0.67*
OR	0.56	0.83	4.17*	-4.21*	0.52	-0.61	-0.88
TX	0.34	1.01*	2.13*	-2.14*	0.71*	-0.76*	-0.68
WA	0.92*	-0.47	2.47*	-4.07*	0.22	0.47	-0.12

Notes: The notes to Figure 8 apply. The symbol † is to denote that statistical significance is not reported for the "All (avg)" values because they are not directly generated from a regression (they are average values of the state-by-state regression coefficients). Statistical significance at the 5 percent level or better for each coefficient in the state-by-state results is denoted by *. Standard errors are suppressed for brevity.

Appendix Table B7. State-by-state numeric results corresponding to Figure 9.

	Grade-8 Test		HS Test		Grad		Grad +1	
O25 Variance Decompositions								
	Between CZ	Within CZ	Between CZ	Within CZ	Between CZ	Within CZ	Between CZ	Within CZ
All (Avg)	0.14	0.86	0.16	0.84	0.25	0.75	0.30	0.70
GA	0.24	0.76	0.38	0.62	0.25	0.75	0.27	0.73
MA	0.10	0.90	0.08	0.92	0.12	0.88	0.13	0.87
MI	0.06	0.94	0.03	0.97	0.12	0.88	0.16	0.84
MO	0.12	0.88	0.18	0.82	0.28	0.72	0.37	0.63
OR	0.15	0.85			0.59	0.41	0.64	0.36
TX	0.14	0.86	0.13	0.87	0.24	0.76	0.31	0.69
WA	0.17	0.83	0.16	0.84	0.12	0.88	0.21	0.79

Notes: The notes to Figure 9 apply.

Appendix Table B8. Average correlations of district-level mobility metrics across outcomes for the sample states, without adjusting for estimation error.

	8th Grade Test	High School Test	On-Time Grad	Grad W/in One Year of On Time
8th Grade Test	1.00	-	-	-
High School Test	**0.74**	1.00	-	-
On-Time Grad	0.25	0.20	1.00	-
Grad W/in One Year of On Time	0.24	0.20	**0.92**	1.00

Notes: These correlations are averages of the uncorrected correlations and correspond to Table 4 in the main text.

Appendix C: Imputation procedure

We retain the full entering third-grade cohorts throughout our analysis by imputing missing later-grade outcomes. The imputation is performed on an outcome-by-outcome basis—e.g., for a student with an eighth-grade test score, but no high school test score and no data on high school graduation, we retain the observed eighth-grade score for use in our analysis and impute the latter three outcomes.

Imputed values for each missing outcome are a function of student demographics in the third grade (race-ethnicity, gender) along with information on FRL status, English as a second language (ESL) status, IEP status, and available test scores from grades 3-7. For example, for a student who exits one of our sample states after the fifth grade, we impute the four focal later-grade outcomes using information from her profile during grades 3-5. We do not use student characteristics or test scores after the seventh grade for imputation for any student in order to enforce consistency of the imputation procedure across all later-grade outcomes, the first of which is recorded in the eighth grade.

We begin the imputation process by using data for students with all observed later-grade outcomes to estimate a series of regressions of the following form:

$$O_{iq} = \beta_0 + \mathbf{Y_{iq}}\boldsymbol{\beta_1} + \mathbf{X_{1i}}\boldsymbol{\beta_2} + \mathbf{X_{2iq}}\boldsymbol{\beta_3} + \varepsilon_{iq} \qquad \text{(C1)}$$

where O_{iq} is a later-grade outcome for student *i* predicted using student characteristics and test-score records through grade *q* (*q*=3, 4, 5, 6, 7). $\mathbf{Y_{iq}}$ is a vector of test data for student *i* in math and ELA (with test scores standardized by subject-grade-year), the length of which depends on *q*, $\mathbf{X_{1i}}$ is a vector of racial-ethnic and gender designations based on the third-grade record, and $\mathbf{X_{2iq}}$ is a vector containing year-by-year student designations for FRL, ELS, and IEP. We estimate versions of equation (C1) for each later-grade outcome and all five values of *q*, using the samples of students in each state for whom all later-grade outcomes are observed.

The parameters from equation (C1) can be applied to predict later-grade outcomes for students who are missing these outcomes with *q*-values ranging from 3-7 (inclusive). These predictions form the basis of our imputation procedure, to which we make two additional adjustments and extend for sensitivity testing.

The first adjustment is that we add an indicator for within-state district mobility to equation (C1), which we expand as follows:

$$O_{iq} = \delta_0 + \mathbf{Y_{iq}}\boldsymbol{\delta_1} + \mathbf{X_{1i}}\boldsymbol{\delta_2} + \mathbf{X_{2iq}}\boldsymbol{\delta_3} + Z_i\delta_4 + \eta_{iq} \tag{C2}$$

Like terms in equation (C2) are defined as in equation (C1). The addition to equation (C2), Z_i, is an indicator variable equal to one if student *i* is observed changing districts within the state at least once prior to the time at which the outcome is assessed, and zero otherwise. Therefore, the coefficient δ_4 captures the additional predictive power of cross-district mobility within a state over the outcome. Using this adjusted equation, we impute later-grade outcomes for students who are missing these outcomes with the following predictions, where *q* indicates the last grade in which student *i* is observed with a test record in the state data through grade-7:

$$\hat{O}_{iq} = \hat{\delta}_0 + \mathbf{Y_{iq}}\hat{\boldsymbol{\delta}}_{\mathbf{1}} + \mathbf{X_{1i}}\hat{\boldsymbol{\delta}}_{\mathbf{2}} + \mathbf{X_{2iq}}\hat{\boldsymbol{\delta}}_{\mathbf{3}} + Z_i\hat{\delta}_4 \tag{C3}$$

Equation (C3) is the imputation equation used in the primary analysis in the paper. If all students with missing later-grade outcomes were state exiters, these imputed values would be accurate under the assumption that within-state district mobility and cross-state mobility are equally predictive of student outcomes. Treating this as a working assumption is approximately accurate because *most* students with missing outcomes are state exiters (although not all; e.g., in practice, some students miss the tests each year).

The second adjustment is needed because shrinkage is inherent in the predictions in equation (C3). If left unaccounted for, the shrinkage would result in compressed distributions of imputed outcomes relative to the distributions observed in the real data. This is problematic for the test-score outcomes because they are ranked and increasing the weight in the middle of the distribution (due to the shrinkage) will have implications for the measurement of outcomes for all students. We address this issue by inflating the variance of the imputed test scores by a factor θ to align the variance of the imputed values with the variance observed in the real outcome data. The variance-inflation adjustment is not necessary for the graduation outcomes because they are not ranked.

Finally, we extend the imputation framework to examine the sensitivity of our findings to the potential for additional selection into state exit along unobserved dimensions. To do this, we parameterize different levels of selection into state exit, above and beyond what is captured by δ_4. Specifically, we produce four alternative sets of imputed values assuming that the true state-exit mobility parameter is (1) 10 percent larger than δ_4, (2) 25 percent larger than δ_4, (3) 10 percent smaller than δ_4, and (4) 25 percent smaller than δ_4. That is, we allow for varying degrees of positive and negative selection into cross-state mobility, relative to within-state-cross-district mobility.

We suppress the results from the sensitivity analysis for brevity because our findings are not meaningfully sensitive in any way to changing the selection conditions as described in the previous paragraph.

Appendix D: Estimating district value added

We use the larger state data samples of all students in grades 4-8 to estimate district value added with a two-step model based on Parsons, Koedel, and Tan (2019):

$$Y_{ijdkt} = \gamma_0 + \mathbf{Y_{i(t-1)}}\boldsymbol{\gamma_1} + \mathbf{X_{it}}\boldsymbol{\gamma_2} + \mathbf{S_{kt}}\boldsymbol{\gamma_3} + \mathbf{L_{dt}}\boldsymbol{\gamma_4} + \varepsilon_{ijdkt} \tag{D1}$$

$$\varepsilon_{ijdkt} = \phi_d + \nu_{ijdkt} \tag{D2}$$

In equation (D1), Y_{ijdkt} is the test score of student *i* in subject *j* taken at district *d* in school *k* at time *t*, which is standardized by subject, grade, and year within each state. $\mathbf{Y_{i(t-1)}}$ is a vector of test scores in math and ELA taken by student *i* the previous year. $\mathbf{X_{it}}$ is a vector of characteristics of student *i* in time *t* that includes information on the student's FRL status, IEP status, gender, race, English as a second language (ELL) status, and geographic mobility. $\mathbf{S_{kt}}$ and $\mathbf{L_{dt}}$ contain the variables included in $\mathbf{Y_{i(t-1)}}$ and $\mathbf{X_{it}}$ aggregated at the school and district levels, respectively, and ε_{ijdkt} is the error term.

In equation (D2), the error term from equation (D1) is regressed on a vector of district indicators to recover district value added, ϕ_d, by subject *j*. We then combine the subject-specific estimates to summarize district value-added to both subjects using the weighting approach of Lefgren and Sims (2012). The Lefgren and Sims (2012) approach also inherently shrinks the value-added estimates toward the mean in a regression-based framework (similarly to Chetty, Friedman, and Rockoff, 2014).

A desirable feature of the two-step modeling structure described by equations (D1) and (D2) is that variation in achievement attributable to student and district characteristics is partialled out in the first equation. The resulting value-added estimates from the second equation are orthogonal to these characteristics by construction. This is useful when we correlate the value-added metrics to our measures of academic mobility at the district level, as it rules out some explanations for the relationships we find. Parsons, Koedel and Tan (2019) also show that estimates from a two-step model of this form are less biased than more common "one-step" models under student-teacher sorting conditions that have been shown to be the most prevalent in practice.

Data for students in grades 4-8 from the entire panel period in each state are used to estimate district value added. All students in the analysis cohorts are omitted from the models in order to remove any mechanical correlation between our academic-mobility and value-added metrics. That is, the value-added models are jackknifed around the focal cohorts we use to study academic mobility, but otherwise cover the timeframe of their enrollment.

Appendix E: Connecting our intragenerational academic mobility metrics to intergenerational economic mobility metrics from CHKS

In this appendix we briefly expand on the calculations we performed to assess the prospects for connecting our intragenerational measures of academic mobility (AM) to the intergenerational economic mobility (EM) metrics from CHKS. We focus specifically on the issue of statistical power, noting that other challenges remain even if an association could be established (namely the time-inconsistency between the measures and the lack of a research design for use in establishing a causal connection).

As noted in the main text, CHKS publish O25 metrics that capture intergenerational EM for commuting zones (CZs) across the United States. Specifically, these place-based metrics indicate the earning percentile of children whose parents were in the 25th percentile of the income distribution. Our seven sample states include at least some coverage of 188 CZs for which economic mobility metrics are available from CHKS. Here we focus on the 165 of these CZs for which at least 50 percent of the population resides in one of our sample states.[27]

Turning to our metrics, and using districts in the sample of 165 CZs, the text (Figure 9) shows that of the total variance in academic mobility across districts, the cross-CZ variance shares for our four focal outcomes—eighth grade test scores, high school test scores, on-time graduation, and late graduation—are 0.14, 0.16, 0.25 and 0.30, respectively. It is straightforward to calculate that a one-standard-deviation move in the CZ-level distribution of academic mobility for each outcome is equal to the district-level standard deviation multiplied by the square root of the cross-district variance share. For our four focal outcomes, this calculation indicates effect sizes corresponding to moves in the CZ-level distributions that are only 37-55 percent as large as analogous moves in the district distributions. Thus, approximate "effect sizes" of a one-standard deviation improvement in the CZ-level distribution imply test percentiles that are 1.8 and 2.0 percentile points higher, respectively, for the eighth grade and high school tests; and graduation rates that 2.7 and 2.6 percentage points higher, respectively, for on-time and late graduation.

[27] The CHKS metrics do not provide any way of splitting CZs by state boundaries. The use of any CZs that are not fully contained by our state boundaries will introduce measurement discrepancies between our metrics and theirs. We chose the 50-percent cutoff to trade off comparability in measurement and the sample size.

In order to link variation in our AM metrics to variation in CHKS' EM metrics, we must convert changes in education outcomes to changes in earnings. Many studies map improvements in test scores into earnings later in life. These studies typically report values based on test standard deviations, not percentiles, to benchmark values, so we must convert our percentile-based AM numbers to standard deviations to use them. If we assume test scores are normally distributed, a two-percentile-point gain on the high school test—i.e., roughly one cross-CZ standard deviation of academic mobility—assessed at the 25th percentile of the distribution maps to a 0.06 standard deviation increase in test scores. The extant literature indicates that higher (late-grade) test scores of this magnitude correspond to higher earnings on the order of about 0.7-0.8 percent.[28]

Similarly, the literature suggests that graduation rate gains of 2.6-2.7 percentage points—again, about one cross-CZ standard deviation of academic mobility—would be expected to correspond to higher earnings of 0.0-0.2 percent.[29]

These earnings gains can be further converted into *mobility in the earnings distribution*, which is the metric used by CHKS. The standard deviation of O25 across the commuting zones in our sample states is just over 3 percentile points, which supplementary data files from CHKS show corresponds to a change in income of about 9.5 percent (taken at the mean O25 value of the 44th percentile of the income distribution). This implies that a one-standard-deviation move in the academic mobility distribution across CZs, converted to income gains, would map to a move in the economic mobility distribution estimated by CHKS of about 0.07-0.08 standard deviations for test scores, and 0.00-0.02 standard deviations for graduation outcomes. These are small numbers given the modest features of the CZ-level dataset (most notably, its small size and the fact that there is imperfect spatial overlap between the AM and EM metrics for CZs that cross state lines). Simple *ad hoc* tests confirm they are well below the thresholds at which statistical relationships could be detected using CZ-level data.[30]

---

[28] These back-of-the-envelope calculations are based on the correspondence between later-grade test scores and earnings reported in Lazear (2003), Mulligan (1999), and Murnane et al. (2011). These studies report that a one-standard-deviation increase in later-grade test scores corresponds to higher earnings on the order of 11-14 percent.
[29] These back-of-the-envelope calculations are based on estimates of the earnings returns to high-school graduation reported in Castex and Kogan Dechter (2014), Clark and Martorell (2014), and Ferrer and Riddell (2008). These studies estimate that the earnings-returns to obtaining a high school diploma ranges between 0 and 8 percent.
[30] Based on *ad hoc* power calculations, regressions of the EM metrics on the AM metrics using the CZ-level data will generate standard errors on the parameters linking the metrics of approximately 0.10 and 0.07 for the test-score and graduation-based AM metrics, respectively. Thus, the 95 percent confidence intervals are sufficiently large that

Again, we emphasize that these results should not be taken to imply that schools do not matter—indeed, across districts we identify substantial variation in academic mobility, and there is likely even greater variation at the school level—however, at the commuting-zone level, which is the level at which economic mobility is measured by CHKS, we conclude that there is not sufficient variation in academic mobility for it to be a major explanatory factor.

the effect sizes we would be looking for in these regressions—even assuming we could resolve other inference problems—are far from the thresholds for statistically detectable values.