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

Where Are Initially Low-performing Students the Most Likely to Succeed? A Multi-state Analysis of Academic Mobility (Preliminary Draft)

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 • February 2020

Where Are Initially Low-performing Students the Most Likely to Succeed? A Multi-state Analysis of Academic Mobility (Preliminary Draft)

Wes Austin
Georgia State University

David Figlio
Northwestern University/CALDER

Dan Goldhaber
American Institutes for Research/CALDER
University of Washington

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

Tara Killbride
Michigan State University

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/CALDER

Contents

Acknowledgments

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (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 and we make the following acknowledgments and disclaimers regarding the provision of data:

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
1000 Thomas Jefferson Street NW, Washington, DC 20007
202-403-5796 • www.caldercenter.org

Where are Initially Low-performing Students the Most Likely to Succeed? A Multi-state Analysis of Academic Mobility (Preliminary Draft)

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
February 2020

Abstract

We use administrative microdata from six states to study academic mobility in K-12 education, by which we mean the extent to which students' ranks in the distribution of academic performance change during their schooling careers. We find that there is substantial heterogeneity across districts in the academic mobility of students. The heterogeneity across districts is largest in terms of absolute mobility—i.e., in some districts, students throughout the distribution, including initially low performers, gain on other students in the statewide distribution, and vice-versa—whereas there is less cross-district variation in relative mobility. The most prominent correlates of high-mobility districts include value-added to achievement and the socioeconomic status of the student population.

1. Introduction

A number of studies examine the degree to which early cognitive and non-cognitive characteristics, including measures of achievement and engagement, predict long-term outcomes such as later test achievement, high school course-taking, high school graduation, college-going and labor market earnings.[1] Evidence shows that the large gaps in test achievement between advantaged and disadvantaged students, evident at the elementary level, continue to persist throughout students' K-12 educational careers (Betts, Zau and Rice, 2003; Clotfelter, Ladd and Vigdor, 2009; Reardon, 2011; Goldhaber et al., 2018). However, research also shows that school and teacher quality matter for educational achievement (Rivkin, Hanushek, and Kain, 2005; Chetty, Friedman, and Rockoff 2014; Jackson, Johnson, and Persico, 2016; Jackson 2018; Brunner, Hyman, and Ju, forthcoming), which implies that differences in school quality can impact (positively or negatively) the degree of persistence of students' early-grade outcomes.

Yet while there are numerous studies that examine the effects of specific educational interventions, the literature on the scope for schools and districts to influence student academic mobility broadly within the performance distribution is surprisingly thin. Jang and Reardon (2019) document the extent to which test achievement for cohorts of students changes as they progress from grades 3 to 8 in states and school districts. They show that in most states higher-SES students, who already outperform lower-SES students by the third grade on state tests, continue to gain on lower-SES students through the eighth grade.[2] Jang and Reardon (2019) also show that there is considerable heterogeneity across states in the rate of at which achievement diverges by SES.

[1]See, for instance, Murnane, Willett and Levy (1995), and Cawley, Heckman, and Vytlacil (2001), Cunha et al. (2006) Todd and Wolpin (2007).
[2] They use student achievement data from the Stanford Education Data Archive (SEDA), which contains about 11,000 school districts.

To date, however, little is known about how much variation exists across individual districts in the expansion or reduction of achievement gaps and what district characteristics might contribute to or predict district differences in this regard. Using state-level administrative microdata from six states, and couched within an "academic mobility" framework, we explore district-level heterogeneity in the extent to which students' academic performance in the third grade maps to their own longer-term educational outcomes through high school.

Our analytic framework is based on tools developed in recent work on intergenerational economic mobility by Chetty, Hendren, Kline, and Saez (CHKS, 2014) and Chetty, Hendren, Jones, and Porter (CHJP, 2018).[3] CHKS and CHJP document children's and their parents' economic mobility over time, assessing whether families are economically mobile and how this varies by geographic place in the United States. In our application, we use detailed student-level data to construct measures of *intragenerational academic mobility* for recent cohorts of students. We focus on how achievement in the third grade maps to the following longer-term educational outcomes: eighth grade test performance, performance on high school end-of-course exams and the ACT, and high school graduation, on-time and within one year of on-time. Following CHKS and CHJP, our test-based measures of academic mobility are based on the rank-rank relationships between performance percentiles in third grade and performance percentiles on the later tests. These measures describe the extent to which students' late-grade placements in the outcome distribution are pre-determined based on early-grade performance, and allow us to assess how distributional "stickiness" varies across schooling systems. For graduation outcomes, our mobility metrics measure graduation rates for students who start at different points in the early-grade achievement distribution.

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

Like in CHKS and CHJP, our access to student-level data facilitates the division of total mobility into its two subcomponents: absolute mobility and relative mobility. Our measures of relative mobility answer the following question for each district "Do initially low-achieving students gain in the district performance distribution compared to initially high-achieving students during the K-12 career?" Alternatively, our measures of absolute mobility answer the question "Do initially low-achieving students in the district gain in the statewide performance distribution during the K-12 career?"[4] Disentangling these two sources of mobility is an important first step toward understanding what drives district-level variation in academic mobility.

In what follows, we first summarize academic mobility across outcomes 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 show that there is statistically and economically meaningful heterogeneity in academic mobility across school districts within states. The heterogeneity is driven primarily by differences across districts in absolute mobility, and less so by differences in relative mobility. We also identify the characteristics of districts that predict greater academic mobility. The two most prominent predictors are districts' value-added to achievement (estimated out of sample) and the socioeconomic status of students, primarily as measured by free and reduced-price meal eligibility. In future extensions we will link our place-based intragenerational academic mobility metrics to location-matched intergenerational economic mobility metrics published by CHKS, which will allow us to empirically explore the connections between these two types of mobility.

---

[4] 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. (2018), Elsner and Insphording (2017a, 2017b), Elsner et al. (2019), and Murphy and Weinhardt (2018).

2. Data and Measurement of Academic Mobility

2.1 Data

We use state administrative microdata from public schools in six states—Massachusetts, Michigan, Missouri, Oregon, Texas, and Washington. We assemble cohorts of all students who have third grade standardized test scores (the initial statewide testing grade in K-12 public schools) and follow them throughout their K-12 schooling careers. 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 the entire U.S. (for comparison). We track academic mobility for two- to four-year cohorts of students who were in the third grade between the 2005-06 and 2008-09 school years (hereafter, including in Table 1, we identify school years by the spring year, e.g., 2006 for "2005-06"). The earliest cohort year is 2006 because this is the first year of consistent testing in grades 3-8 in most states, and the latest cohort-year is 2009 because this is the oldest cohort for whom we can track lagged graduation outcomes using our data (which go through 2019, one year after the on-time graduation year for the 2009 cohort).

In total our analysis involves more than 2.5 million students (Table 1). The states in the sample exhibit substantial heterogeneity in terms of their populations. For example, the percent of black and Hispanic students across states ranges from 3.0-19.0 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.[5] Finally, the structure of the education system differs significantly across the states in terms

[5] 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.

of the distribution 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 six states we examine are diverse along many dimensions and provide substantively different evaluation contexts.

Under the No Child Left Behind Act (NCLB) and the 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. Thus, while each state administers a unique test, our analysis of academic mobility between grades 3 and 8 is fairly uniform across states.[6] At the high school level, however, the flexibility of federal testing requirements means that the grades in which a test outcome is observed varies somewhat across states. To assess academic mobility based on high-school achievement, in each state we identify an exam with near-universal coverage administered in a common grade (see Table 2).[7] 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 school is more rigidly structured than in other subjects (e.g., in math and science, students are much more likely to take the same classes in different grades). Table 2 shows that the focal high school tests are administered in grades 10 or 11, have very high coverage rates, and are overwhelmingly taken in the common grade.[8] We also assess the likelihood of high school graduation and consider both on-time and delayed graduation. We define the latter as graduating within one year of on-time.

[6] 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.

[7] 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).

[8] Note that in Washington a test change during the data panel changed the test and focal grade, as noted in the table.

2.2 Measuring Academic Mobility

2.2.1 Rationale and Basic Framework

As noted above, our methodological approach follows on the framework developed by CHKS and CHJP to study intergenerational economic mobility. The mobility metrics we construct are based on percentile rankings in the performance distribution for various outcomes at different points in the schooling career. In their similar percentiles-to-percentiles analysis of place-based intergenerational economic mobility, CHKS note that "absolute mobility may be of greater normative interest than relative mobility" (p. 1562) because the former must reflect an unambiguous improvement. This logic applies to our measures as well. For instance, relative mobility could be high in a district with underwhelming performance among initial high-performers, even in the absence of exceptional performance of initial low-performers. Many researchers and education systems use the district achievement gap as a measure of performance, and comparisons between relative and absolute mobility highlight potential limitations of this measure. Specifically, a rightward shift in achievement across the 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 potentially reduce the achievement gap even though initially low-achievers actually fare worse in comparison to peers in other districts.

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, we follow their lead in the use of parsimonious summary measures of mobility. A key insight from CHKS permitting parsimonious presentation is that the rank-rank relationship between intergenerational economic outcomes is functionally linear. This allows them to summarize the relationship in each state with just the slope—the measure of relative mobility—and intercept—the measure of absolute mobility—parameters from a linear regression. This is also true in our application—i.e., the rank-rank relationships between early- and late-career student outcomes at various points during K-12 education are linear. Documentation of the linearity of the rank-rank relationships in our data is provided in Appendix Figure A1.[9]

Equation (1) summarizes a student's late-career rank given the early-career (grade-3) rank with just the slope (β) and intercept (α) parameters from the following regression:

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

In equation (1), O_i is a late-career academic outcome for student *i* and R_i is student *i*'s initial academic rank (assessed in the third grade).

We define O_i in four ways: test score rank as measured by combined performance on math and English Language Arts (ELA) grade-8 statewide assessments (i.e., average of the grade-8 ranks), test score rank on the HS test (using the tests indicated in Table 2), on-time high school graduation, and high school graduation within one-year of on-time. For the graduation outcomes, O_i is not a rank, but simply an indicator for whether graduation occurred.

In our primary specifications we set the initial rank, R_i, as the average rank on third grade math and ELA statewide tests. In extensions omitted for brevity we also construct subject-specific versions of R_i based on third grade math and ELA ranks that we link to subject-specific math and

[9] Results are reported in the appendix for three states; results for the other states are pending. Linearity is strongly upheld in the achievement-based rank-rank relationships in all three states. For the relationships involving high school graduation outcomes, which are binary, 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. At lower percentile values the relationship is not linear, which is attributable to strong floor effects in graduation combined with the binary outcome.

ELA ranks in grade-8 (no unique insights come from the subject-specific models and we suppress the results, they are available upon request).

Figure 1 illustrates two hypothetical, extreme mobility scenarios within our percentiles-to-percentiles framework, which applies for the mobility assessments based on the eighth-grade and high-school tests. The first graph in the figure shows a case where $\alpha = 0$ and $\beta = 1$. This is a situation 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 in the state achievement distribution. This would mean that students who enter in the first percentile achieve outcomes that are, on average, equivalent to those of students who enter the panel in the 99th 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 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, which is given by $\alpha = 50 - 50\beta$.

When we disaggregate the data below the state level—e.g., 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 a subpopulation. To illustrate this point, consider the following modified versions of equation (1) that permit subgroup-level analyses:

$$O_{is} = \alpha_s + \beta_s R_{is} + \epsilon_{is} \quad (2)$$

$$O_{id} = \alpha_d + \beta_d R_{id} + \epsilon_{id} \quad (3)$$

In equation (2), the subscript *s* indicates the group membership of student *i*—e.g., within a state we define groups *s* by race/ethnicity, FRL eligibility, and the urbanicity (urban, suburban or rural) of the school attended in the third grade. In equation (3), the subscript *d* identifies students who attend school district *d*. The dependent and independent variables in equations (2) and (3) continue to be defined by the full statewide distributions (i.e., the ranks are not type-specific). α_s and α_d capture absolute mobility in the statewide distribution at the bottom of the panel entry rankings for students in group *s* and district *d*, respectively. Similarly, β_s and β_d capture differences in relative mobility across groups of students defined by their subgroup membership and districts.

Total mobility for a subgroup of students in equation (2) or (3) is defined by the values of both mobility parameters, as illustrated by the hypothetical scenarios in Figure 2. In each scenario, the slopes of the two lines are held constant (i.e., neither β_{solid} nor β_{dashed} change across panels), with the solid line exhibiting more relative mobility, shown by the slope of the solid line as less steep than the dashed ($\beta_{solid} < \beta_{dashed}$). Hence, two students from the solid group who enter the panel with a given performance gap will have a smaller later outcome gap, on average, than two students from the dashed group 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 than for those represented by the dashed line. Moreover, in each scenario absolute mobility, as indicated by the intercept α, is larger for the solid group ($\alpha_{solid} > \alpha_{dashed}$)—i.e., the lowest performing students from the solid group perform better on the outcome measure than the lowest performing students from the dashed group.

The size of the gap in absolute mobility in Figure 2 results in very different outcomes for the two groups. In the first panel, the gap in absolute mobility is large, and as a result, students in the solid group have greater upward mobility throughout most of the distribution. This is apparent in the figure by their higher outcome ranks on the vertical axis at most entry-rank values. When the gap in absolute mobility is moderate in the second panel, the dashed group overtakes the solid group more quickly. The third panel shows that if the gap in absolute mobility is small enough, the greater relative mobility of the solid group leads to a situation where students from that group underperform the dashed group with respect to the outcome measures at most entry percentiles. This latter situation is an example of why CHKS argue that absolute mobility is a more useful measure than relative mobility conceptually—the higher relative mobility of the solid line simply reflects the underperformance of initial high achievers in the hypothetical district.

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

$$\bar{O}_{ps} = \alpha_s + \beta_s p \tag{4}$$

Equation (4) gives the average outcome rank for students in group *s* for any given starting percentile *p*. Similarly, $\bar{O}_{pd}$ gives the district-level analog. Following CHKS, we focus on the mobility of students at the 25th percentile of the initial performance distribution to produce summary measures of academic mobility for initially low-achieving students, denoted by $\bar{O}_{25}$. From equation (4), $\bar{O}_{25}$ for student group *s* can be estimated as $\hat{\alpha}_s + \hat{\beta}_s * 25$.

The interpretation of the α 's and β 's as described thus far applies to their estimation on outcomes that are percentile-ranked. The graduation outcomes are not ranked. Rather, they 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. For example, $\bar{O}_{ps}$ in equation (4) for on-time graduation indicates the likelihood of graduation for a student in the 25th percentile of the third grade performance distribution who belongs to group *s*. This likelihood can be compared to the likelihood of graduation for a student in the 25th percentile in group *r*, $r \neq s$, to compare the mobility across groups as measured by graduation. This parallels the approach of CHJP for the binary variables they consider in their analysis of intergenerational economic mobility.

2.2.2 Complicating Issues: Measurement Error and Sample Attrition

In this section we discuss two complications associated with the measurement of academic mobility. First, the initial third grade percentile rankings are measured with error as they are based on noisy test scores (Boyd et al., 2013; Lockwood and McCaffrey, 2013). This will attenuate our estimates of β and correspondingly inflate our estimates of α .

We reduce the influence of test measurement error by using the average of the ranks in math and ELA in third grade to set the initial percentiles, rather than using ranks based on a single test. However, this does not fully address the problem. Thus, in addition, we leverage data on measurement error in the testing instruments to further adjust our estimates. Specifically, the state tests we use to set the initial ranks include publisher-provided Standard Errors of Measurement (SEMs), which give estimates of the measurement-error variance in student scores. We use the SEMs to construct reliability ratios for students' initial percentiles following Wang and Stanley (1970), who give formulas for reliability ratios based on multiple assessments (in our case, the third grade math and ELA tests).[10] We then estimate standard errors-in-variables regressions to correct the β coefficients for attenuation bias.[11]

The measurement-error correction work has only been completed for Missouri thus far. Results from the other states are pending. For this reason, we show results throughout the main text based on analyses of initial percentile ranks that are not corrected for measurement error. Noting this limitation, the Missouri results show that while the magnitudes of α and β are affected by attenuation bias owing to measurement error, none of our findings with respect to heterogeneity in α and β are qualitatively affected by the measurement-error adjustment. We elaborate on these results below and show the estimates from Missouri in the appendix.

A second complication is that we can only observe students who remain enrolled in their initial state's public schools because we only have access to the state administrative data panels. This presents a challenge because it means we cannot track longer-term educational outcomes for the full cohorts of third grade students in our initial sample—we can only track students who remain in public schools in the home state long enough to be assessed. We impose this restriction conditional on the outcome assessed—e.g., for the grade-8 test outcome we track academic mobility

---

[10] 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: $r_c = \frac{0.25r_m + 0.25r_e + 0.50\theta_{m,e}}{0.50 + 0.50\theta_{m,e}}$. This formula is specific to calculating the reliability ratio for a combined assessment based on two tests with equal weight, which is the application here; the general formula from which it derives is available from Wang and Stanley (1970). Note that the SEMs of the tests taken by different cohorts in the sample in each subject vary slightly from year-to-year—in our calculations, we use the average SEM across all cohorts in each subject for the reliability ratios.

[11] 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 annual earnings in their investigations of economic mobility. This approach is feasible for us in that we can use tests from later grades, but this comes with additional complications—most notably, as we use tests from later grades our 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 broadly similar to what we find using the SEM-based approach, acknowledging that there is some ambiguity owing to the aforementioned compression of time between ranking periods.

for students who remain in the state through grade-8, for the high school test we track academic mobility for students who remain through the focal testing grade, and so on.

Our focus on students who remain in the data long enough to be assessed necessitates two caveats. One caveat is that this implies a modest concession in terms of the generalizabilty of our findings. We view the concession as "modest" because most students who are in the third grade samples remain within the states throughout the rest of their K-12 schooling careers. For instance, as shown in Table 3, we observe high school graduation (the most distant outcome from third grade) for 80-88 percent of our initial entering cohorts across states.

A potentially more substantive caveat is that our inability to observe state exiters raises concerns about bias from changes to the sample composition over time in each state. The validity issue stems from the fact that mobile students—whether across schools, districts, or states, are negatively selected on average (e.g. Grigg, 2012; Mehana and Reynolds, 2004). The negative selection is apparent in comparisons between state stayers and state movers in the data from each of the six states in Table 3—i.e., the average entry percentiles of state exiters are consistently below 50, as are the average outcome percentiles of students who were not in the focal state in the third grade but appear later (shown for the test outcomes only in Table 3).[12]

There are two types of bias concerns due to the negative selection of state leavers. The first is reference bias. Given that state leavers are negatively selected, on average, if their departure from the distribution is unaddressed the reference bias will lead to understated academic mobility among state stayers.[13] We address this concern by using in-migrators to counterbalance out-migrators when

[12]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.

[13] This is easiest to illustrate if one considers the extreme scenario in which there is no academic mobility in the full population (i.e., the rank order does not change), but the bottom 10 percent of students in the entry cohort disappear by

we set the rankings for our focal third grade cohorts of state stayers at different points in the schooling career. That is, we set the initial ranks based on all third grade students, then for a future outcome in grade g (where $g > 3$), we calculate the percentile ranks for state stayers within the full distribution of all observed students in grade g, inclusive of in-migrators. If in-migrators and out-migrators are similarly selected, the rank mobility of state stayers over time can be reliably measured with this substitution.

Table 3 shows that in several states—Missouri, Oregon, and Washington—the in-migrator and out-migrator ranks are similar for the grade-8 and high school tests, although there are small differences. In Massachusetts, Michigan and Texas, in-migrators and out-migrators are both negatively selected, but the differences between them are more pronounced.[14] This suggests that backfilling the distribution with in-migrators to replace the lost out-migrators is a reasonable, though imperfect, solution to address the reference bias problem. That said, there is no indication that differential selection between state entrants and exiters drives our findings substantively. For example, our results below are qualitatively similar in states with different selection directionally (e.g., Massachusetts and Michigan); moreover, the results are qualitatively similar in states with more and less differential selection (see Appendix C). Explanations for the insensitivity of our findings across states to differential in-migrator and out-migrator selection conditions include (a) that differences in selection are fairly modest in all states, even in the states with the largest differences, and (b) the majority of the sample in each state remains in-state over the course of the

the outcome year and no adjustment is made. In this case, the mobility curve (e.g., as in Figure 2) would be truncated such that it would cross the horizontal axis at the 10th percentile but would still end at coordinates (100,100). It would always be at or below the full population equivalent curve. Less extreme selection will generate a less extreme but directionally similar response.

[14] Note that with an underlying continuous distribution of scores, the mean of each rank distribution should be 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.

full period during grades 3-12. Finally, note that reference bias is not a concern for our analysis of graduation outcomes because we assess academic mobility in terms of graduation directly rather than a rank-based measure.

The second concern is of the more typical variety. Namely, because we only directly analyze the ranks of state stayers, 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 issue is not likely to be very important at the state level because it reduces to the generizability caveat above—i.e., our results are only valid for individuals who remain in their states. However, for our subgroup analysis is it potentially more problematic. For example, if state exiters are negatively selected conditional on their initial ranks, and district *A* has a higher proportion of exiters than district *B*, the differential attrition between districts will cause a compositional bias in the comparison.

For the time-being we leave this second type of potential biasing concern unresolved. In future iterations of this work we will implement an imputation procedure that will permit us to retain the entire initial third grade cohort for analysis of academic mobility. We will impute missing longer-term outcomes for students who exit the state using available data prior to the point of exit. The imputation procedure will also be designed to allow us to parameterize additional unobserved selection of out-of-state movers and examine the sensitivity of our findings to different parameterizations.

3. Findings

We conduct the analyses described above for each state separately. The state-by-state results show great consistency and accordingly, we consolidate the findings in the main text by reporting simple-average values across states.[15] State-by-state breakouts of the summary findings are available in Appendix C, and we note in text the handful of instances where there are notable differences across states.

3.1 Broad Patterns of Academic Mobility at the State Level

Figure 3 reports averages of the state-level estimates of β and $\overline{O}_{25}$ 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 8th grade and high school test rankings. The average estimates of β for the eighth grade and high school tests in Figure 3 are 0.75 and 0.73, even before correcting for attenuation due to measurement error. Using a projection based on how the Missouri results are affected by the measurement error correction (see Appendix Figure A2), we estimate that the average corrected $\beta's$ for the six-state sample will be around 0.80-0.82 for these outcomes. Put more plainly, where students start in the distribution, when tested in the third grade, is highly predictive of where they end up in the distribution in eighth grade and high school.

The estimates of β for the graduation outcomes are much lower—0.32 for on-time graduation and 0.23 for lagged graduation—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 shown in Appendix Figure A1 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 (most students graduate), students' early-career performance ranks are not strong predictors of success.

[15] We report simple averages rather than weighted averages in order to preserve as much of the state-level sample heterogeneity as possible.

The $\overline{O}_{25}$ values for the test-score outcomes are similar to each other, at 31.7 and 32.9 for the eighth-grade and high-school tests, respectively. The high school $\overline{O}_{25}$ values are expected to be higher if academic mobility is a continuous process during K-12 schooling. These numbers indicate that the average 25th percentile entrant in our data scored at the 31.7th percentile of the combined eighth-grade math and ELA tests, and at the 32.9th percentile of the high school test. The graduation-based $\overline{O}_{25}$ values, which capture on-time and delayed graduation likelihoods for the average 25th percentile student, are 78.4 and 83.8.

Returning to the measurement-error issue, and folllowing on the discussion above, the measurement error correction increases the value of β and reduces the value of α (as can be seen for Missouri in Appendix Figure A2). Because the directions of the attenaution bias in α and β are opposite, the effect on $\overline{O}_{25}$, which combines information from both parameters, is muted. The effect should be essentially offsetting at the 50th percentile of the distribution because this is where the magnitude of the bias in β will equal the magnitude of the bias in α in terms of setting the ranks. At values below the 50th percentile, such as at $\overline{O}_{25}$, the upward bias in α will exceed the bias toward zero in β, leading to overstated $\overline{O}_{25}$ values in the absence of the measurement-error correction. Another way to think about the uncorrected $\overline{O}_{25}$ values is that they are biased upward because of mean reversion due to test measurement error. When we correct for the test measurement error to account for the mean reversion, the estimates of $\overline{O}_{25}$ should decline.

We again project the likely effects of the measurement error correction in all six of our states based on the correction results that are currently available from Missouri. The projections suggest that the $\overline{O}_{25}$ values in Figure 3 for test ranks in eighth grade and high school will fall by about 2 percentage points once we correct for measurement error in the initial ranks in all six states.

For graduation outcomes, we project that the $\overline{O}_{25}$ values will fall by about half of a percentage point.

Finally, Appendix C shows that the graduation-based mobility metrics exhibit more cross-state variability than the test-based mobility metrics. The distributions of test score ranks are forced into alignment across states by the percentile conversions. But the distribution of graduation outcomes are quite different across states, primarily because there are notable 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. Hence, states with higher graduation rates overall have higher graduation-based $\overline{O}_{25}$ values.

This creates is an important source of ambiguity in interpreting the mobility findings with respect to graduation outcomes across states. One interpretation of a high $\overline{O}_{25}$ value is that it reflects a state's success in pushing initially low-achieving students through high school. But an alternative interpretation is that high graduation rates 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.

3.2 Academic Mobility for Sub-Groups Within States

In Figures 4, 5 and 6 we report on variants of Equation (2) where we define groups *s* by students' third grade racial/ethnic designations, FRL designations, and the urbanicity of the third grade school (urban, suburban, rural). Again, once we split the sample into subgroups within states, α_s and β_s are separately identified for each group and contribute unique information. $\overline{O}_{25s}$ serves as a summary measure of academic mobility. We continue to report simple-average values across the sample states in the figures. The charts in row 1 of each figure show the α_s, β_s, and $\overline{O}_{25s}$

parameters for the test-score outcomes and the charts in row 2 show the same parameters for the graduation outcomes.

We begin with Figure 4, which shows results for the splits by race/ethnicity. We compare Asian, Black, Hispanic, and White students.[16] The gaps in $\overline{O}_{25s}$ in the charts in the third column of Figure 3 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. This result is shown on average across the sample states in Figure 4; in Appendix C we show that it holds with great consistency across states individually as well.

The academic mobility differences between the other three racial/ethnic groups are less stark but still clearly present—within these groups, White students are the most mobile, followed by Hispanic and then Black students. The Black-White and Hispanic-White $\overline{O}_{25s}$ gaps in terms of grade-8 achievement, for example, are 4.4 and 1.9 pecentage points, respectively. For on-time graduation these same gaps are 6.3 and 3.0 percentage points. The Black-White mobility gaps shown 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 (Todd & Wolpin, 2007). In contrast, Clotfelter, Ladd, and Vigdor (2009) find that the Hispanic-White achievement gap narrows in North Carolina in 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

[16] 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.

representative sample, and Todd and Wolpin (2007), who find that it remains flat or widens modestly.[17]

A broad takeaway from Figure 4 is that variation between race/ethnic groups in absolute mobility (α_s) drives the variation in total mobility ($\overline{O}_{25s}$), and there is substantially less variation in relative (within district) mobility (β_s). The top row in Figure 4 shows that for the 8th grade and high school test outcomes, there is little variation in relative mobility (β_s) across groups—all hover around 0.7 (grade 8) and slightly lower (high school test). Most of the variation is in absolute mobility (α_s). 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, again noting that $\overline{O}_{25s}$ is a summary measure of the information in α_s and β_s for initially low-performing students. The variance in relative mobility (β_s) across racial/ethnic groups is somewhat more pronounced when we examine graduation outcomes, but even then the predominant variation driving differential total mobility by race/ethnicity ($\overline{O}_{25s}$) is in absolute mobility (α_s). This strong correspondence between α_s and $\overline{O}_{25s}$ is visually reinforced in the state-by-state results in Appendix Figure C2.

Another way to illustrate the importance of absolute versus relative academic mobility in explaining total mobility by race/ethnicity is to decompose the total change in $\overline{O}_{25}$ between groups into the portions that reflect 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 grade-8 test, the $\overline{O}_{25}$ gap is about

[17] 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.

12.4 percentile points. Of this total gap, just over 10.5 points is accounted for by the gap in α between Asian and Black students, and only 1.9 points is accounted for by the gap in β (which is roughly 0.07 in the chart, multiplied by 25 to map to $\overline{O}_{25}$).[18] This comparison overstates the value of α_s by focusing at a point in the distribution below the 50th percentile; still, even at $\overline{O}_{50}$, α_s is the dominant explanatory factor over the total mobility gap. Broadly speaking, Our findings suggest that when enrollment in a district matters for mobility of initial low performers—that is, when these students appear to be less anchored to their third grade test achievement—it is more so because all boats are raised in the district rather than particular subgroups of students are impacted.

Next, Figure 5 shows analogous splits by third grade FRL status. Compared to FRL students, non-FRL students have much higher absolute mobility (α) and similar relative mobility (β) (their β estimates are slightly higher for the test-based metrics and somewhat lower for the graduation metrics). These combine to result in much higher $\overline{O}_{25}$ values for non-FRL students, especially in terms of graduation outcomes. For example, the on-time and lagged graduation gaps among students who differ by FRL status and start at the 25th percentile in grade 3 are 13.4 and 11.1 percentage points, respectively, as shown in the bottom-right chart in Figure 5.

The last subgroup comparison we make is by the urbanicity of the school attended in the third grade, shown in Figure 6. The categories are urban, suburban, and rural. Here there is much less heterogeneity across groups than in the preceding figures, although it is again true that what heterogeneity does exist is concentrated in the absolute-mobility metrics. The most notable variation in Figure 6 is in terms of graduation rates. This is observed most easily in the bottom-right chart. It shows that graduation rates for initially low-performing students who attend urban schools are

[18] Note that the numbers from this calculation are depicted in the charts, but sometimes difficult to infer precisely visually. The raw data underlying the charts are available in Appendix C.

significantly below those of their suburban and rural counterparts (which have similar graduation rates) in the sample states.

In Appendix Figures A3 and A4 we show measurement-error corrected analogs to Figure 4 and 6 for Missouri (the measurement-error corrected version of Figure 5 is forthcoming). The appendix figures confirm the comparative patterns presented in Figures 4 and 6 are substantively unaffected by the measurement-error correction in Missouri, and strongly suggest this result will generalize to Figure 5 and the analogous output from other states.

Appendix Figures C2, C3, and C4 provide the state-by-state results that underlie the consolidated results in Figures 4, 5, and 6. For the comparisons along all three dimensions, there is strong consistency across individual states in the results.

3.3 *District-Level Variation in Mobility*

Figure 7 documents within-state, cross-district heterogeneity in α and β as estimated by equation (4). Specifically, the figure reports the averages of the estimated standard deviations of α_d and β_d for each outcome in each state. These estimates capture the extent to which rates of absolute and relative academic mobility vary across school districts within our sample of states.

An issue with these calculations is that the raw variance of $\hat{\alpha}_d$ and $\hat{\beta}_d$ will overstate the true variance across districts in α_d and β_d because the raw variance includes 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 $\hat{\alpha}_d$ and $\hat{\beta}_d$ 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 illustrate, define $\sigma^2_{\hat{\alpha}}$ as the unadjusted variance of $\hat{\alpha}_d$ using the real data, 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:

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

A similar procedure is applied to obtain estimates of σ_β. The null distributions from the randomized inference procedure also allow for direct tests of statistical significance of the variability in α_d and β_d. We say that the variance of a given parameter across districts is statistically signficant in a given state if the variance estimate using the observed data falls outside of the 95-percent confidence interval of the null-distribution values.

Thus far we have only been able to perform the randomized-inference procedure in Missouri. The variances of α_d and β_d across districts for all outcomes are statistically signficant at the 5 percent level or better, although the p-values are lower for β_d. To produce Figure 7, and to ensure that we do not overstate the standard deviations of α_d and β_d in the sample states, we use the magnitude of the error-variance correction in Missouri, in percentage terms, to adjust the raw variance values for all states. This is an approximation that serves as a placeholder until the state-by-state randomized-inference results are available to make a more precise correction.

A second, smaller issue with the variance comparisons is that α_d and β_d are not in the same units, making an assessment of the magnitudes of the variances difficult in raw terms. For comparability purposes we report estimates based on 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.

At a high level, the results in Figure 7 show that on average across the sample states, the cross-district variance of α_d consistently exceeds the magnitude-aligned variance of β_d. Specifically, the average standard deviation of α_d is about 50-100 percent larger across outcomes. Appendix Figure C4 shows that this relationship holds fairly steadily on a state-by-state basis. The randomized inference procedure also reveals that the estimation-error share of the raw variance in α_d and β_d is much larger when these metrics apply to graduation outcomes as opposed to test-based outcomes (results suppressed for brevity).

In terms of interpretation, the results indicate that attending a district that is one standard deviation higher in the distrubibution of absolute academic mobility (α_d) corresponds to an increase in student rank on the grade-8 test of 5.8 percentile points, on average. For the high-school test the same change corresponds to a 5.4 percentile point increase, and for on-time and delayed graduation rates the gains are 7.8 and 5.3 percentage points, respectively. The variances in relative mobility are more difficult to interpret in isolation. More broadly, a challenge in mapping the variances in α_d and β_d directly to student outcomes is that their combined movements determine total academic mobility, inclusive of co-movement. Future iterations of this work will replicate our procedure using the total mobility measure—i.e., $\bar{O}_{25}$—to gain clearer insight into the meaning of cross-district variability in total academic mobility.

3.4 *District-Level Correlates of Academic Mobility*

In order to explore linkages between academic mobility and district characteristics, we construct estimates of $\bar{O}_{25d}$ using the estimates of α_d and β_d for each district, then regress $\hat{\bar{O}}_{25d}$ on a host of district characteristics. The district characterisicts we consider include the percentage of students who are (a) Black, (b) Hispanic, (c) free/reduced-price lunch eligible (FRL), (d) participants in an individualized education plan (IEP), and (e) geographically mobile. Recall that we define a student as geographically mobile if she spends less than the full school year in the school where she is tested in that year (i.e., if she is a mid-year school mover). We also construct a district-level measure of school segregation using a Theil index, following CHKS.[19] 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. Value-added captures district contributions to student test score growth conditional on student characteristics. 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):

$$Y_{ijdkt} = \gamma_0 + \mathbf{Y_{i(t\text{-}1)}}\boldsymbol{\gamma_1} + \mathbf{X_{it}}\boldsymbol{\gamma_2} + \mathbf{S_{kt}}\boldsymbol{\gamma_3} + \mathbf{L_{dt}}\boldsymbol{\gamma_4} + \varepsilon_{ijdt} \quad (6)$$

$$\varepsilon_{ijdt} = \phi_d + \nu_{ijdt} \quad (7)$$

In equation (5), Y_{ijlkt} 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, $\mathbf{Y_{i(t\text{-}1)}}$ 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 free/reduced-price lunch status, gender, race, English as a second language (ELL) status, geographic mobility, and whether or not the student has

[19] 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.

an individualized education plan (IEP), $\mathbf{S}_{\mathbf{kt}}$ and $\mathbf{L}_{\mathbf{dt}}$ contain the variables included in $\mathbf{Y}_{\mathbf{i(t-1)}}$ and $\mathbf{X}_{\mathbf{it}}$ aggregated at the school and district levels, respectively, and ε_{ijlt} is the error term.

In equation (6), the error term from equation (5) is regressed on a vector of district indicators to recover district value-added estimates, 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, as in Chetty, Friedman, and Rockoff (2014). A desirable feature of the two-step modeling structure described by equations (5) and (6) 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 a 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.[20]

Data from the entire panel period for students in grades 4-8 in each state are used to estimate district value-added. However, 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 but otherwise cover the timeframe of their enrollment.

Figure 8 shows the predictors of district-level academic mobility, again summarized by $\bar{O}_{25d}$, averaged over the states in our sample for each focal outcome. Specifically, we report average

[20] Parsons et al. (2019) show that estimates from a two-step model of the form shown in equations (5) and (6) 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.

coefficients across the six 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.[21] The independent variables are standardized in each state to have a mean of zero and variance of one to facilitate comparability. 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.[22] The detailed state-by-state regression output summarized by Figure 8 is provided in figure and table form, including information about statistical significance, in Appendix C.

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 (URM, or Hispanic and Black students) shares and higher FRL shares correspond to lower academic mobility. More broadly, all indicators of student disadvantage in the univariate regressions—student shares by URM, FRL, IEP, and 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. Notably, the value-added associations are similar in magnitude regardless of whether mobility is measured by test scores or graduation.

[21] The multivariate regressions also include indicator variables for urbanicity for completeness (for which the coefficients are suppressed), but this detail is inconsequential to the results given the limited variability in academic mobility by district urbanicity documented in Figure 6.

[22] 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, 2014).

The multivariate regressions show that district value-added to student achievement and the share of students who are FRL are the most important factors for predicting academic mobility. Both of these factors are strongly related to academic mobility and the relationships are statistically significant individually in most states and for most outcomes in both the univariate and multivariate regressions (Appendix C). The FRL share is clearly the primary factor among the indicators of student disadvantage—in the multivariate regressions, the associations with the other disadvantage metrics attenuate substantially and even flip signs in some cases, whereas the negative associations between the FRL share and academic mobility become stronger. Unsurprisingly, the association between value-added and academic mobility is essentially unchanged in the multivariate regression. This follows from the construction of the value-added models, which partial out variation in outcomes due to student characteristics in the first-step equation.

Finally, Figure 9 shows the average correlations between district value-added (ϕ_d) and α_d and β_d to gain insight into what aspect of academic mobility drives the association with value-added. The correlations in the figures are partially corrected for estimation-error variance in the metrics via shrinkage in the value-added estimates. However, a correction has not yet been made to account for estimation error in α_d and β_d, which implies that the correlations as presented are attenuated somewhat relative to their true values.[23] Noting this caveat, the figure shows that the correlation between ϕ_d and α_d is consistently much stronger than the correlation between ϕ_d and β_d in our sample states. While the correlations between ϕ_d and α_d are uniformly positive, they are larger when α_d is estimated using a test-score outcome, which is intuitive given that our value

[23] In future iterations of this work we will apply empirical Bayes shrinkage to the mobility metrics using the formula described in Koedel, Mihaly, and Rockoff (2015) to adjust this correlation.

added measures are test-based. For the graduation-based mobility metrics, the correlations between ϕ_d and α_d are positive, but they are negative for β_d.

The results in Figure 9 show that higher academic mobility in high value-added districts reflects higher absolute academic mobility. Put another way, school districts that produce strong mobility outcomes for initially low-performing students via the link between $\bar{O}_{25d}$ and district value-added, as shown in Figure 8, do not seem to achieve this result by re-shuffling their students within their own performance distributions, but rather by improving outcomes for all of their students, including initially low performers. This result is consistent with high-value-added districts raising performance for all students. It does not support the hypothesis that districts promoting mobility among low performers do so by being particularly effective with low-performing relative to high-performing students (a scenario that would be captured by variability in β_d).[24]

4. Discussion and Conclusion

There is a robust literature showing that early academic performance measures are highly predictive of upstream educational success (e.g. Silver, Saunders, and Zarate, 2008; Easton, Johnson, and Sartian, 2017). Our work adds to this literature by examining long academic panels from six states to show that test scores as early as third grade are highly predictive of 8th and 10th grade test scores and of high school graduation. In particular, in regressions of students' percentile ranks on 8th grade and high school tests, we estimate that the coefficient on the third grade percentile rank is roughly 0.80 after correcting for test measurement error. Third grade performance is less predictive, but still highly predictive, of high school graduation, with analogous coefficients of between 0.25 and 0.35 when used to predict on-time and delayed graduation, also after correcting

[24] Also see Parsons (2016).

for measurement error. The weaker predictive power of the third grade rank over high school graduation is driven largely by the fact that most students graduate, which means that variability in much of the third grade performance distribution maps only weakly to variation in outcomes. Interpreting this is challenging: it could be interpreted as showing that school systems are doing a good job of helping most students graduate, or as showing that states have low standards for graduation. Our data are ill-suited to disentangle these competing mechanisms.

Our analysis of differences in academic mobility by student race and ethnicity, FRL status, and school urbanicity replicate largely familiar patterns in the literature—White students, and especially Asian students, have relatively high academic mobility whereas Hispanic students, and especially Black students, have lower academic mobility. A novel and troubling finding, given the policy emphasis on achievement of high-poverty students, is that FRL students have much lower academic mobility than non-FRL students. Differences in academic mobility by urbanicity are smaller than along the other two dimensions, although a notable result is that low-performing students who attend urban schools in third grade are much less likely to graduate from high school than their suburban or rural counterparts.

We also show that school districts exhibit statistically and economically significant variation in academic mobility. The predominant driver of cross-district variation in total academic mobility is *absolute* mobility, not *relative* (within district) mobility. That is, districts differ much more by whether they are effective in raising achievement throughout the entire distributions of their students than they do in their ability to improve lower-performing students' relative ranks internally. Indeed, we do not find evidence of large differences across districts in relative mobility, which suggests that districts do not, in fact, differentially specialize in educating students at different achievement levels within their distributions (e.g., high versus low achievers).

When we consider total mobility (absolute and relative combined), we find that districts where initially low-performing students rise the most in the statewide distribution are characterized predominantly by having high value-added to student achievement in grades 4-8, and low proportions of FRL-eligible students. Because of the way we construct the value-added models for our analysis, these two district characteristics are effectively orthogonal. If we consider high value-add as a proxy for quality, as it measures districts' unique contributions to student achievement gains conditional on the student population served, then together our results suggest that "better" districts are, perhaps unsurprisingly, better at helping initially low-performing students improve. However, initially low-performing students also improve faster in districts with lower proportions of low-income students (proxied by FRL eligibility), pointing to larger economic factors outside of the control of school systems that influence academic mobility.

The analytic framework we draw on from the economic mobility literature is useful for documenting academic mobility across states, across subpopulations of students within states, and across school districts. We view this work as a first step in a lager research agenda focused on academic mobility, and specifically how and why students experience different levels of mobility within and across education systems. But, we caution that our descrptive findings are not intended to show causal relationships and should not be interpreted as such.

There are several fruitful directions for future work building on our study. Descriptively there is still much to do. For example, the racial/ethnic and FRL gaps in academic mobility can be further decomposed into their within- and across-district components to improve our undersanding of how these gaps evolve and identify pressure points for future interventions to address them. Another area ripe for additional research involves pushing on the causality of district-level academic mobility. A natural analog to CHJP's work on geographic economic mobility, and recent related work on the geography of health (e.g., Deryugina and Molitor, 2019), would leverage district movers to assess the extent to which students' academic mobility trajectories change when they move. Finally, it is reasonable to hypothesize that intragenerational academic mobility is a factor that drives variation in intergenerational economic mobility as documented by CHKS. School systems in which initial low-achievers improve fastest in the academic performance distribution plausibly contribute positively to generation-by-generation improvements in economic well-being. In future work we will incorporate more states into our analysis and aggregate our district-level academic mobility metrics to the commuting zone level. These aggregate data, combined with commuting zone data on economic mobility published by CHKS, will facilitate exploration of the connection between place-based intergenerational economic mobility and intragenerational academic mobility.

<u>References</u>

Betts, J.R., Zau, A.C., & Rice, L.A. (2003). *Determinants of Student Achievement: New Evidence from San Diego*. San Francisco, CA: Public Policy Institute of California.

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.

Brunner, E., Hyman, J., & Ju, A. (forthcoming). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics*.

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

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., 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*.

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

Denning, J. T., Murphy, R., and Weinhardt, F. (2018). Class rank and long-run outcomes. IZA Discussion Paper.

Deryugina, T., & Molitor, D. (2019). Does when you die depend on where you live? Evidence from Hurricane Katrina. National Bureau of Economic Research Working Paper No. 24822.

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.

Easton, J.Q., Johnson, E., & Sartain, L. (2017). The Predictive Power of Ninth-Grade GPA. Research Report. Chicago, IL: University of Chicago Consortium on School Research.

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.

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., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal* 55(1), 171-201.

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.

Jackson, K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy* 126(5), 2072-2107.

Jackson, K., Johnson, R., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics* 131(1), 157-218.

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.

Koedel. C., Mihaly, K., & Rockoff, J.E. (2015). Value-added modeling: A review. *Economics of Education Review* 47: 180-195.

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.

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.

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

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. (2018). Top of the class: The importance of ordinal rank. National Bureau of Economic Research Working Paper No. 24958.

Parsons, E. (2016). Does attending a low-achieving school affect high-performing student outcomes? *Teachers College Record* 118(8), 1-36.

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.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica* 73(2), 417-58.

Silver, D., Saunders, M., & Zarate, E. (2008). What Factors Predict High School Graduation in the Los Angeles Unified School District? California Dropout Research Project Report #14. Santa Barbara, CA: UC Santa Barbara Gevirtz Graduate School of Education.

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.

Stark, P., Noel, A.M, & McFarland, J. (2015). Trends in High School Dropout and Completion Rates in the United States: 1972-2012. Compendium Report. Washington, DC: United States Department of Education.

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).

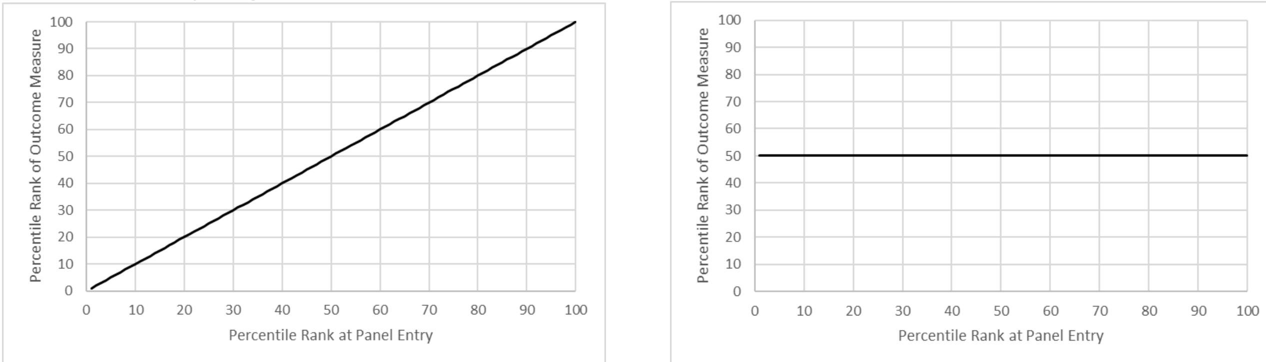


Figure 2. Comparison of two hypothetical student groups, one with higher relative mobility (solid lines) and one with lower relative mobility (dashed lines), with differing gaps in absolute mobility.

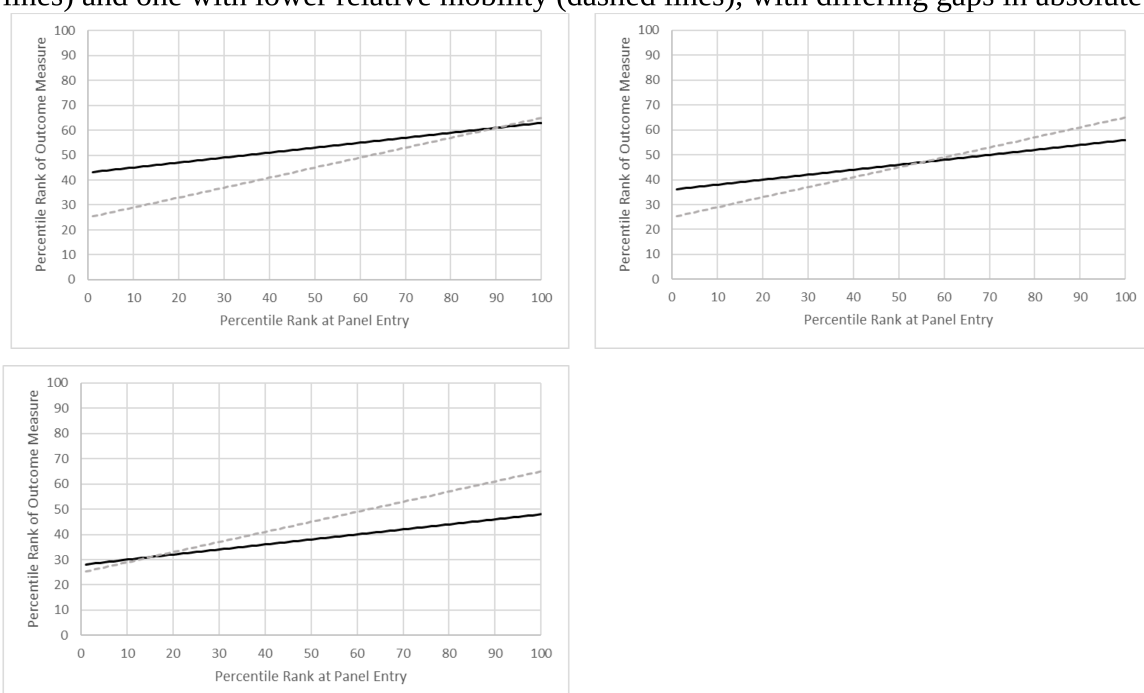
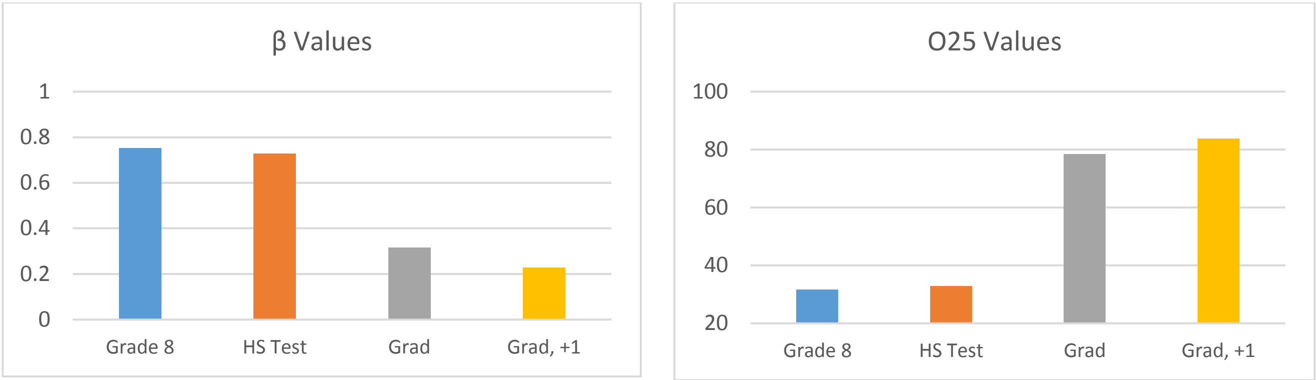


Figure 3. Simple averages of state-level estimates of β and $\overline{O}_{25}$ for each outcome, baseline estimation conditions.

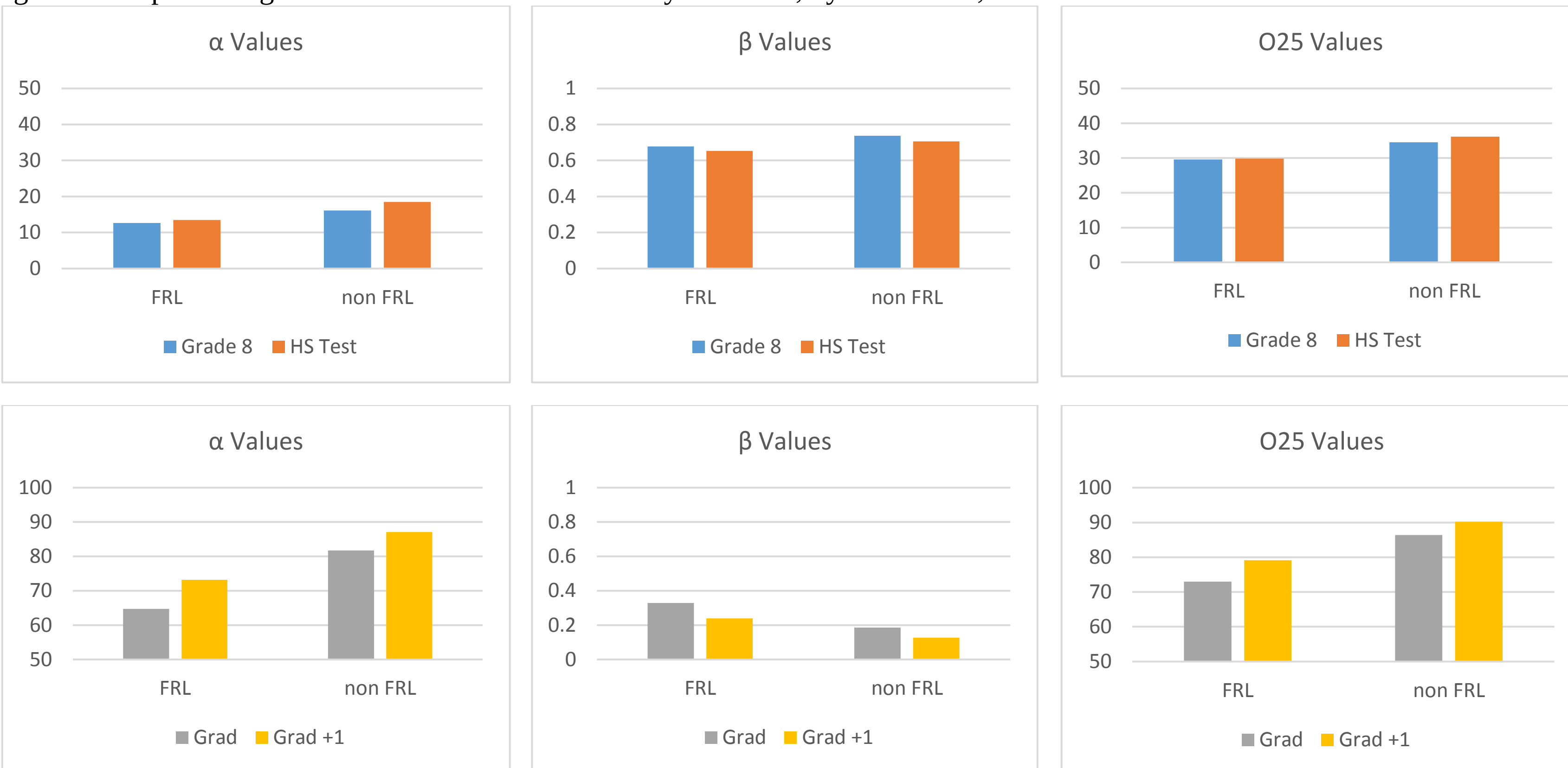


Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. α is redundant when all statewide data are used, as described in the text. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted 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. The estimates of β and $\overline{O}_{25}$ in the baseline conditions are without correcting for measurement error in the initial percentile ranks. We have yet to run the measurement-error correction for all states, but results for Missouri with the error correction are reported in Appendix Figure A2. The results in the appendix are consistent with the predicted effects of measurement error as described in the text. Full results broken out by each state individually are reported in Appendix C.

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

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. These estimates are from the baseline conditions without correcting for measurement error in the initial percentile ranks. We have yet to run the measurement-error correction for all states, but results for Missouri with the error correction are reported in Appendix Figure A3. The results in the appendix are consistent with the predicted effects of measurement error as described in the text. Full results broken out by each state individually are reported in Appendix C.

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

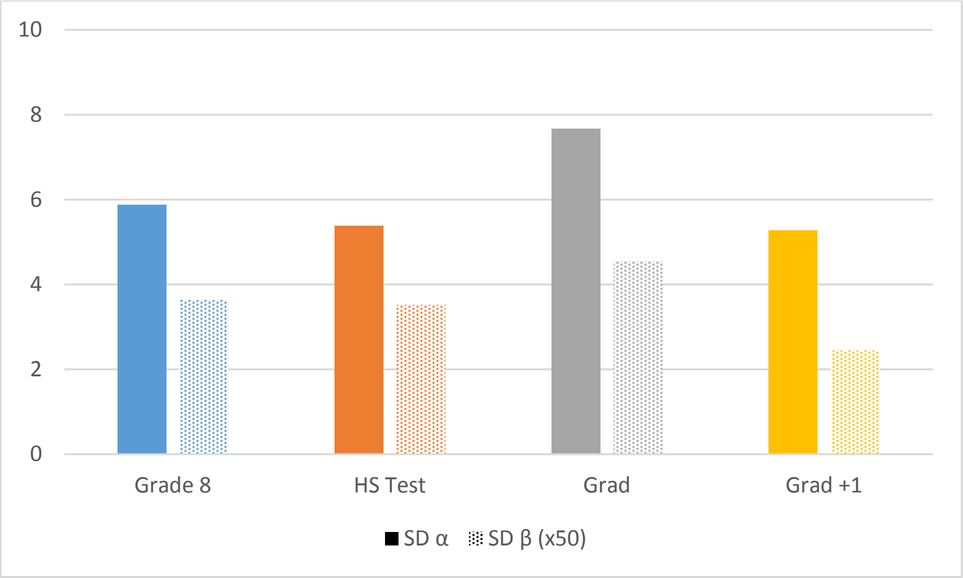


Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. These estimates are from the baseline conditions without correcting for measurement error in the initial percentile ranks. Measurement-error corrected results are pending. Full results broken out by each state individually are reported in Appendix C.

Figure 6. Simple averages of statewide academic mobility measures, by the urbanicity designation of the school, baseline conditions.

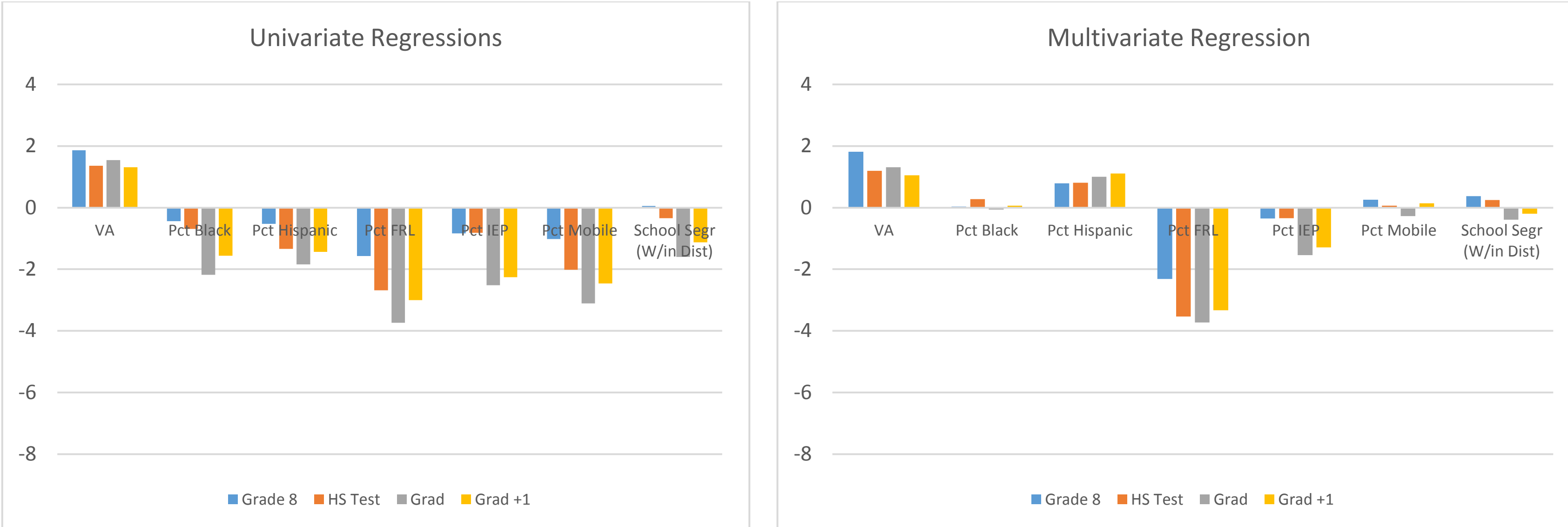
Notes: $\overline{O}_{25}$ for the graduation outcomes is the graduation rate at the 25th percentile of the entering-rank distribution. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. These estimates are from the baseline conditions without correcting for measurement error in the initial percentile ranks. We have yet to run the measurement-error correction for all states, but results for Missouri with the error correction are reported in Appendix Figure A4. The results in the appendix are consistent with the predicted effects of measurement error as described in the text. Full results broken out by each state individually are reported in Appendix C.

Figure 7. Simple averages of the within-state, cross-district standard deviations of α_d and β_d .



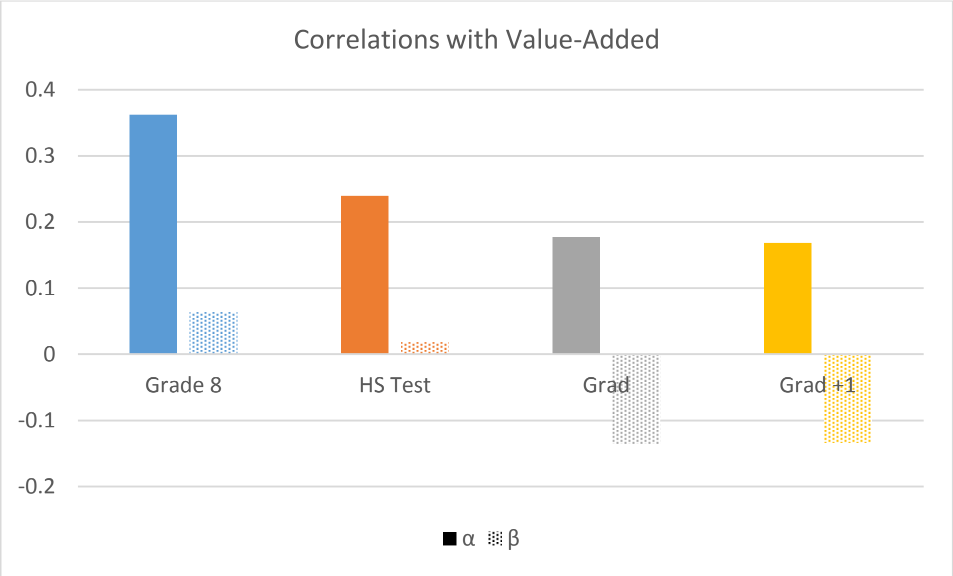
Notes: The standard deviations of β_d are multiplied by 50 to align the magnitudes of variances of α_d and β_d for comparative purposes in this figure. Oregon does not offer a high school test taken in a (near) universal grade, so Oregon is omitted from the HS Test results. In Michigan, data are unavailable for the 2009 cohort to assess the late-graduation outcome, so just three cohorts are used for that outcome. The standard deviations are corrected for estimation-error variance using the randomized-inference procedure described in the text. Current estimates are *projections* for the all-state sample based on the magnitude of the estimation-error correction in Missouri, which is the only state we have run the randomized inference procedure thus far. State-by-state results are omitted from Appendix C pending the application of the estimation-error correction procedure in each state.

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



Notes: The multivariate regressions also include urbanicity indicators but whether these are included does not influence the results. 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 C.

Figure 9. Simple averages of the correlations between district value-added, and α_d and β_d, within states.



Notes: 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. These correlations are not corrected for measurement error in the academic mobility metrics; the value-added metrics are corrected for measurement error via implicit shrinkage. Full results broken out by each state individually are reported in Appendix C.

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
Massachusetts	2007-2008	139,337	7.83	13.94	31.65	17.30	2.32*	20.11	68.19	304	1,116
Michigan	2006-2009	453,733	18.99	5.71	40.97	10.92	11.11	21.29	44.99	755	2,040
Missouri	2006-2009	264,612	18.17	4.00	46.34	15.16	6.62	18.79	30.87	548	1,200
Oregon	2006-2008	123,833	3.03	16.83	47.59	15.37	4.03	30.69	25.60	208	1,086
Texas	2006-2009	1,309,114	13.54	47.68	57.84	5.86	6.68	42.27	27.90	1,173	4,338
Washington	2006-2008	218,051	5.70	15.80	42.26	11.44	1.04*	26.12	45.30	296	1,254
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 grade-3. The spring year is used to indicate the academic year (e.g., 2009 = 2008-09 school year). Students who took both the Math and Reading grade-3 state tests are included in the core sample described here. 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 reported in the final column 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 2007-08 school year from the 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
Massachusetts	MCAS ELA	10	99.5	0.2
Michigan	ACT/SAT	11	99.3	0.7
Missouri	English II EOC	10	93.1	2.6
Oregon	--	--	--	--
Texas	Reading/English II EOC	10	94.13	5.66
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. State level attrition and in-migration.

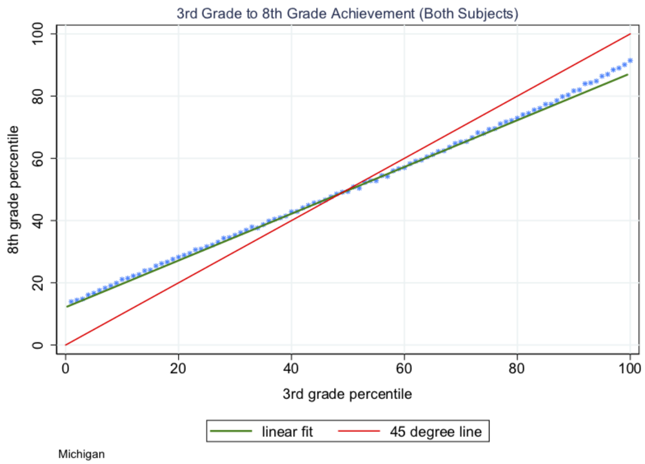
		Original Cohort Members						
		Panel Entry	Observed with Outcome		Out-Migrators: Assessed in Grade-3 But Exited Prior to Outcome Year		In-Migrators: Assessed with Outcome but Not Present in Grade-3	
		N	N	Avg. Outcome Pctl. or Grad Rate	N	Avg. Entry Pctl.	N	Avg. Outcome Pctl.
Grade 8 – Combined Math and ELA	Massachusetts	139,337	124,606	50.45	14,731	46.65	19,525	38.79
	Michigan	453,733	394,721	50.69	59,012	40.84	48,912	47.93
	Missouri	263,910	227,459	50.69	34,907	47.68	39,834	46.54
	Oregon	123,833	105,674	49.74	18,159	45.87	24,530	44.56
	Texas	1,280,996	1,094,987	48.73	186,009	49.29	169,263	44.94
	Washington	218,051	185,609	50.51	32,442	45.23	43,600	42.73
High School Exam	Massachusetts	139,337	114,374	50.53	24,963	46.28	28,794	36.54
	Michigan	453,733	346,217	50.52	107,516	40.83	61,658	47.98
	Missouri	262,366	205,634	51.23	56,732	42.73	51,302	48.24
	Oregon	--	--	--	--	--	--	--
	Texas	1,280,996	1,095,603	50.57	185,393	41.11	284,238	42.35
	Washington	218,051	175,625	53.75	42,426	43.25	61,031	46.681
Graduation (On-Time)	Massachusetts	139,337	114,413	93.92	24,924	--	--	--
	Michigan	453,733	385,359	85.49	68,374	--	--	--
	Missouri	262,366	210,423	91.08	51,943	--	--	--
	Oregon	123,833	101,692	83.80	22,141	--	--	--
	Texas	1,280,996	1,129,684	84.27	151,312	--	--	--
	Washington	218,051	176,505	82.48	41,546	--	--	--
Graduation (Within One Year of On Time)	Massachusetts	139,337	114,413	94.18	24,924	--	--	--
	Michigan	341,465	284,947	89.25	56,518	--	--	--
	Missouri	262,366	210,423	93.59	51,943	--	--	--
	Oregon	123,833	101,692	85.77	22,141	--	--	--
	Texas	1,280,996	1,129,684	87.73	151,312	--	--	--
	Washington	218,051	176,505	86.44	41,546	--	--	--

Table Notes: Sample sizes and entry percentiles for state stayers and out-migrators are based on the average of the grade 3 math and reading percentiles (i.e., percentiles at entry). For the test outcomes, with an underlying continuous distribution of scores the mean of each rank distribution should be 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. For graduation outcomes, we report the percent of students who graduate among stayers because percentiles are not informative. We focus on stayers because it is difficult to track graduation rates for state leavers across state lines, and we cannot match new entrants to a specific grade-3 cohort with certainty (because the grade-3 year is unobserved). In Michigan, data for lagged graduation outcomes are not yet available for the 2009 grade-3 cohort and the numbers reported in the bottom panel omit that cohort (they are for the 2006-2008 grade-3 cohorts only).

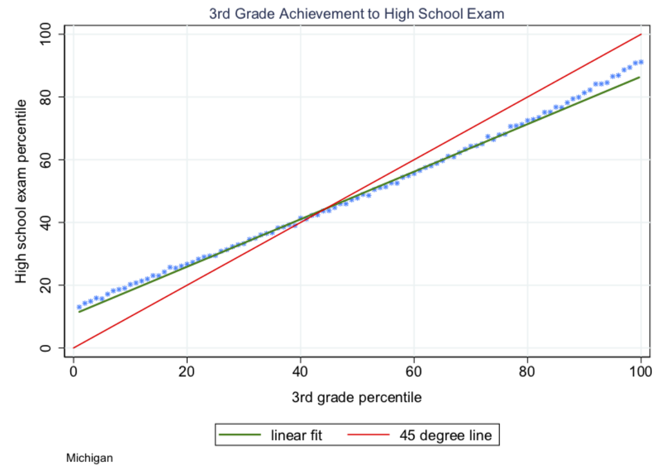
Appendix A
Supplementary Analyses

Appendix Figure A1. Binned scatter plots of percentiles on percentiles for each outcome to assess the linearity of the rank-rank relationships. Michigan, Missouri and Texas plots are available, others are pending.

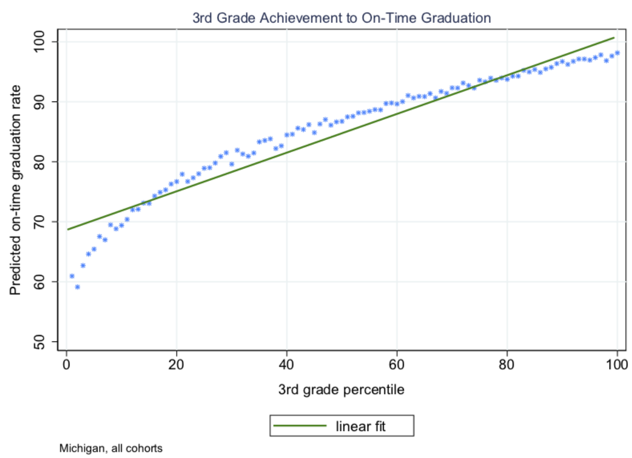
Michigan: Grade-8 Math & ELA



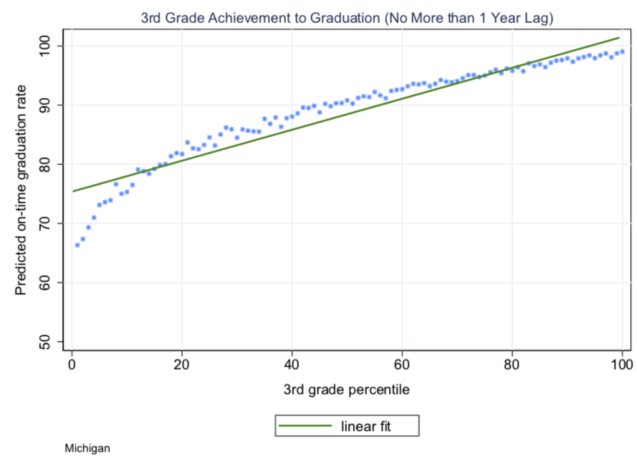
Michigan: HS Test



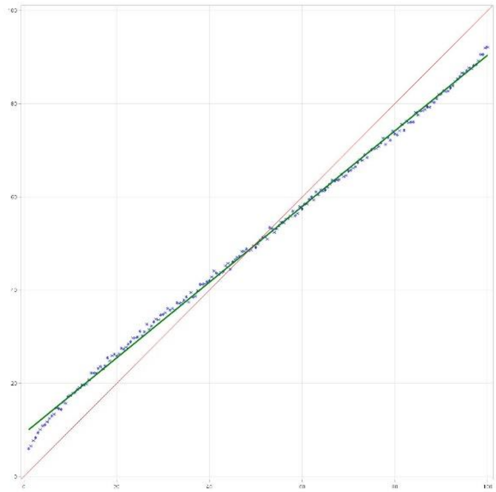
Michigan: Graduation



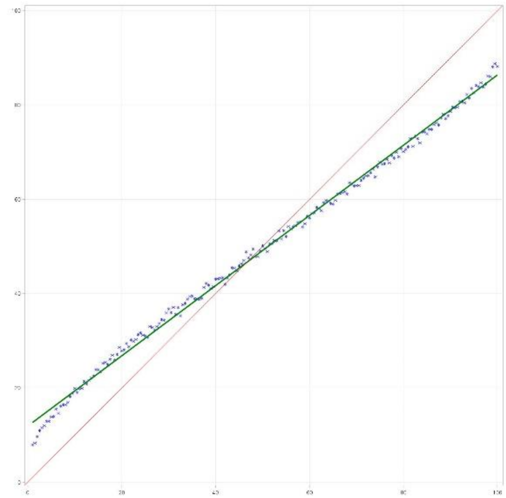
Michigan: Graduation +1



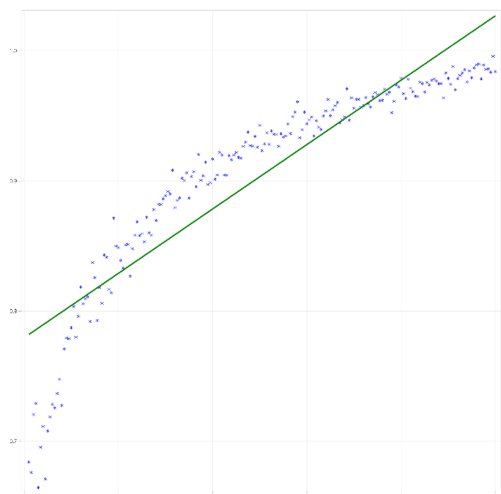
Missouri: Grade-8 Math & ELA



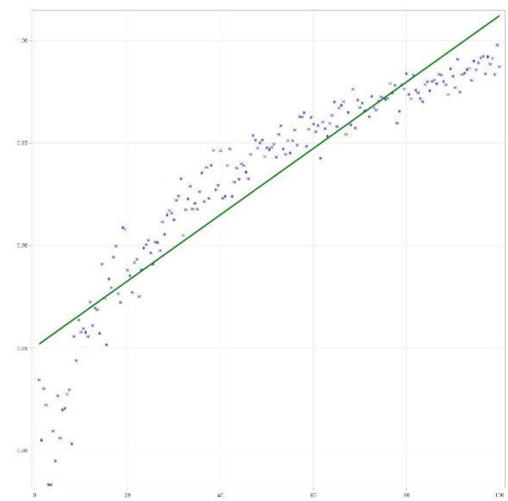
Missouri: HS Test



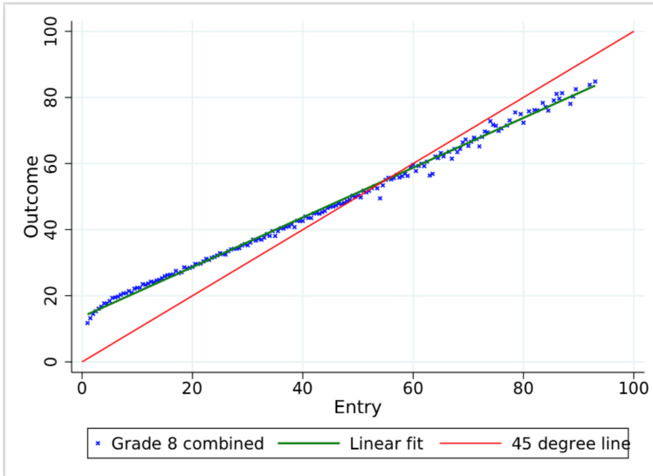
Missouri: Graduation



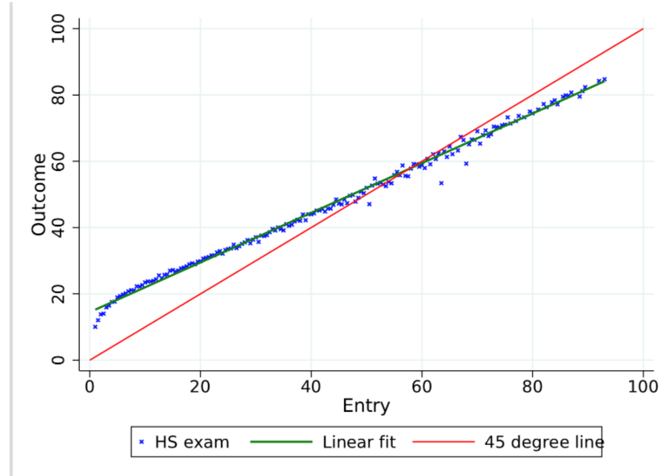
Missouri: Graduation +1



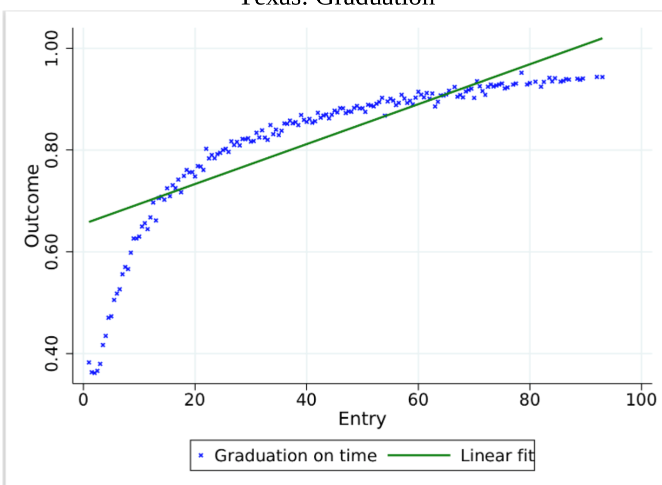
Texas: Grade-8 Math & ELA



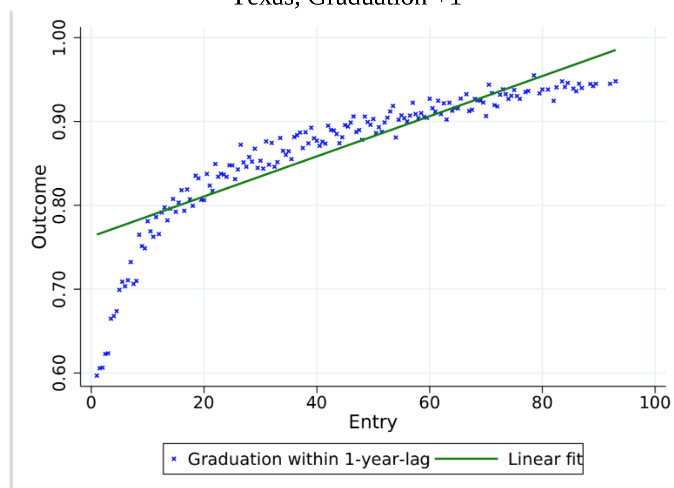
Texas: HS Test



Texas: Graduation

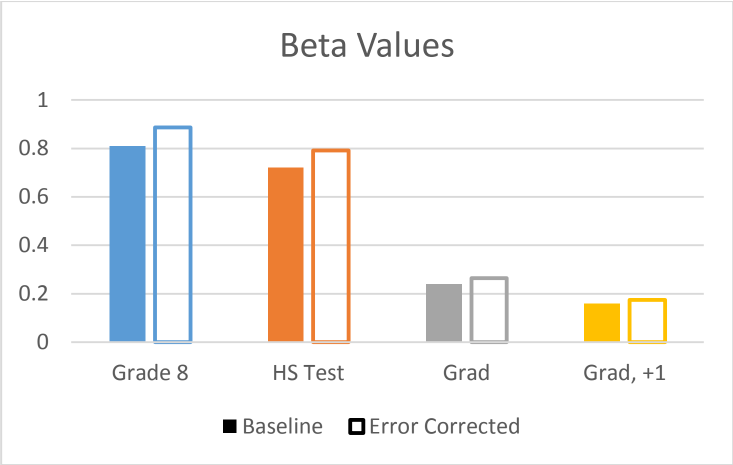


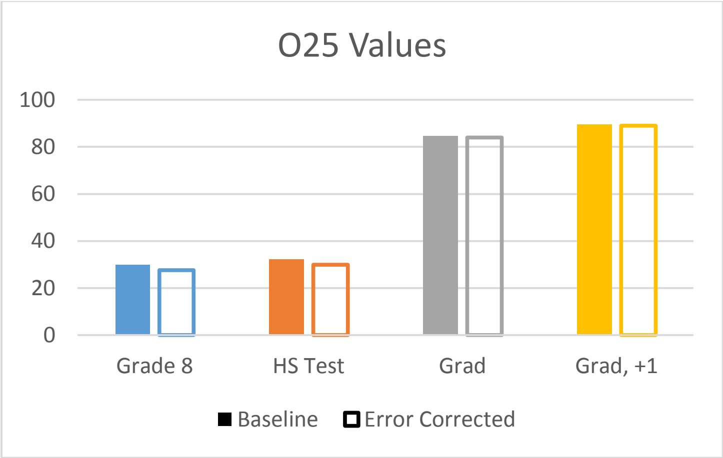
Texas, Graduation +1



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

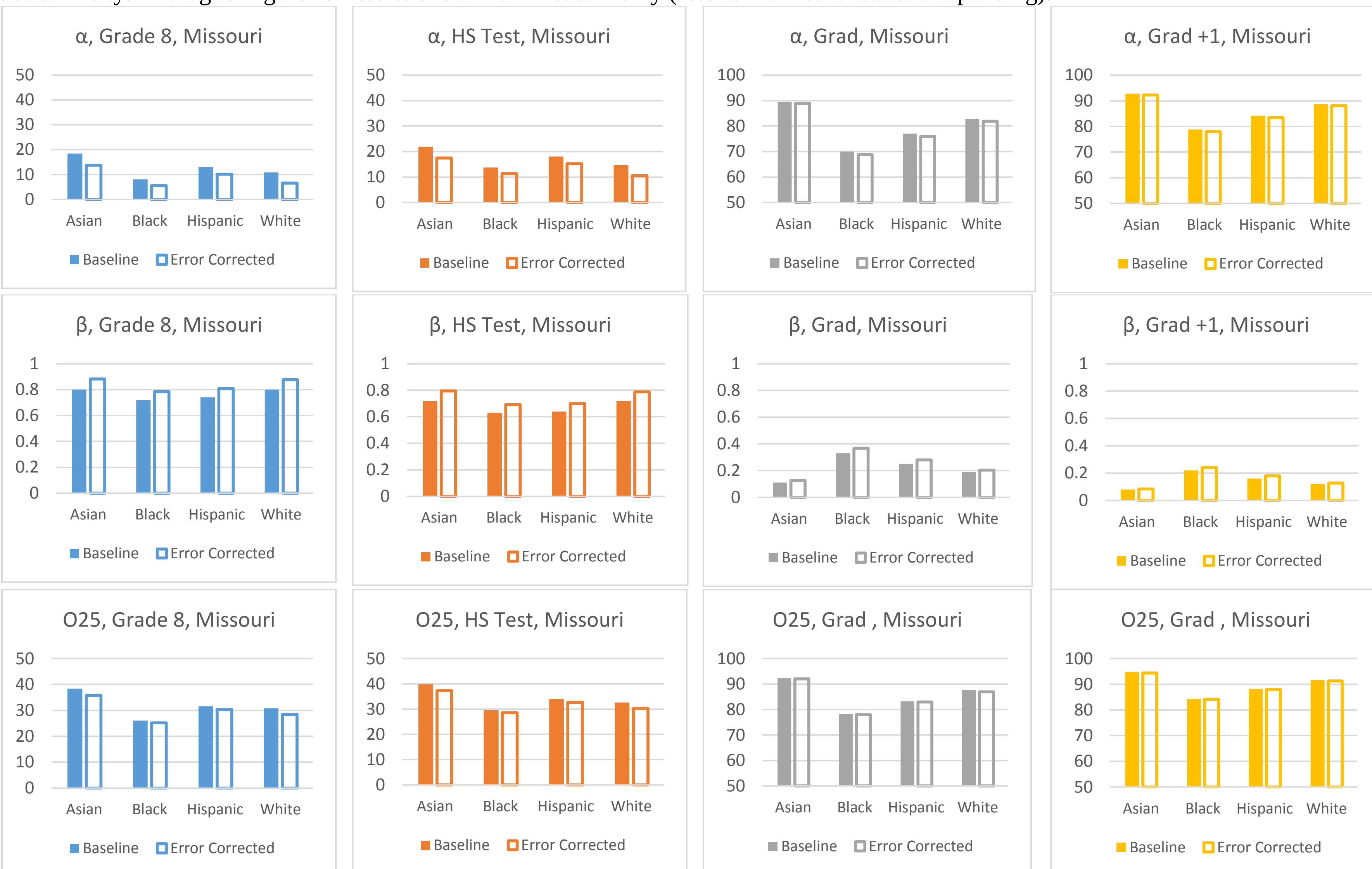
Appendix Figure A2. Estimates of β and $\overline{O}_{25}$ with and without correcting students' initial percentile ranks for measurement error. Analog to Figure 3. Results shown for Missouri only (results from other states are pending).



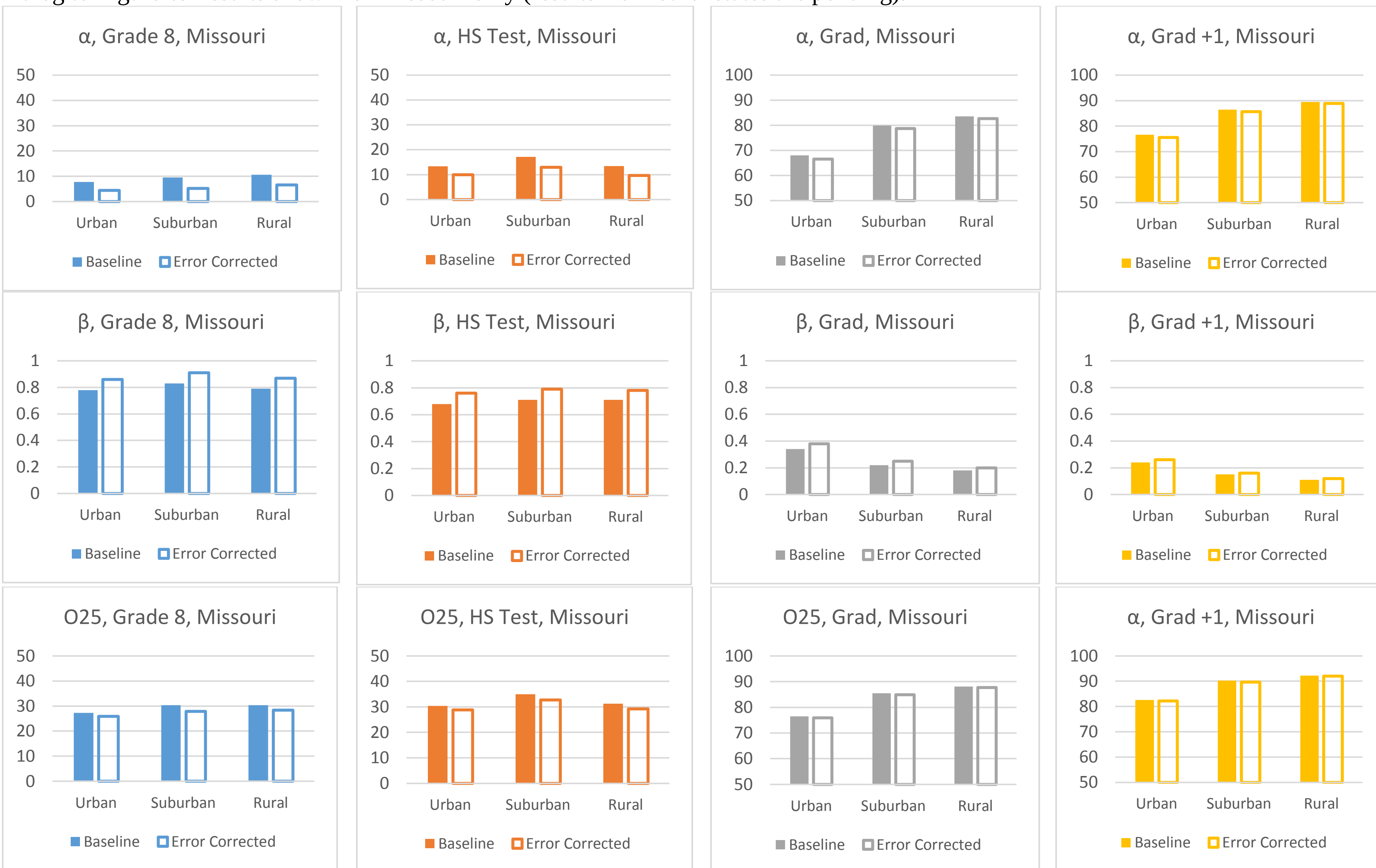


Notes: To correct for measurement error in the initial percentile rank based on grade-3 tests, we estimate an errors-in-variables regression where we specify the reliability ratio of the initial percentile rank using test-publisher provided Standard Errors of Measurement (SEM) for the grade-3 tests. As discussed in the text, the nature of the measurement error in the baseline (uncorrected) condition is such that β should be attenuated and $\overline{O}_{25}$ should be overstated, which is consistent with these results.

Appendix Figure A3. Comparison of baseline and measurement-error-corrected values of the mobility parameters ($\alpha, \beta, \bar{O}_{25}$) by race/ethnicity. Analog to Figure 4. Results shown for Missouri only (results from other states are pending).



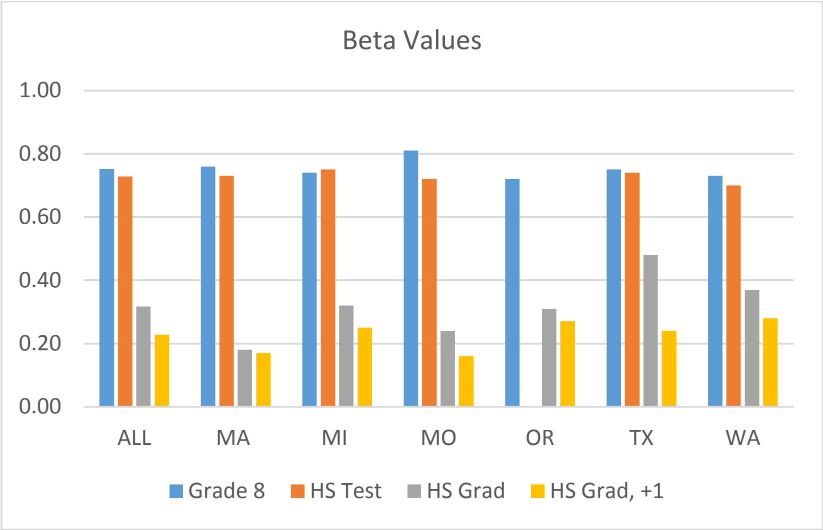
Appendix Figure A4. Comparison of baseline and measurement-error-corrected values of the mobility parameters ($\alpha, \beta, \bar{O}_{25}$) by urbanicity. Analog to Figure 6. Results shown for Missouri only (results from other states are pending).

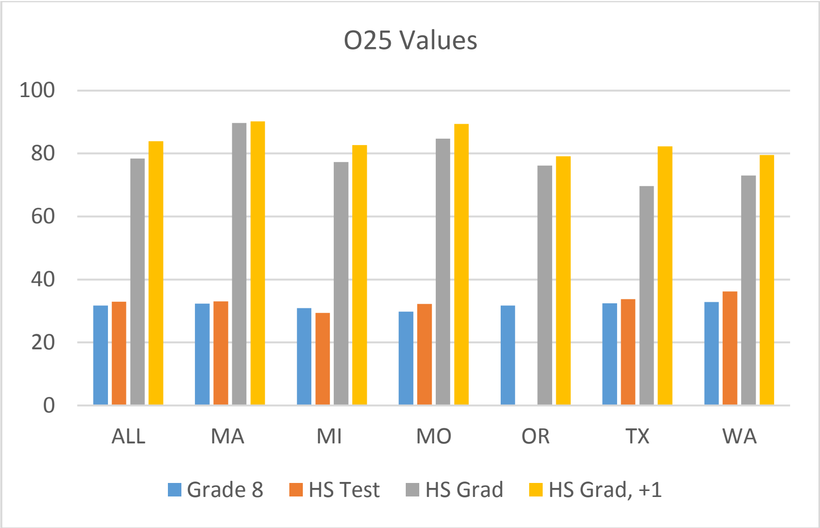


Appendix B
Rank Imputation Procedure for State Leavers
(TBD)

Appendix C
State-by-State Results Expansion
Figures and Data Tables

Figure C1. Estimates of β and $\overline{O}_{25}$ by state and outcome measure corresponding to Figure 3, baseline conditions.



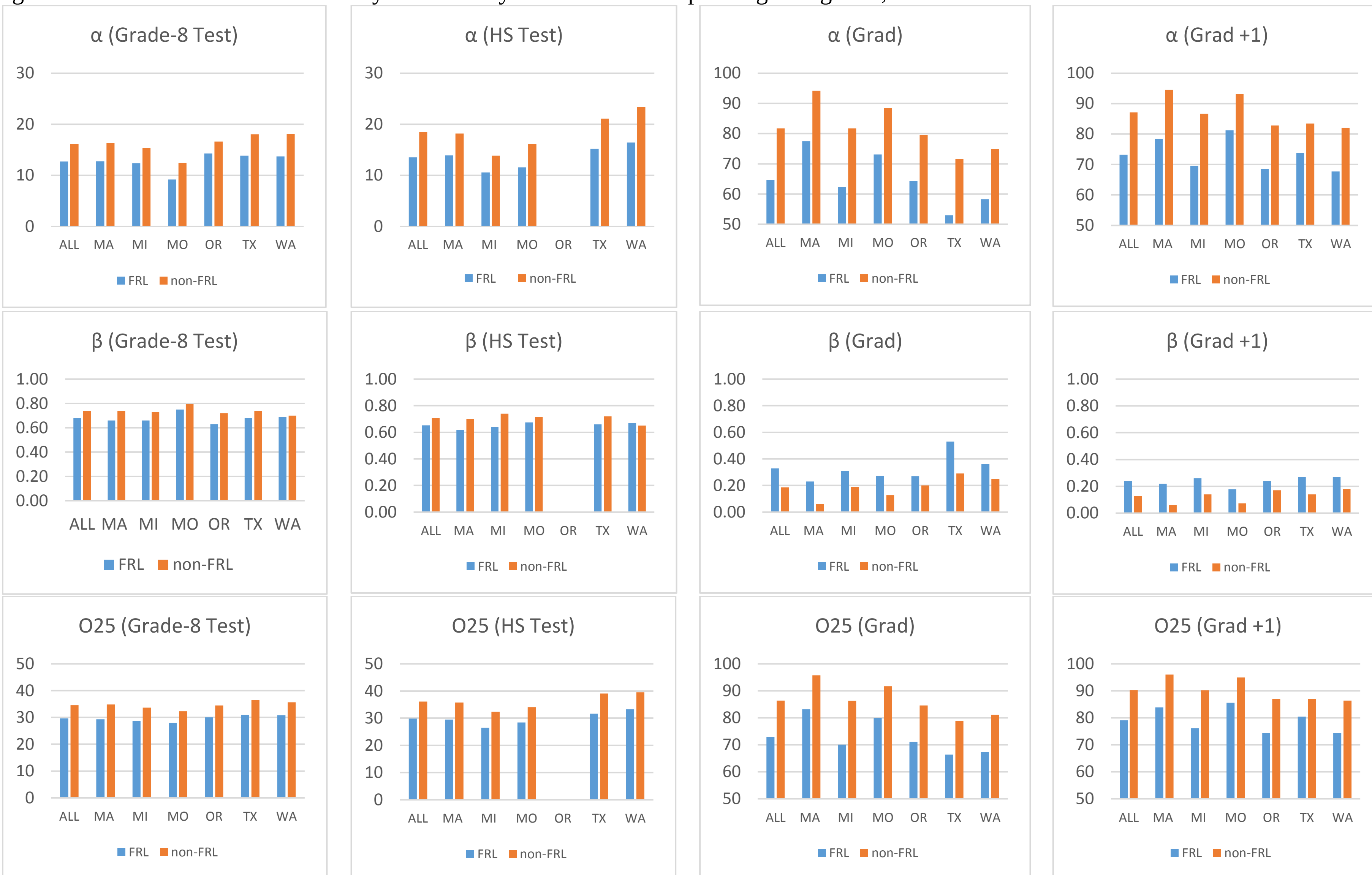


Notes: See notes to Figure 3.

Figure C2. Statewide academic mobility measures by race/ethnicity corresponding to Figure 4, baseline conditions.

Notes: See notes to Figure 4.

Figure C3. Statewide academic mobility measures by FRL status corresponding to Figure 5, baseline conditions.

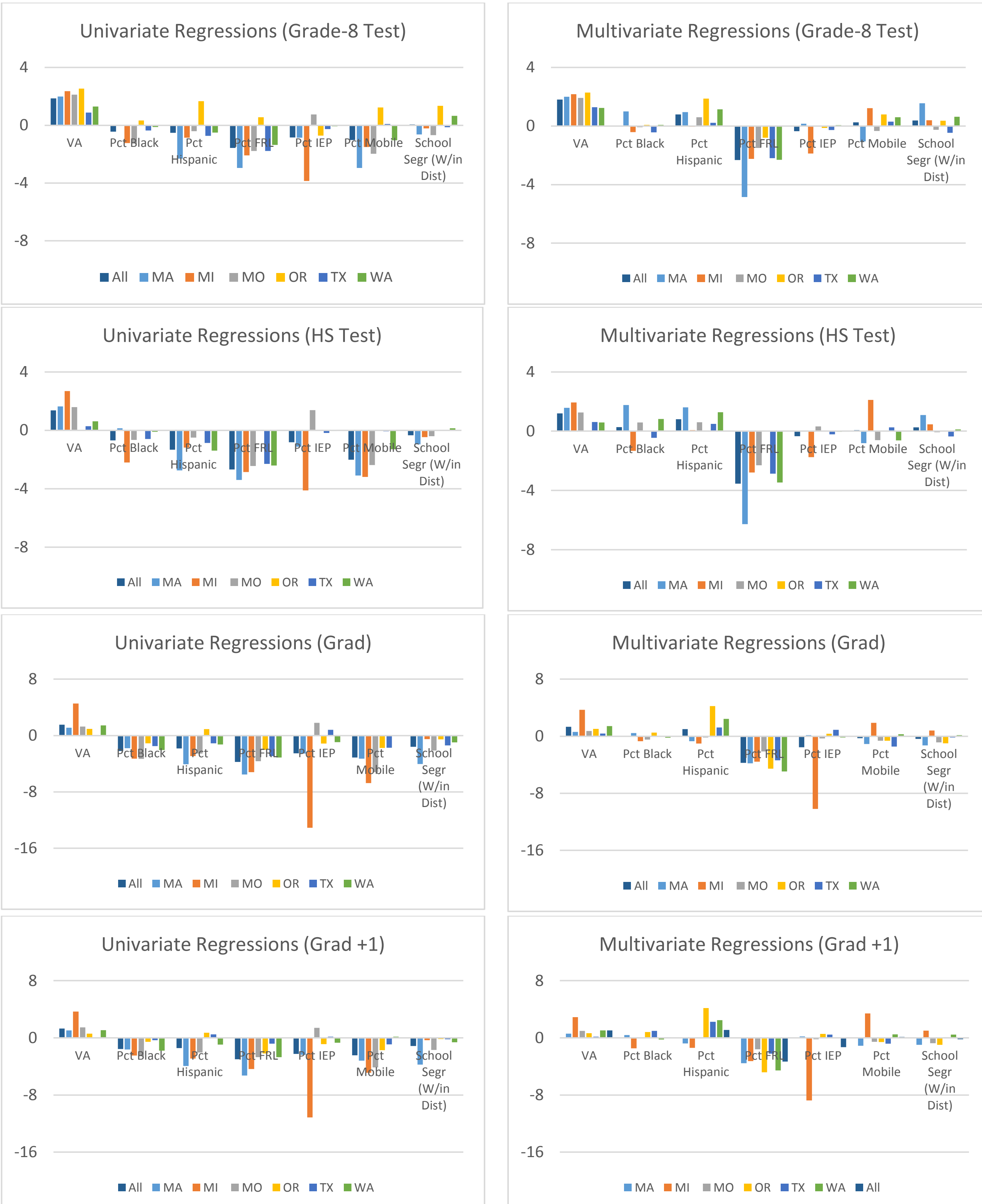


Notes: See notes to Figure 5.

Figure C4. Statewide academic mobility measures by school urbanicity corresponding to Figure 6, baseline conditions.

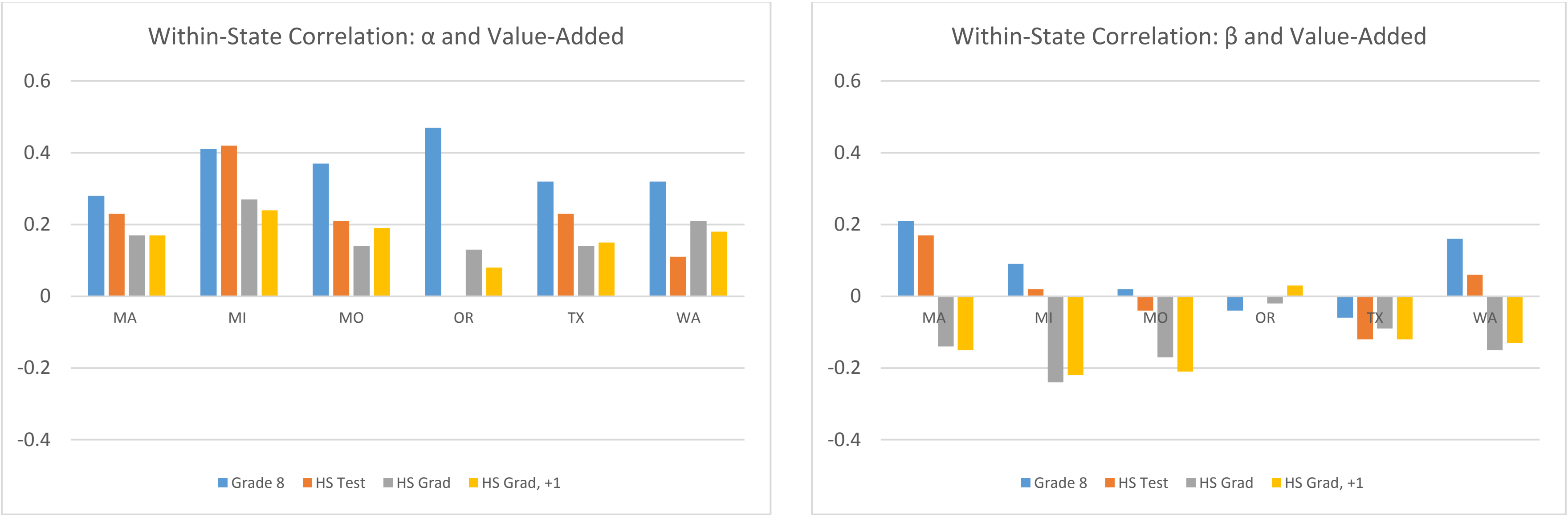
Notes: See notes to Figure 6.

Figure C5. Predictors of district-level $\bar{O}_{25d}$ corresponding to Figure 8.



Notes: See notes to Figure 8.

Figure C6. Correlations of district value-added with α_d and β_d in each state corresponding to Figure 9.



Notes: See notes to Figure 9.

Appendix Table C1. 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)	0.75	31.67	0.73	32.93	0.32	78.41	0.23	83.85
MA	0.76	32.36	0.73	33.05	0.18	89.71	0.17	90.17
MI	0.74	30.92	0.75	29.37	0.32	77.28	0.25	82.62
MO	0.81	29.74	0.72	32.26	0.24	84.66	0.16	89.41
OR	0.72	31.72			0.31	76.14	0.27	79.1
TX	0.75	32.43	0.74	33.78	0.48	69.64	0.24	82.25
WA	0.73	32.87	0.7	36.21	0.37	73	0.28	79.54

Appendix Table C2. 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)	22.54	0.74	41.05	25.44	0.71	43.20	85.30	0.16	89.18	89.65	0.11	92.30
MA	24.31	0.74	42.78	26.67	0.71	44.44	93.24	0.08	95.14	93.85	0.07	95.57
MI	22.88	0.74	41.29	23.2	0.76	42.18	88.49	0.12	91.51	92.77	0.08	94.68
MO	18.4	0.8	38.40	21.83	0.72	39.83	89.53	0.11	92.28	92.78	0.08	94.78
OR	18.52	0.75	37.22				82.11	0.19	86.87	84.87	0.16	88.87
TX	29.65	0.7	47.10	31.27	0.70	48.77	82.18	0.17	86.5	91.1	0.06	92.52
WA	21.5	0.72	39.51	24.22	0.66	40.78	76.23	0.26	82.78	82.5	0.19	87.38
Student Group: Black												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	11.51	0.68	28.39	13.96	0.64	29.93	64.80	0.37	73.99	74.21	0.26	80.67
MA	13.73	0.67	30.43	15.08	0.63	30.82	81.23	0.2	86.32	82.5	0.18	87.16
MI	11.7	0.62	27.14	9.43	0.56	23.37	63.73	0.36	72.84	71.78	0.3	79.22
MO	8.04	0.72	26.04	13.78	0.63	29.53	69.99	0.33	78.24	78.88	0.22	84.38
OR	10.01	0.65	26.36				60.76	0.34	69.24	68.12	0.27	74.76
TX	13.45	0.7	30.85	15.08	0.69	32.24	54.83	0.55	68.61	76.31	0.26	82.92
WA	12.11	0.7	29.54	16.42	0.69	33.69	58.23	0.41	68.71	67.64	0.32	75.6
Student Group: Hispanic												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	13.74	0.69	30.91	15.14	0.66	31.67	68.59	0.34	76.96	76.52	0.24	82.56
MA	12.93	0.67	29.61	13.9	0.63	29.64	76.37	0.28	83.33	77.4	0.27	84.04
MI	13.44	0.68	30.53	11.99	0.67	28.68	68.21	0.29	75.51	74.49	0.24	80.37
MO	13.1	0.74	31.6	18.03	0.64	34.03	77	0.25	83.25	84.19	0.16	88.19
OR	14.76	0.64	30.65				70.81	0.28	77.73	74.44	0.25	80.56
TX	14.07	0.7	31.53	15.48	0.68	32.48	55.48	0.54	68.98	75.64	0.27	82.5
WA	14.14	0.7	31.55	16.31	0.69	33.5	63.66	0.37	72.97	72.98	0.27	79.68
Student Group: White												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	14.26	0.74	32.81	16.67	0.71	34.47	73.47	0.27	80.26	79.69	0.20	84.78
MA	14.05	0.76	32.93	15.81	0.72	33.75	89.74	0.11	92.65	90.13	0.11	92.93
MI	14.09	0.73	32.3	13.22	0.73	31.5	73.05	0.27	79.83	79.32	0.22	84.71
MO	10.82	0.8	30.82	14.66	0.72	32.66	82.91	0.19	87.66	88.72	0.12	91.72
OR	14.47	0.71	32.28				68.3	0.31	76.04	72.04	0.27	78.85
TX	16.55	0.74	35.12	19.18	0.72	37.27	61.68	0.39	71.45	74.45	0.24	80.41
WA	15.55	0.71	33.38	20.47	0.67	37.18	65.13	0.35	73.94	73.48	0.26	80.06

Appendix Table C3. 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)	12.68	0.68	29.61	13.51	0.65	29.86	64.76	0.33	73.00	73.17	0.24	79.13
MA	12.74	0.66	29.27	13.87	0.62	29.48	77.47	0.23	83.16	78.43	0.22	83.84
MI	12.35	0.66	28.73	10.55	0.64	26.44	62.27	0.31	70.04	69.5	0.26	76.08
MO	9.18	0.749	27.9	11.55	0.675	28.43	73.14	0.272	79.94	81.18	0.177	85.59
OR	14.29	0.63	30.03				64.28	0.27	71.1	68.48	0.24	74.37
TX	13.86	0.68	30.92	15.17	0.66	31.66	53.03	0.53	66.39	73.73	0.27	80.46
WA	13.68	0.69	30.81	16.42	0.67	33.27	58.35	0.36	67.39	67.71	0.27	74.41

Student Group: non-FRL												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	16.13	0.74	34.57	18.52	0.71	36.15	81.72	0.19	86.40	87.09	0.13	90.26
MA	16.34	0.74	34.86	18.17	0.7	35.77	94.19	0.06	95.76	94.54	0.06	96.01
MI	15.33	0.73	33.6	13.84	0.74	32.39	81.71	0.19	86.33	86.61	0.14	90.12
MO	12.4	0.796	32.3	16.12	0.717	34.03	88.48	0.128	91.69	93.15	0.072	94.95
OR	16.61	0.72	34.5				79.5	0.2	84.56	82.82	0.17	87.02
TX	18.03	0.74	36.53	21.1	0.72	39.06	71.56	0.29	78.89	83.42	0.14	87.03
WA	18.07	0.7	35.6	23.35	0.65	39.48	74.89	0.25	81.16	82.02	0.18	86.4

Appendix Table C4. 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)	12.27	0.74	30.89	13.41	0.72	31.39	64.79	0.38	74.20	73.66	0.28	80.50
MA	12.66	0.72	30.76	13.63	0.7	31.1	77.24	0.26	83.74	78.4	0.25	84.53
MI	10.89	0.75	29.73	7.22	0.77	26.59	63.56	0.37	72.88	71.58	0.3	79.07
MO	7.78	0.78	27.28	13.39	0.68	30.39	68.01	0.34	76.51	76.62	0.24	82.62
OR	14.29	0.73	32.65				66.66	0.33	74.85	71.61	0.28	78.57
TX	13.42	0.74	31.94	14.84	0.74	33.35	53.73	0.53	67.05	74.76	0.26	81.38
WA	14.59	0.74	33	17.97	0.7	35.54	59.55	0.42	70.16	69	0.33	76.84

Student Group: Suburban												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	13.40	0.76	32.39	16.06	0.72	34.20	72.63	0.29	80.06	80.05	0.21	85.26
MA	13.89	0.77	33.01	15.61	0.73	33.8	88.47	0.14	91.89	88.93	0.13	92.22
MI	12.84	0.75	31.52	11.63	0.75	30.49	71.88	0.29	79.19	78.29	0.23	84.1
MO	9.56	0.83	30.31	17.19	0.71	34.94	79.94	0.22	85.44	86.45	0.15	90.2
OR	14.19	0.74	32.59				69.27	0.31	77.09	73.22	0.27	80
TX	14.71	0.75	33.5	15.91	0.75	34.72	61.46	0.43	72.26	79.34	0.2	84.4
WA	15.23	0.73	33.42	19.97	0.68	37.04	64.75	0.35	74.51	74.08	0.26	80.62

Student Group: Rural												
	α	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*	*α*	*β*	*O25*
All (Avg)	13.26	0.74	31.64	14.91	0.71	32.74	73.67	0.27	80.41	80.20	0.20	85.08
MA	14.04	0.75	32.69	15.4	0.72	33.38	89.77	0.11	92.55	90.22	0.11	92.87
MI	13.76	0.72	31.65	12.47	0.72	30.36	71.96	0.28	78.85	77.98	0.23	83.63
MO	10.62	0.79	30.37	13.5	0.71	31.25	83.58	0.18	88.08	89.5	0.11	92.25
OR	13.72	0.68	30.78				69.6	0.28	76.68	72.86	0.25	79.15
TX	13.25	0.76	32.26	15.32	0.73	33.61	60.9	0.43	71.6	76.03	0.23	81.8
WA	14.18	0.72	32.08	17.87	0.69	35.11	66.23	0.34	74.71	74.59	0.25	80.78

Appendix Table C5. Numeric results corresponding to Figure 8. Coefficient values on standardized variables. Statistical significance at the 5 percent level or better is denoted by *. Standard errors are suppressed for brevity.

	Univariate, Grade-8 Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.98*	0.02	-2.31*	-2.95*	-0.86*	-2.95*	-0.64
MI	2.36*	-1.24*	-0.85*	-2.09*	-3.85*	-1.51*	-0.21*
MO	2.11*	-1.26*	-0.41*	-1.78*	0.75*	-1.98*	-0.69*
OR	2.54*	0.32	1.67*	0.55	-0.72*	1.24*	1.35*
TX	0.88*	-0.36*	-0.73*	-1.78*	-0.27	0.09	-0.14
WA	1.29*	-0.12	-0.51*	-1.36*	-0.07	-1.03*	0.65*
All (Avg) †	1.86	-0.35	-0.48	-2.04	-1.30	-1.15	0.08

	Multivariate, Grade-8 Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.99*	1.01*	0.95	-4.85*	0.15	-1.04*	1.56*
MI	2.18*	-0.42	-0.05	-2.24*	-1.88*	1.22*	0.39*
MO	1.92*	-0.08	0.60*	-1.50*	-0.02	-0.34	-0.25*
OR	2.28*	0.07	1.87*	-0.79	-0.13	0.79	0.37
TX	1.28*	-0.44	0.23	-2.20*	-0.28	0.30	-0.47
WA	1.24*	0.07	1.14*	-2.31*	0.06	0.60*	0.64*
All (Avg)†	1.82	0.00	0.85	-3.68	-0.50	0.45	0.27

	Univariate, HS Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.63*	0.13	-2.75*	-3.40*	-1.13*	-3.10*	-0.97*
MI	2.68*	-2.21*	-1.19*	-2.86*	-4.11*	-3.19*	-0.48*
MO	1.59*	-0.66*	-0.50*	-2.44*	1.38*	-2.38*	-0.41*
OR							
TX	0.28	-0.61*	-0.86*	-2.30*	-0.18	-0.08	0.03
WA	0.62*	-0.10	-1.39*	-2.41*	-0.04	-1.30*	0.14
All (Avg)†	1.36	-0.50	-1.26	-3.46	-1.40	-2.33	-0.26

	Multivariate, HS Test						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.58	1.77*	1.60*	-6.27*	0.00	-0.80*	1.09*
MI	1.94*	-1.33*	0.07	-2.78*	-1.75*	2.11*	0.47*
MO	1.26*	0.59*	0.60*	-2.31*	0.32	-0.61	-0.07
OR							
TX	0.62*	-0.45*	0.49*	-2.86*	-0.21	0.25	-0.35
WA	0.59*	0.81*	1.29*	-3.46*	-0.05	-0.62*	0.12
All (Avg)†	1.20	0.25	0.70	-5.31	-0.65	0.19	0.13

	Univariate, Grad						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.12*	-1.79*	-4.08*	-5.52*	-2.55*	-3.27*	-4.04*
MI	4.55*	-3.29*	-2.93*	-5.20*	-13.08*	-6.75*	-0.50*
MO	1.28*	-3.33*	-2.55*	-3.63*	1.79*	-5.18*	-2.12*
OR	0.93	-1.09*	0.91	-2.01*	-1.15*	-1.76*	-0.55
TX	-0.07	-1.51*	-1.11*	-2.92*	0.82*	-1.72*	-1.41*
WA	1.44*	-2.07*	-1.28*	-3.12*	-0.93*	0.02	-0.97*
All (Avg)†	1.54	-1.95	-1.68	-4.92	-4.07	-3.67	-1.53

	Multivariate, Grad						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	0.62*	0.46	-0.68	-3.79*	0.16	-1.08*	-1.27*
MI	3.71*	-0.69	-1.03	-3.56*	-10.18*	1.88	0.82
MO	0.73*	-0.46	-0.18	-2.14*	-0.31	-0.63	-0.85*
OR	1.04*	0.50	4.24*	-4.55*	0.34	-0.63	-0.99
TX	0.37	0.02	1.22*	-3.36*	0.89*	-1.43*	-0.19
WA	1.43*	-0.20	2.42*	-4.94*	-0.17	0.27	0.16
All (Avg)†	1.32	-0.06	1.31	-7.18	-2.56	-0.48	-0.48

	Univariate, Grad+1						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	1.05*	-1.64*	-3.93*	-5.27*	-2.39*	-3.20*	-3.74*
MI	3.67*	-2.47*	-2.91*	-4.36*	-11.12*	-4.90*	-0.34*
MO	1.45*	-2.56*	-2.06*	-2.72*	1.41*	-4.15*	-1.70*
OR	0.59	-0.57	0.72	-2.18*	-0.88	-1.73*	-0.17
TX	0.02	-0.32	0.50*	-0.80*	0.15	-0.93*	-0.21
WA	1.06*	-1.81*	-0.95*	-2.70*	-0.69	0.17	-0.61
All (Avg)†	1.31	-1.39	-1.28	-4.00	-3.58	-2.87	-1.08

	Multivariate, Grad+1						
	Value Added	Pct Black	Pct Hispanic	Pct FRL	Pct IEP	Pct Mobile	School Seg. Index
MA	0.58*	0.38	-0.77	-3.56*	0.19	-1.11*	-0.99*
MI	2.89*	-1.46	-1.41	-3.26*	-8.77*	3.42	0.99*
MO	0.96*	-0.08	-0.07	-1.60*	-0.20	-0.56	-0.75*
OR	0.66	0.82	4.17*	-4.81*	0.54	-0.58	-0.97
TX	0.17	0.97*	2.25*	-2.21*	0.45*	-0.81*	0.11
WA	1.05*	-0.25	2.48*	-4.58*	0.04	0.47	0.44
All (Avg)†	1.05	0.11	1.44	-6.73	-1.73	0.06	-0.32

Note: † Statistical significance is not reported for the "All (avg)" values because they are not directly generated from a regression (they are simple average values of the state-by-state regression coefficients). Note that the multivariate regressions also control for the grade-3 school urbanicity category but these coefficients are suppressed for brevity (and they prove no new insights beyond what is shown in Figure 6).

Appendix Table C6. Numeric results corresponding to Figure 9.

	Grade-8 Test		HS Test		Grad		Grad +1	
	Correlation of VA with:							
	α	*β*	*α*	*β*	*α*	*β*	*α*	*β*
All (Avg)	0.36	0.06	0.24	0.02	0.18	-0.14	0.17	-0.13
MA	0.28	0.21	0.23	0.17	0.17	-0.14	0.17	-0.15
MI	0.41	0.09	0.42	0.02	0.27	-0.24	0.24	-0.22
MO	0.37	0.02	0.21	-0.04	0.14	-0.17	0.19	-0.21
OR	0.47	-0.04			0.13	-0.02	0.08	0.03
TX	0.32	-0.06	0.23	-0.12	0.14	-0.09	0.15	-0.12
WA	0.32	0.16	0.11	0.06	0.21	-0.15	0.18	-0.13