CEDR Working Paper | No. 0522204-1

CEDR
Center for Education Data & Research

Progression and Predictors of Public-School Student Outcomes in Washington State

Dan Goldhaber
University of Washington
American Institutes for Research

Katherine Baird
University of Washington

Suvekshya Gautam
University of Washington

The Center for Education Data & Research
University of Washington Bothell
3876 Bridge Way N. Ste. 201 Seattle, WA 98103
(206) 547-5585 cedr@u.washington.edu
www.cedr.us

Acknowledgements: We are grateful to Washington's Office of Superintendent of Public Instruction for providing the data that made the analyses in this report possible. We also wish to acknowledge helpful comments from Maia Goodman Young, Stephanie Liddle, and Rebecca Goodvin about the data and assessments and research assistants from Rafia Nisat.

This work is supported by the Bill & Melinda Gates Foundation. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the institutions to which the authors are affiliated or the study's funder.

Correspondence regarding this article should be addressed to Dan Goldhaber, Center for the Analysis of Longitudinal Data in Education Research, American Institutes for Research, 3876 Bridge Way N, Suite 201, Seattle, WA 98103. Email: dgoldhaber@air.org • www.caldercenter.org

Suggested citation:
Goldhaber, D., Baird, K., & Gautam, S. (2024). Progression and Predictors of Public-School Student Outcomes in Washington State. CEDR Working Paper No. 05222024-1. University of Washington, Seattle, WA.

Abstract: In this paper we analyze the extent to which a mandated kindergarten assessment predicts 3rd grade outcomes, and the academic progression for students from 3rd grade to high school. We find that the kindergarten assessment strongly predicts 3rd grade outcomes, with the math skills assessment being especially predictive of 3rd grade academic outcomes. The kindergarten assessments also illustrate the degree to which there are large inequities in skills when students are assessed in kindergarten. Students from historically disadvantaged groups enter kindergarten with significantly fewer readiness standards met. Our analysis of student academic progression from 3rd grade through high school echoes the kindergarten to 3rd grade results. The 3rd grade test assessment is strongly predictive of all high school outcomes, and we see that those students eligible for free- or reduced-price lunches and students of underrepresented racial or ethnic groups are less likely to have upward academic mobility. In sum, we observed limited academic mobility; students who start out behind generally stay behind.

1 INTRODUCTION

Understanding the factors explaining variation in students' academic performance has been a focus of policy and research going all the way back to the *Coleman Report* (Coleman et al., 1966). Finding what drives academic achievement and achievement trajectories is key to informing policy and practice decisions designed to address student needs. Thus, tracing the typical academic progression of student groups is important to both identify which students may benefit from interventions as well as to discern the points in a student's academic career when interventions are most likely to be necessary.

It is well-known that measures of students' academic achievement at one point in school are highly predictive of academic performance at later stages in schooling (e.g., Gray-Lobe et al., 2022), and post-schooling outcomes (e.g., Chetty et al., 2014). Also, there are various policies and interventions that are predicated on the notion that the achievement of key skills should be achieved by certain grades in order to help ensure later schooling success. For instance, it has become a norm to assert that reading by the 3rd grade is a necessary skill (Fiester, L., 2010). Consequently, several states require students to achieve test thresholds or risk being retained in the 3rd grade (Jacob, 2017; LiCalsi et al., 2019). Various states and localities have also developed early warning systems and diagnostic tools to help alert practitioners that individual students are at risk of not meeting specific skills or benchmarks such as high school graduation (e.g., Allensworth, 2013; Curtin et al., 2012).

But while there is extensive research linking academic and demographic factors to K-12 schooling outcomes, most of it relies on linkages across short grade spans (e.g., elementary to middle, or middle to high school) and focuses mainly on test outcomes. There is also only a nascent literature connecting *pre*-3rd grade school readiness indicators (e.g., Herring et al. 2022, Justice et al., 2019) to later schooling outcomes.[1]

In this research, we utilize rich administrative data from Washington State to assess how demographic characteristics, Kindergarten Readiness Indicators (KRIs), and early achievement indicators such as 3rd grade test scores and absences predict the academic progression of K-12 public school students. This descriptive paper contributes to the literature in several ways. First, it is amongst a small number of studies that connects multiple cohorts of students throughout their K-12 progression in public schools. Second, it is one of the few studies that considers outcomes beyond test scores, examining a variety of non-test outcomes including absences, disciplinary incidences, and advanced course taking, among other outcomes. Finally, we leverage multiple cohorts of students to explore the extent to which predictors of high school outcomes change over time. Specifically, we address the following questions:

1. What is pattern of student test and non-test outcomes for different groups of students as they progress through grades in the Washington state public education system?
2. To what extent do different indicators, or combinations of indicators, predict different outcomes?

[1] This is not surprising given that state-level testing requirements typically begin in the 3rd grade.

3. Is there evidence of changes across cohorts in the predictive power of early indicators (e.g., test scores, demographics) of later academic outcomes?

We find that student performance on kindergarten readiness assessments is highly predictive of performance on 3rd grade reading and math assessments, and predictive of the absences in 3rd grade and probability of 3rd grade disciplinary incidents. But we also find that the various KRIs used by the state are differentially predictive of outcomes, suggesting there is scope for improvement in using them to make predictions about later academic outcomes.

Not surprisingly given existing research on differences in readiness of students entering kindergarten (e.g., Farkas and Beron, 2004; Fryer and Levitt, 2004; Reardon and Portilla, 2016), we document large differences in kindergarten readiness measures, 3rd grade performance, and academic mobility by student race/ethnicity and free- and reduced-priced lunch (FRPL) status. In other words, much of the inequity in educational achievement is present when students begin in kindergarten.

In exploring academic progression from grades 3 to 12, we find, again consistent with prior research, that 3rd grade test scores are highly predictive of various middle and high school outcomes; 3rd grade tests are highly correlated with math and reading scores in grades 8 and 10 and more moderately (though still significantly) predictive of high school GPA and graduation rates. There is little difference in the predictive power of 3rd grade and 8th grade tests. This suggests that schooling experiences and interventions that occur between 3rd and 8th grade do relatively little to alter the trajectory of student achievement.

Underrepresented minority students and those who receive FRPL have worse high school outcomes even controlling for 3rd grade test achievement, this is especially true for disciplinary incidents. While demographic indicators and 3rd grade test performance are highly predictive of high school outcomes across all cohorts, the strength of these relationships tends to be diminished for later cohorts of students in our sample.

2 BACKGROUND ON THE FACTORS PREDICTING PROGRESSION AND SUCCESS THROUGH PUBLIC SCHOOLS

Numerous studies have highlighted the importance of various student characteristics (e.g., measures of student achievement, demographics, and socio-economic classifications) in predicting long-term outcomes such as later test achievements, high-school course taking, GPA, high school graduation, and post-secondary outcomes such as college-going and labor market earnings. Specifically, test scores in early grades have been shown to not only measure the performance of the students at that grade but also predict of future performance. Testing students annually and using the results to identify and communicate about students' needs and to inform policy decisions has been one of the primary strategies to battle educational inequities. While the use of tests as a measure of identifying learning gaps is contentious (e.g., Forte, 2021; Koretz, 2017; Strauss, 2015), past research illustrates the importance of early academic indicators in predicting future academic success and that tests can serve this predictive functions.[2]

[2] See Goldhaber and Ozek (2019) for a review.

Additionally, differences in academic performances between students during their K-12 education predict post-secondary outcomes, including earnings (Cawley, Heckman, and Vytlacil, 2001; Cunha et al., 2006; Murnane et al., 2000).

As noted above, many studies linking demographics, measures of school readiness, or achievement at one point in a student's career to later student outcomes only consider a relatively short period of time between the intervention and the outcome of interest or are focused on limited samples of students (Goldhaber et al., 2020). Thus, it is challenging to know what points in a student's academic career are most consequential to their progress. Much of the existing research focuses on the predictability of achievement in early grades—more specifically 3rd grade test scores—in determining outcomes such as test scores at higher grade levels, high school GPA, and graduation rate (Goldhaber et al., 2022; Easton et al., 2017). Austin et al. (2023) use the concept of academic mobility to describe the extent to which students' ranks in the distribution of academic performance change during their school careers and show the existence of considerable heterogeneity in the academic mobility across different student subgroups and school districts.

There is also significant academic focus on inequality in educational achievement and the degree to which gaps in student achievement change as students move from grade to grade. Research indicates that these disparities mainly emerge during the early stages of childhood, preceding the commencement of formal education (von Hippel and Hamrock, 2019), and persist throughout the students' academic progression (Clotfelter et al., 2009; Finkelstein and Fong, 2008). The findings across both reading and math scores (Betts et al., 2003; Clotfelter et al., 2009; Reardon, 2011) and non-test outcomes (Gamoran, 1987; Kelly, 2009; Lee, 2002) document persistence in student achievement gaps.

Several recent studies have shown that racial/ethnic gaps in student achievement have been declining throughout the last decade or so, though the magnitude of change in these educational gaps vary for different races and for different outcomes. For instance, the White-Hispanic gap for test scores has narrowed more than the White-Black gap and the evidence for other non-test school outcomes are less pronounced (Reardon and Portilla, 2016). Studies have also shown that the magnitude of changes in gaps vary by other factors such as neighborhood segregation, lagged school and home inputs, cultural and social factors, and parental characteristics (Card and Rothstein, 2007, Gamoran 2001; Kao and Thompson 2003; Todd and Wolpin, 2007; Hanushek and Rivkin 2009; Domina et al., 2017; Duncan and Magnuson, 2005). The degree to which schools mitigate or amplify overall inequality is a topic of significant discussion, as highlighted by varying viewpoints (Dumont and Ready, 2020; von Hippel et al., 2018; Northrop, 2017; Domina et al., 2017).[3]

While not as widely studied, researchers have also explored the link between early test scores and non-test outcomes such as absenteeism, disciplinary incidences, and course-taking behavior. Studies find that early grade scores are strongly correlated with later grade test and

[3] Teacher quality and within-school sorting of teachers and students may cause some of this mitigation or amplification (Goldhaber et al., 2022; Northrop, 2017), and there is some evidence that gaps decrease over time if schools and teachers cater to the needs of disadvantaged students by offering targeted instruction (Downey et al., 2004; von Hippel et al., 2018).

non-test outcomes. This research has also shown that variables such as race. gender, income, disability status, are strong predictors of test and non-test outcomes, with a proportion of the differential outcomes explained by these variables (Fryer Jr and Levitt, 2006; Figlio, 2005; Goldhaber, Wolff, and Daly, 2020; Sorenson, 2019; Todd and Wolpin, 2007).

Research also documents large disparities in academic performance along racial and ethnic lines when children enter kindergarten which continue through high school (Fryer and Levitt, 2004; Reardon and Galindo, 2009; Hemphill, Vanneman, and Rahman, 2011). Nitardy et al. (2014) found that the mean GPA in high school was significantly lower for minority racial and ethnic groups. Moreover, evidence of an engagement-achievement paradox has been found where Black students report more engagement and motivation, but lower GPA compared to White students (Kao and Tienda, 1998; Shernoff and Schmidt, 2008). Racial differences and socioeconomic status have been identified as a primary contributor to achievement gaps (Fryer and Levitt, 2004; Rothstein and Wozny, 2013), and some research suggests that while education can enhance measured abilities, it doesn't necessarily bridge the gaps between children from different racial and economic backgrounds and may, if anything, exacerbate them (Cunha et al. 2006). Black and Hispanic students also have the highest dropout rates (Lofstrom, 2007). Cameron and Heckman (2001) found that family factors such as family composition, parental education and family income explain all the Black-White gap in high school graduation rates and most of the Hispanic-White gap.

Less is known about very early indicators of student readiness for K-12 schooling. A few states have systems in place to assess kindergarten readiness (Garver, 2020; Weisenfeld et al., 2020). We found two studies that assess the extent to which kindergarten readiness indicators predict later schooling outcomes. Specifically, a recent study of Virginia (Herring et al, 2022) examines the probability that a given child with a certain level of performance on their early Phonological Awareness Literacy Screening (PALS) assessments would meet advanced standards on their 3rd grade Standards of Learning (SOL) exam, conditional on their literacy skills at kindergarten entry. It finds that there are large disparities, by race and socio-economic status, in the likelihood that children achieve reading proficiency in 3rd grade even when accounting for students' literacy skills in kindergarten. However, the study only focuses on literacy-related kindergarten readiness and only considers 3rd grade state assessment scores as the outcome of interest. A similar study in Ohio (Justice et al., 2019) examines 3rd grade reading scores as the outcome of interest by comparing students with various levels of kindergarten readiness scores. The researchers find that 25 percent of the variation in 3rd-grade ELA scores was explained by the children's kindergarten readiness assessment scores. However, other studies using national and international data on assessments at school entry show mixed evidence of how persistent levels of early school readiness are as the student progress along their education.

3 DATA, MEASURES, AND ANALYTICAL SAMPLE

The data for this study comes from several datasets maintained by the Office of Superintendent of Public Instruction (OSPI). The primary sources of the data are Core Student Record System, the Comprehensive Education Data and Research System, and the WaKIDS data inventory described below. Individual student records are longitudinally linkable across years based on a unique student record ID that the state maintains.

3.1 WaKIDS Data

WaKIDS was piloted in SY 2010/11 and was expanded over the next few years to schools that volunteered to participate. By SY 2014/15, over 50% of the kindergarteners participated in the WaKIDS assessments, with the program being made mandatory for all schools by the SY 2017/18 school year. It is worth noting that WaKIDS was initially rolled out in schools reporting the highest rates of children qualifying for FRPL (Goodvin, et al., 2020). [4] While data from WaKIDS readiness assessments are available from SY 2010/11, we only utilize SY 2014/15 and SY 2015/16 data as these cohorts of kindergartners had over 50% of students took the assessment and we can link students' WaKIDS assessments to their 3rd grade state assessments. [5]

As part of the WaKIDS assessments, teachers observe the child's skills along six domains: **socio-emotional, physical, cognitive, language, literacy,** and **mathematics**. For each domain, children are deemed as "kindergarten ready" if they meet/exceed their age-appropriate benchmark score. [6] Due to the possible differences in the assessment measures and benchmark scores, we standardize the WaKIDS assessment scores by year and assessment type for these years. The comprehensive data on kindergarten readiness includes multiple assessment measures including:

1. Scale scores for each of the six domains: Students are scored based on their performances on certain tasks for each of the domains. Teachers observe students perform these tasks and score them based on their performance. The distribution of the scale scores for each of the domains along with the threshold [7] that characterizes if the student is deemed ready for that domain is shown in **Figure 1**. For analysis using the scale scores, the scores were standardized by year and assessment type.

2. Readiness flag for each individual domains: For each domain, a readiness flag indicates whether the students meet/exceed the set age-appropriate developmental expectation. A student is deemed "kindergarten ready" in each domain if their observed behavior falls within the skill level expected of a kindergartener.

---

[4] Given that we are utilizing the early cohorts of students with WaKIDS assessments, students from low-income families are likely overrepresented in our analysis samples.

[5] Kindergarteners who were assessed with WaKIDS in SY 2016/17 do not have 3rd grade state assessment scores due to the lack of state testing in the spring of the SY 2019-20 during the COVID-19 pandemic. Because of how disruptive COVID-19 was to early childhood education, we do not use WaKIDS data from students who entered 3rd grade after the 2019-20 school year.

[6] Kindergarten teachers rate the students within each of the domains, with scores based on a provided rubric. If the scores for a given domain exceed specific readiness threshold, students are deemed kindergarten ready in that domain. Refer to Appendix A1 for scoring objectives for each domain.

[7] Note that the thresholds shown in the figure are not "official" threshold. Due to lack of official documentation, these were identified based on the data and there seems to be a range of score that acts as a threshold rather than just a cut off score. For instance: for cognitive readiness the threshold is between 597 and 603. Data show that for every year, anything below 597 was deemed as not ready and above 603 was deemed ready. There are no observations between 597 to 603.

3. Number of overall indicators met: The “number of indictors met” variable represents the number of domains for which the child exceeds or meets the set benchmark. The number ranges from 0 to 6 where 0 denotes a student not meeting the benchmark for any domain and 6 denotes a student meeting benchmark in all six domains.

OSPI states that the data from the WaKIDS assessment might serve a wide range of purposes in furthering education (Washington Office of Superintendent of Public Instruction, n.d.). The data help identify students who need extra support in kindergarten and can guide resources, instructional planning for specialists and other district staff. They can also be used to explore students' strengths and needs, potentially impacting decisions about special education or advanced programs, and they can help inform families about their children's progress. Finally, the data could help inform systematic needs for professional development and support conversations with early learning providers, school boards, and community stakeholders.

3.2 Core Student Record System (CSRS) and Core Student Record System (CEDARS) Data

The CSRS and CEDARS datasets we utilize include information on student demographics (e.g. race and gender), classifications (e.g., FRPL, special education, and limited English proficiency status), and various educational outcomes; CSRS includes this information in early school years (2004-05 to 2008-09) and CEDARS for later years (SY 2009/10 to SY 2018/19). Combined, the datasets include annual information for 15 cohorts of students. They also include test outcomes for students in 3rd through 8th grade in math and reading/language arts (ELA).[8] As tests might have changed through the years, we use yearly information on test achievement to standardize these tests within subject, year and grade to be mean zero and standard deviation of one.

For selected years the CEDARS data also have information on other student outcomes. We use the courses data from CEDARS’ Course Catalog file, starting in the 2008-09 school year to track student course-taking and, in particular, the likelihood of taking advanced courses. We look at students’ probability of taking at least one advanced math and at least one advanced ELA course as the literature finds significant benefits from taking even just one advanced course (Avery et al, 2018; Smith et al., 2018; Conger et al., 2021). Beginning in the 2009-10 school year, the data include information on high school grade point average (on a four-point scale) and whether students graduated.

Starting in SY 2011/12 the state collected information on excused and unexcused absences; we use two measures of absenteeism for each student in each year: 1) total absences and 2) unexcused absences only. Unexcused absences are defined as school days in which the student was not present, and a parent or guardian did not inform the school of the reason for the absence. Finally, starting in the SY 2012/13 school year the state began collecting data on student disciplinary actions in schools. Using this, we generated a discipline variable, which includes all reported disciplinary incidences. In addition, we define a ‘Suspension’ subgroup based on whether those incidences resulted in the student missing instructional time in the

[8] The tests in Washington changed several times over the period for which we have data. ELA tests changed in the 2014/15 school year while math tests changes twice over the period, in the 2009/10 and 2014/15 school years.

classroom. Total suspensions are coded as a sum of emergency expulsion, expulsion, in-school suspension, long term suspension, short term suspension, and classroom exclusion, while incidents that result in either a non-suspension disciplinary action or no action are excluded. Previous research suggests that school disciplinary incidences, especially exclusionary practices, in earlier grades are positively correlated with such incidences in later grades and negatively correlated with educational outcomes for students (LiCalsi et al., 2021).

Figure 2 summarizes the student information that is available for different years of CSRS and CEDARS data. The data we use for our analytic samples are anchored by the 15 cohorts for which we have data on student demographics.[9] Student outcomes that we examine include test scores, advanced course-taking, high school GPA and graduation, absences, and disciplinary incidences. **Table 1** provides more details on all the outcome variables' definitions and construction.

3.3 Analytical Samples and Sample Statistics

The availability of data by cohort is determined by the specific outcome examined. As shown in Figure 1, there are two cohorts of students for whom both kindergarten readiness assessments (WaKIDS) and data on demographics and classifications (e.g. LEP status) can be linked to 3rd grade test scores (three years later). These cohorts form what we call the *kindergarten readiness analytic sample*.

There are 15 cohorts of students whose demographics and can be linked to various outcomes: 14 cohorts can be linked to test scores; 11 cohorts to course-taking; 10 cohorts to high school GPA and graduation; 8 cohorts to absences; and 7 cohorts to disciplinary incidences. These cohorts form what we refer to as the *high school outcomes analytic sample*.[10]

Figure 3 reports the change in proportion of the race/ethnicity category in each year of the data, representing students entering and leaving the Washington state public education system. The figure shows the overall trend of changing demographics and increasing diversity in the state of Washington. Because we follow students as they progress through the Washington public school system, we restrict the sample to students that can be linked longitudinally to the outcomes of interest. Consequently, the samples we utilize for the analysis described below look demographically different than the composition of high school students in the state in recent years.

In **Table 2** we describe the analytic sample we utilize for the kindergarten readiness analysis and how it compares to the full sample of students from the same cohorts with KRIs (students who enter or leave the public education system and/or do not appear in 3rd grade with test scores are not included in the analytic sample). Panel A of Table 2 provides sample statistics for the overall kindergarten cohort and Panel B of Table 2 provides the sample statistics for our analytic sample of kindergarteners who have 3rd grade test scores. While the analytic sample size

[9] We use the earliest classification into different groups such as FRPL status, LEP services etc. For cohorts/students without the classification at grade K, we consider the earliest grade-level for which such a classification is available.
[10] The linkages across these cohorts and grades for different years for our kindergarten readiness analytical sample and high school outcomes analytical sample are also shown visually in Appendix A2.

is about 12% smaller than the full sample of students with KRIs in each cohort, there is little evidence of demographic differences between the full and restricted (analytic) samples.

Table 3 reports analogous sample statistics for analytic sample used to assess high school outcomes. In Panel A we report the means for the full sample of students from cohorts that could *potentially* be linked from 3rd grade to high school and in Panel B we report the means for the students in those cohorts that were longitudinally linked. Given the availability of data over a larger span of years than the for the kindergarten readiness sample, the sample sizes are considerably larger. And here we see greater differences between the unrestricted and restricted samples with Asian and White students more likely to appear in the analytic samples and fewer students with learning disabilities and students who received FRPL in the analytic sample. These findings are not terribly surprising given that student groups that tend to be more at risk for poor academic outcomes also tend to be more mobile (Goldhaber et al., 2022).

One concern with measuring the predictive power of different indicators of progress and success over students' public schooling careers is the possibility of non-random attrition of students from the sample. We consider this for both analytic samples. In **Figure 4** we report the attrition for kindergarten readiness sample. This analysis is based on the two cohorts of students that link with the 3rd grade test outcomes data. The figure shows attrition across race/ethnicity when students proceed from kindergarten to the 3rd grade.[11] The figure suggests no substantial attrition by groups across our sample, though it is only across a short time span.

Similarly, **Figure 5** reports the attrition for our high school outcomes samples. This stacked area graph shows the relative proportion of student race/ethnicity overall as student cohorts move from 3rd grade to 12th grade, which allows us a visual inspection of attrition over time. Given the stability of each demographic proportion over the grade progression, there is little evidence of differential attrition by student demographic from 3rd grade to 12th grade sample. These findings do not rule out the possibility of bias arising from sample attrition, but they do mitigate worry about this issue. Additionally, as Austin et al. (2023) show, the predictive power of early tests are relatively insensitive to a large degrees of unobserved differences between those students whose high school outcomes are observed and those whose outcomes are not observed due to attrition out of state administrative databases.

In **Table 4**, we provide information from the two analytic samples (described above) about differences in outcomes by race/ethnicity for kindergarten (Panel A), 3rd grade (Panel B), and high school (Panel C). For outcomes at each level there are stark differences across student subgroups. For instance, at the kindergarten level, Hispanic students are scoring well below the overall average, whereas Asian and White students are scoring slightly above the average on every readiness domain.[12] Out of the six kindergarten readiness domains, White and Asian students are achieving the most readiness indicators, closely followed by students in the

---

[11] We also look at attrition by FRPL status, disability status, LEP Services as students move from kindergarten to 3rd grade- we do not observe any differential attrition. The logit regression that predicts the probability that a student in a certain group proceeds to 3rd grade suggests that student mobility is more prominent across traditionally disadvantaged groups of population.

[12] The differences noted above are statistically significant. This table presents the non-standardized raw scale scores for context. For other analysis, we use standardized scale scores.

Native/Pacific Islander/Other category and Black students, with Hispanic students meeting the fewest readiness indicators. These early academic differences among student subgroups are still apparent in 3rd grade where we see sizable differences in 3rd grade test achievement across subgroups. For example, the Black-White and Black-Hispanic achievement gaps on both math and reading assessments are statistically significant at more than a half of a standard deviation.

Finally, Table 4 also provides sample statistics for high school GPA, absences, disciplinary actions, and graduation rates by student race/ethnicity. The average high school GPA is 2.58 on a 4-point scale, with Asian and White students averaging significantly higher GPAs than Black and Hispanic students.[13] On average, almost 70% of our sample graduates from high school within four years, and over 72% of the sample graduates within five years, but here too there are significant differences across race/ethnicity subgroups in graduation rates.

4 METHODOLOGY

To answer our research questions and assess links between early grade assessments and later grade outcomes, we use a combination of descriptive analyses of student pathways as well as statistical tools that help us verify that the findings we report reflect relationships that also show up in a multivariate setting. We use different indicators, such as early assessments, test scores, and socio-economic variables, to predict the intermediate outcomes and high school outcomes.

The descriptive findings in the paper are derived using various methods. For all of the outcomes we provide a pathway analysis (Sankey plots) showing the flow of different student subgroups (or students with different baseline levels of readiness or academic achievement) into different outcomes.

Regression Analysis: We verify that the findings we describe from the pathway analysis also show up when we use a regression estimate to compare the magnitude and direction of change in the outcome variables as various control variables change. More specifically, to measure the relationship between the independent variables and the dependent variable we use regression models[14] described generally as:

$$Y_i = \alpha_0 + \beta T_i + \tau_i X_i + \varepsilon_i$$

where Y_i is the outcome variable. This includes variable such as test, scores, chronic absenteeism, etc. T_i is the independent variable of interest, such as the number of KRIs achieved, and β is our coefficient of interest that describes measures the correlation between our outcome variable and the independent variable. X_i is a vector of control variables that includes student demographics such as race/ethnicity, FRPL status, disability status, and LEP Services.

[13] Asian students have around 0.3 point higher GPA compared to White students while White students have around 0.4 point higher GPA compared to Black and Hispanic students. All noted difference are statistically significant.
[14] We also used different model specifications (like *probit* model) and the findings from these models are similar to the findings from the OLS model.

Analysis of Variance and Marginal Effects: We examine the R squared and the ANOVA analysis to determine what proportion of the variation in the outcome variables are predicted by different indicators. We use the adjusted R squared to compare the explanatory variables in terms of the degree to which they explain the variation in the outcome.

Dominance Analysis: We conduct dominance analysis to rank the predictive power of each indicator to predict the outcome of interest. A dominance analysis[15] runs a series of regressions with all possible combinations of the six KRIs and tracks the changes in R-squared values that result from iteratively adding or subtracting the explanatory variables. Then the explanatory power for each of the six indicators is averaged across the series of regressions. This average explanatory power can then be used to construct a weight for each variable to optimize predictive power. We compare the R-squared value of the regressions with different explanatory variables: 1) The number of readiness indicators met (0-6) vs. 2) A weighted number of readiness indicators met ($\sum_{i=1}^{6} weight_{ij} \times I_{i[1,0]}$ for the i^{th} domain indicator and the j^{th} outcome).

Cohort Mobility Analysis: We compare the "academic mobility" of cohorts of students by assessing the extent to which variation in early grade academic performance is still present in later grade test scores. Here transitions in relative academic standing are a measure of academic mobility. To visualize this transition, we divide the early grade distribution of scores into percentile rankings (effectively bins of scores from the first percentile to the 100th percentile). Within each percentile bin, the average later grade percentile rank is calculated. This average rank score is then compared to the early grade rank as a measure of academic mobility. For example, we observe that students who score in the lowest percentiles early in their education tend to score higher in their class distribution several years later - an example of upwards mobility. We conduct this exercise for both our WaKIDS sample and our high school outcomes sample and explore a variety of academic outcomes and heterogeneity for student subgroups.

In addition to within-student academic mobility over time, we also analyze the performance of student cohorts over time for our high school outcomes sample (note again that there are more cohorts for the high school outcomes sample than for the kindergarten readiness sample). This is accomplished by mapping the trends of high school GPA, Grade 10 math standardized math scores, and Grade 10 standardized reading scores for each cohort of students that graduated in the SY 2014/15 – 2018/19. We also combine a cohort analysis with a dominance analysis (described above) to test how the predictive ability of various student indicators changes across cohorts. For each of the outcomes (GPA, 10th grade math scores, and 10th grade reading scores) we run a dominance analysis run for each cohort, and then compare the predictive power of each variable (3rd grade test scores and demographics) over time across groups of students. This analysis helps us to assess the extent to which changes in cohorts or school interventions (instructional methods, policies, etc.) might influence the predictive power of high school outcomes over time.

[15] See Budescu (1993) and Azen and Budescu (2003) for a review on Dominance Analysis.

5 FINDINGS

In this section we present findings about the achievement of students as they progress through Washington public schools. We begin in Section 5.1 with focusing on the kindergarten readiness analytic sample (the sample that tracks cohorts from kindergarten to 3rd grade) by describing differences amongst subgroups in measures of kindergarten readiness. In 5.2, we map these readiness indicators to 3rd grade outcomes, and in 5.3 we explore the extent to which inequity in readiness observed in kindergarten persists to inequity in 3rd grade outcomes, a concept we refer to as "academic mobility." We conclude the focus on the kindergarten analytic sample in Section 5.4 by assessing the degree to which the KRIs might be used in different ways (reweighted according to the dominance analysis described in Section 4) to increase their predictive power.

In Section 5.5 we turn to the high school outcomes sample (the sample that tracks cohorts from 3rd grade to high school) and begin by showing variation across student subgroups in 3rd grade test scores. In Section 5.6 we map these test scores, along with demographic information about students to high school outcomes, and in Section 5.7 we assess the degree of academic mobility from 3rd grade to high school and whether it has changed across the cohorts in our sample.

5.1 VARIATION ACROSS STUDENTS IN KINDERGARTEN READINESS INDICATORS

As described above, within several months of students entering kindergarten, their teachers assess their readiness for the grade across six different domains: Math, Literacy, Social-Emotional, Physical, Cognitive, and Language. Teachers rank each student on a scale that varies by domain (see Figure 1), and then based on whether each domain score is above or below a pre-determined threshold, the student is marked as kindergarten ready in that domain. Therefore, a student could meet between zero to six of the six total KRIs. We are unaware of any statewide intervention linked to the measures of kindergarten readiness, but the system was designed "to observe, collect documentation, and level children's knowledge skills and abilities across the 6 areas of development and learning" (Washington Office of Superintendent of Public Instruction, n.d.).

As we show in **Figure 6**, there is significant variation in this measure of kindergarten readiness by student race/ethnicity. For instance, about half of Asian and White students are found to meet all six readiness indicators, while less than one third of Hispanic students do the same.

When exploring the demographic differences by kindergarten readiness domain, several patterns emerge that cannot be captured by examining only the number of readiness indicators met. For instance, as we show from the regression results reported in **Table 5**, compared to their White classmates, Black students score lower (by a statistically significant amount) in the Cognitive, Language, Physical, and Social-Emotional domains, but not in the Literacy or Math domains. Asian students score significantly higher in the Literacy, Math, and Physical domains, but lower in the Language and Cognitive domains compared to their White classmates. Finally, Hispanic students tend to score lower on all domains compared to their Asian and White classmates, which the exception of the Social-Emotional domain on which they tend to outperform students from all other racial/ethnic backgrounds. In short, there is evidence of significant racial/ethnic gaps in most readiness indicators when students enter kindergarten.

Controlling for racial/ethnic background and gender, students who are FRPL eligible, utilize LEP services, or have a learning disability perform lower across all six kindergarten readiness domains compared to their peers who don't fall into these categories. And except for the Math domain, female students on average score higher on all readiness domains compared to their male peers.

5.2 MAPPING KINDERGARTEN READINESS TO 3RD GRADE OUTCOMES

We begin in this subsection by examining the connection between readiness indicators and 3rd grade test scores. As shown in **Figure 7**, KRIs predict test scores in that students who met more indicators were more likely to have a higher level of performance on the average of math and reading test scores; this is visually apparent from looking at the width of the streams that map the number of indicators met (on the left hand side) to the quartile of 3rd grade test scores (on the right hand side). However, the likelihood of students in different test quartiles having met different numbers of indicators in quite different. For instance, nearly 70% of the students who scored in the top quartile of the average 3rd grade math and reading test scores are students that were deemed to have met all 6 indicators in kindergarten (just over 43% of the sample). By contrast, about half the students in the bottom quartile of the average 3rd grade test distribution were deemed to have met three or fewer indicators.[16]

We confirm the relationship between KRIs and 3rd grade test outcomes with regressions that include readiness indicators along with student demographics (**Table 6**). As we can see in the table, the overall number of indicators met predicts significant increase the test scores, and the pattern persists after controlling for other demographic variables. Students meeting all six indicators (relative to no indicators) correlates to over 1 standard deviation gain in both math and reading scores. Even after controlling for demographic variables, the gains persist to 0.9 standard deviations in math and 0.8 standard deviations in reading as shown in columns 2 and 4 of Table 6.

We also look at non-test outcomes such as absences and disciplinary incidences resulting in suspensions. **Figure 8** shows the Sankey plots with the flow of students from number of KRIs met to total absences.[17] Here we see that those students meeting all six readiness indicators are slightly more likely to have fewer absences (i.e., be in the bottom quartile of absences) compared to their peers. The findings for the relationship between demographic and readiness indicators and 3rd grade non-test measures can also be confirmed from the regression results in Table 6. Having a lower number of KRIs met correlates to higher absences and suspensions controlling for other demographic variables. Generally, the more indicators met, the less likely the student is

[16] While not reported, we also confirmed that these patterns are similar when we consider students' distribution on 3rd grade math and reading tests separately.

[17] We do not present the Sankey for disciplinary incidences as the proportion of students with any such incident is very low to conduct such an analysis.

to be absent in the 3rd grade. This trend is present for both unexcused and total absences. The patterns look similar for other non-test outcomes like suspensions and disciplinary incidences.

To assess the explanatory power of student demographics and KRIs we use the R-squared from the regression results in Table 6. We further show in **Figure 9**, about 6% of the variation in test scores is explained by race and race/ethnicity. This rises to about 23% when we also include FRPL, gender, learning disability, and LEP status in the model. Students' scores for the six kindergarten readiness domains alone account for about 20% of the variation we see in test scores. When including only the number of readiness indicators met (0-6), we lose a small amount of explanatory power. This suggests that there may be variation in the degree of importance of each readiness category (e.g., math readiness may be more predictive than physical readiness for example), as well as implying that the raw score in each category may be more useful than the readiness indicator itself. We more formally explore this idea further in section 5.4 below.

The increase in predictive power associated with the use of the individual domains, instead of just using the overall number of indicators met, shows that the domains have differential ability to predict outcomes. We explore this further in **Table 7**, which reports the association between a one standard deviation increase in each of the scores students received in each readiness domain on various 3rd grade outcomes, both when we include student demographics and omit them. The findings for test and non-test outcomes reflect what we had earlier presented on the overall indicators, but also show the relative importance of the individual domains separately for different outcomes (all else equal). The results show that the different domains are differentially predictive of tests and non-test outcomes. Not surprisingly, the math domain is highly predictive of 3rd grade math score and the literacy domain is more predictive of 3rd grade reading tests. A standard deviation increase in student scores in the math domain leads to 0.274 standard deviation increase in 3rd grade math score and a standard deviation increase in literacy leads to 0.296 standard deviation increase in reading scores. Both math and reading assessments predict non-test outcomes in expected ways (e.g., higher readiness assessment scores in these domains are associated with fewer absences and suspensions). However, higher scores in language readiness are correlated with higher absences, which underscores the importance of looking at these domains separately and with relative weights when analyzing various outcomes.

5.3 ACADEMIC MOBILITY FROM KINDERGARTEN TO 3RD GRADE

While students may enter kindergarten with gaps in learning and variation in readiness, their progress through the school system allows them the opportunity to change their path and hopefully catch up with classmates that were more academically ready in kindergarten. In this section we explore the extent to which this variation in readiness identified by the WaKIDS readiness assessment (as shown above in Section 5.1) is still present as measured by 3rd grade test scores. Again, we consider transitions in relative academic standing as a measure of "academic mobility" (Austin et al., 2023).

In **Figure 10** we present smoothed density plots of standardized math scores in kindergarten (the math domain of WaKIDS), 3rd grade (the annual state test), and the change over time. From this we can see that Hispanic students enter kindergarten with, on average, the lowest scores in math compared to their peers. However, by 3rd grade the separation in distribution of students by student subgroups becomes more evident with White and Asian students performing similarly and on average scoring higher than their Hispanic and Black classmates. The student-level change in standardized score over time is explicitly calculated and plotted in the bottom panel of Figure 10. The measure of the average change in standard deviations by race/ethnicity is reported, with the White distribution showing almost no change over time, the Hispanic and Asian distributions showing some small positive change, and the Black distribution showing a decrease over time.

It is important to note that due to the nature of testing, the normal variation in learning patterns, and the randomness in achievement associated with testing or assessments, there will be movement by individual students in the distribution of test scores. As students move along grades, students whose performance are at the extreme end of the spectrum are likely to revert towards the mean. As such, a student testing in 80th percentile in the 1st grade is likely to test lower in the distribution in the 3rd grade, while a student testing in the 20th percentile will likely test higher in the distribution in the 3rd grade. We explore the extent to which this pattern may explain the above changes (Figure 10) in the performance distributions across student subgroups by examining changes from kindergarten to 3rd grade in students' placement in the distribution of achievement by student demographics, what we have termed "academic mobility." In **Figure 10**, we describe the findings from this exercise.

We look at mobility of different student subgroups by performing the nonparametric percentile ranking exercise whereby each student's kindergarten math assessment is put into a percentile rank bin based on where it falls in the overall distribution of math scores. For each of the 100 bins, we then find the average percentile rank for the 3rd grade math score for each racial subgroup (**Figure 11**). These relationships between kindergarten and 3rd grade scores are shown by the four linear fits on the graph. Holding fixed kindergarten math rank, we can immediately see differences in academic mobility over time by student race/ethnicity. For a given percentile rank in kindergarten math, the *vertical distance between the different lines measures the variation in math mobility by race/ethnicity category*. For students that are ranked the same in math readiness in kindergarten, on average the Asian students have the greatest upwards academic mobility, followed by White students, then Hispanic students, and finally Black students with the lowest upwards mobility.

As previously explored in Section 5.1, kindergarten readiness in literacy varied substantially across student racial/ethnic group, and trends present in math readiness did not perfectly correlate to literacy readiness. The changing distributions of reading scores and resulting academic mobility can be observed in **Figure 12**, which presents smoothed density plots of standardized reading scores in kindergarten, 3rd grade, and the change over time. As with math, Hispanic students enter kindergarten with, on average, the lowest scores compared to their peers. In addition, by 3rd grade the White and Asian students perform similarly and on average score higher in reading than their Hispanic and Black classmates. The student-level change in standardized score over time is explicitly calculated and plotted in the bottom panel of Figure 12.

The measure of the average change in standard deviations by race/ethnicity is reported, and the trends are the same as with the shifting math distributions. The distribution of White students shows almost no change over time, the Hispanic and Asian distributions show some small positive change, and the Black distribution shows an average score decrease over time.

We explore this average mobility in literacy by student racial/ethnic group in **Figure 13**.[18] Holding fixed kindergarten literacy rank, we can immediately see differences in literacy mobility over time by student race/ethnicity, though the gaps are not as large as with math mobility. For students that are ranked the same in literacy readiness in kindergarten, on average the Asian students have the greatest upwards mobility, followed closely by White students, then Hispanic students, and finally Black students with the lowest upwards mobility. Interestingly, the race/ethnicity-mobility gap is the smallest for those students in the lowest quartile of the kindergarten literacy distribution and tends to expand the higher the higher the kindergarten rank.

From exploring academic mobility between kindergarten and 3rd grade, several themes emerge. The first is that we observe a greater degree of both upwards and downwards mobility in math compared to reading. Along with this, distinctions in mobility by student racial/ethnic background tend to be greater in math as well. Underrepresented minority students experience less academic mobility in both subjects compared to their White and Asian classmates. This trend exists even controlling for baseline kindergarten readiness, suggesting evidence of growing achievement gaps as students progress through early elementary grades.

5.4 Improving the Predictive Power of Kindergarten Readiness Indicators

As we described above (see Figures 9 and discussion in Section 5.2), there is evidence that the kindergarten readiness indicators have differential ability to predict 3rd grade outcomes. To assess this in more detail, we perform a dominance analysis to see if the predictive power can be increased if these domains were weighed differently than simply adding up the number of domains met.

To test the predictive power of KRIs separately, we regressed 3rd grade math and reading test scores on the score for each KRI category: cognitive, literacy, math, language, physical, and social-emotional. By comparing the regression coefficients, we can better understand which categories have the largest marginal effects on 3rd grade test scores, holding fixed all other variables. This analysis shows that some of the domains are more predictive of 3rd grade math and reading scores than others. Holding fixed the other domains, the math kindergarten readiness category has the largest marginal effects on both math and, perhaps surprisingly, reading test scores, followed by literacy readiness. The cognitive, language, and social-emotional categories have less predictive power, and the physical category does not have a statistically significant impact on either reading or math scores in the 3rd grade.

Using the information from these regressions, we optimally weight the various indicators such that they maximize the predictive power of the 3rd grade math and reading tests. The

[18] The standards defined in the WaKIDS literacy domain (see Appendix A) do not exactly align with the standards measured by the 3rd grade reading test

optimal weights for predictive power for each outcome are shown in **Table 8.** [19] The increase in R^2 associated with using optimally weighted regressions is shown in **Figure 14**; for both math and reading score outcomes, we can see that the inclusion of the weighted indicators increases the explanatory power by about 0.03, a 17% increase.

As with math and reading outcomes, there are substantial differences in the magnitude and direction of the correlations between the six various kindergarten readiness domains and 3rd grade absences. To explore this, we separately model total absences as function of the six different KRI domains and demographic controls. The marginal effects of each domain are shown in **Table 7.**

As was true for both math and reading test scores, the KRI math domain has the strongest correlation with absences, with students who are deemed kindergarten ready in math being less likely to be absent in the 3rd grade. As shown in Table 7, readiness in the literacy domain has the second strongest correlation. Note that these correlations are present even after controlling for student demographics.

In **Figure 15** we report the explanatory power of the combinations of demographic and academic variables for non-test outcomes. These variables explain substantially less of the variation in 3rd grade student absences than they did of 3rd grade test scores. Student race/ethnicity alone explains less than 1% of the variation, and when other student classifications (FRPL, Gender) are added the explanatory power increases to just under 2%.

5.5 VARIATION ACROSS STUDENTS 3RD GRADE TEST SCORES IN THE HIGH SCHOOL SAMPLE

In this section we continue our assessment of students' academic trajectories, focusing on the high school analytic sample and students' pathways progressing from 3rd grade to 12th grade. We begin by setting a baseline by exploring variation in 3rd grade math and reading test scores by race and ethnicity of our high school sample in **Figure 16**.

The trends reported above for our kindergarten analytic sample are also present in the 3rd grade test scores of our significantly larger high school analytic sample.[20] For instance, White students are overrepresented in the 4th quartile (57% of the total sample, but 69% of the top quartile performers), whereas students who are Hispanic are overrepresented in the bottom test quartile (33% of the sample, but 43% of the lowest performers).[21] This analysis is supported by the results in our descriptive regression with the kindergarten readiness sample. Compared to White students, we observe that Black and Hispanic students tend to score lower on both math

---

[19] Table 8 shows the results of a weighted regression of test and non-test outcomes on all six readiness indicators. The table includes the weights from the dominance analysis as well as total change in the R-squared value once the weights are incorporated.

[20] The high school outcomes sample of students overlaps with the kindergarten readiness sample but is significantly larger as it includes 15 cohorts of students as opposed to just 2 for the kindergarten readiness sample (see Figure 2).

[21] These same patterns exist if we break test scores out separately for math and reading/ELA tests.

and reading, while female students perform slightly worse in math and substantially better in reading than male students. Asian students outperform their White classmates in both reading and math. Holding all other factors such as race/ethnicity fixed, students in the FRPL Eligible, Gender, LEP Service, or Learning Disability categories also perform lower on average in both math and reading.[22] As we show in **Table 9,** the patterns we see from the Sankey charts hold up even accounting for the FRPL, LEP, and learning disability status. All else equal, relative to White students, Asian students score about 0.3 percent of a standard deviation better on 3rd grade math tests and 0.2 percent of a standard deviation better on reading tests. By contrast, Black, Hispanic, and students of other races[23] score worse, ranging from 0.1 to 0.3 of a standard deviation in math to 0.09 to 0.26 of a standard deviation in reading.

5.6 MAPPING 3RD GRADE INDICATORS TO HIGH SCHOOL OUTCOMES

In this section we continue our assessment of students' academic trajectories, focusing on students' pathways progressing from 3rd grade to 12th grade. Our analysis here will vary somewhat from the Kindergarten Readiness section in that we have a larger variety of student outcomes to explore (course-taking, GPA, and graduation, in addition to test scores and absences) over more cohorts and years.

We begin by examining the relationship between 3rd grade tests and later test scores. Because there are test scores in multiple subjects (we focus on math and reading/ELA) and in multiple grades (3rd, 8th, and 10th), in **Figure 17**, we summarize the test score relationships from regressions predicting the relationship between 3rd grade tests and other demographics on 8th (Panel A) or 10th grade (Panel B) test achievement. The points on each graph indicate the marginal effects, or conditional relationships, with later test scores.[24]

Students who score higher in reading and math in 3rd grade are clearly more likely to also have higher test scores in both 8th and 10th grades. The relationships between 3rd grade and later test scores are quite strong, though they are stronger in math than reading. Female students are more likely to score higher in high school reading than their male peers, even controlling for 3rd grade test scores. Students who are identified as having learning disabilities and those eligible for free or reduced lunch are more likely to score lower on both reading and math compared to their peers not identified as having a learning disability or being eligible for free and reduced price lunch. This trend was observed in the pathways portion of this paper, and the trend is still present even when controlling for 3rd grade test scores and student race/ethnicity.

---

[22] We show this with Sankey charts in A4-A8 in the Appendix. Though the prevalence of these demographics varies widely in our sample (about 42% of the students are eligible for FRPL, 6% utilize LEP services, and 5.5% have a learning disability), the trends in the pathways are all similar. Students who fall into these categories are much more likely to score in the bottom quartile of 3rd grade math and reading tests.

[23] The "Other" category of race includes students who are American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, and mixed-race students.

[24] The effects displayed are marginal – that is, they are the findings after controlling for all the student indicators noted in the chart. For example, controlling for gender, race/ethnicity, and other student demographics (holding these factors constant), 3rd grade math and reading scores have strong marginal effects on high school scores.

Finally, the marginal effects of race/ethnicity show the relationship between high school scores and student racial/ethnic identity compared to White students (and controlling all other indicators). Black students and Hispanic students are likely to score lower in high school tests than their White peers, while Asian students are likely to score higher. These effects are present even when controlling for 3rd grade test scores, gender, FRPL eligibility, learning disabilities, and LEP service usage. This suggests further evidence that achievement gaps by race/ethnicity persist and even expand between elementary school and high school. *The similarity of Panel A and Panel B indicate that 3rd grade scores and demographics are roughly as predictive of 10th grade scores as they are of 8th grade scores, showing persistence of these factors over time.*

Next, we explore the links between 3rd grade tests and other student classifications and non-test high school outcomes. But before exploring the statistical relationships from regression models, we present Sankey charts depicting the relationships between race/ethnicity and these various outcomes. Figures **18, 19, 20, and 21** show the relationships for absences,[25] high school grade point average (GPA), advanced course taking,[26] and graduation.

Continuing the trend that we observed in the kindergarten sample, White and Asian students are more likely to have better outcomes in test as well as non-test measures such as absences. Figure 20 shows the variation in graduation rates by student race/ethnicity. Overall, slightly more than 70% of the students in our sample graduated high school. The variation seen by race/ethnicity maps relatively closely to the overall proportions of race/ethnicity in our sample. Two categories of note include Hispanic students who are less likely to graduate and Asian and White students who are more likely to graduate high school compared to their peers. There is a similar variation in advanced course taking where Hispanic students are less likely to take any advanced math course while Asian students are more likely to take one. This trend, along with the previous pathways explored, suggest persistence of lower performing outcomes throughout elementary and high school.

The marginal effects of student demographics on graduation rates are mostly consistent with the trends described above.[27] The trends presented in the pathways analysis are supported by regression analysis, with achievement gaps by student race/ethnicity still present even after controlling for early grade performance, gender, and other characteristics. Underrepresented minority (URM) students and those eligible for free or reduced lunch are less likely to graduate, have lower test scores and GPAs, are more likely to be absent, and are more likely to be involved in disciplinary incidents. URM students experience less upwards academic mobility in terms of reading, math, and other academic outcomes, in comparison to other demographic groups. This is explored further in regression results presented in **Tables 10**. After controlling for

[25] The figure shows the pathway for total absences. The results are similar for excused absences only. Disciplinary incidences are not explored for Sankey plots as the proportion of students with such incidences are few low for a meaningful visual representation.
[26] We also looked at advanced algebra courses separately. The variation by race/ethnicity mimics the variation in advanced math courses. We also found out that taking Algebra I earlier is correlated with the probability of taking Algebra II and White and Asian students are more likely to have taken Algebra I earlier than Black and Hispanic students. We also looked at probability of taking advanced ELA and the results are similar.
[27] As previously noted, these effects represent correlations and should not be interpreted as causal estimates.

race/ethnicity and student demographics, early grade academic performance still has a highly predictive effect upon high school achievement, with persistent effects through high school. Taken together, early grade academic performance and student demographics can predict a significant share of the variation in high school grades, test scores, and graduation rates.

Probability of high school graduation is a highly salient and relevant metric of student success, and predictor of later life outcomes.[28] To explore the pathway of early grade performance to high school outcomes, **Figure 22** takes the distribution of averaged reading and math test scores from the third grade and maps them into the probability of high school graduation. Within each quartile, the majority of students graduate high school. However, the probability of graduation increases successively as a student scores higher on 3rd grade reading and math tests.[29] It is interesting to note that among the students that do not graduate,18% were in the top quartile of the 3rd grade test distribution. On the other side of the distribution, more than a third (35%) of the students in the bottom quartile of 3rd grade scores ended up not graduating high school. This trend, along with the previous pathways explored, suggest persistence of lower performing outcomes throughout elementary and high school.

Here the regression estimates of the effects of student demographics on graduation rates and other outcomes are mostly consistent with the trends described above. In particular, as we report in Table 10, achievement gaps by student race/ethnicity are still present even after controlling for early grade performance, gender, and other characteristics. URM students and those eligible for free or reduced lunch are less likely to graduate, have lower test scores and GPAs, are more likely to be absent, and are more likely to be involved in disciplinary incidents. URM students experience less upwards academic mobility in terms of reading, math, and other academic outcomes, in comparison to majority groups. After controlling for race/ethnicity and student demographics, early grade academic performance still has a highly predictive effect upon high school achievement, with persistent effects from 8th to 12th grade. Taken together, early grade academic performance and student demographics can predict a significant share of the variation in high school grades, test scores, and graduation rates.

5.7 EDUCATION MOBILITY FROM 3RD GRADE TO HIGH SCHOOL

A striking finding from the aforementioned analysis is the extent that 3rd grade tests appear to be strongly predictive of high school outcomes. One might expect that students who struggling the 3rd grade receive interventions that help ameliorate academic struggles, and hence, 8th grade test scores should be far more predictive of high school outcomes than 3rd grade tests. We check this more formally by estimating model specifications that predict high school outcomes as a function of 3rd grade test scores, 8th grade test scores, or both (in models that include other covariates), reported in Table 10. Across all high school outcomes, the 8th grade

[28] Students with a high school degree are not only more likely to pursue advanced degrees (Seftor et al., 2009; Perrone et al., 2010), but are more likely to maintain employment, less likely to be incarcerated (Baker and Lang, 2013; Hjalmarsson, 2008) and earn more in their lifetimes than their peers without degrees (Hauser and Daymont, 1977).

[29] We also find regression evidence that, controlling for student race/ethnicity and other characteristics, there are strong positive marginal effects of taking advanced math or advanced ELA on the probability of graduation.

test score is more predictive, as would be expected given its proximity to high school, but the coefficients on 3rd and 8th grade tests are generally not very different and, for some outcomes, the explanatory power of the model only increases marginally when the 8th grade test is used instead of the 3rd grade test. All of this suggests that students are, to at least some degree anchored to their educational achievement levels all the way back in elementary school.

Next, we assess the degree to which there appears to be differential anchoring/academic mobility for different student subgroups. We do this by assessing the academic mobility patterns for different outcomes. Specifically, we look at the various outcomes for different student subgroups for students with different average (across math and reading) percentile ranks on the 3rd grade test score (see **Figures 23-28)**. For outcomes, like GPA and high school tests, we compare where students fall in the 3rd test distribution (x-axis) to where they fall in high school distributions (y-axis). For high school graduation, we look at the probability of graduation for students at each percentile of the average 3rd grade test (since most students graduate, it does not make sense to focus on the percentiles of graduation).

The slope of the trend lines tells us of the strength of the correlation between early grade test scores and high school outcomes. A few trends jump out. First, the slopes all suggest that better 3rd grade tests result in better high school outcomes, i.e., upward sloping for positive outcomes like test scores and GPA, and downward sloping for negative outcomes, like absences and disciplinary actions. But we also observe that Black and Hispanic students have lower academic mobility between 3rd and 12th grades compared to their White and Asian peers (i.e., a Black or Hispanic student with a given 3rd grade test score is less likely to have a good high school outcome and more likely to have a bad outcome, than Asian or White students with the same 3rd grade test score).

Next, we explore whether student demographics and early grade academic achievement have changed in terms of predictive power across cohorts. We do this by estimating statistical models that allow these indicators to have differential predictive power across the different cohorts and assessing how much of the variation in the outcomes they explain for different cohorts. The results of this exercise are reported in **Figures 29-31** which display the R-squared value of each listed variables in a regression of student demographics and 3rd grade test scores on High School GPA, 10th grade math, and 10th grade reading scores. The R-squared value on the y-axis indicates the relative predictive power of each variable independently. Within each variable, the six columns show how that predictive power changes over time from 2015 to 2019.

As shown by Figure 28, 3rd grade test scores and FRLP eligibility are the strongest predictive variables for variation in high school GPA. However, about 80% of the variation in high school GPA is unexplained, i.e., it cannot be accounted for by the test score variables and demographics included in our analysis. The amount of unexplained variation is decreasing slightly over time, while FRLP eligibility (a proxy for family income) has become increasingly more predictive of GPA over time. After these three factors, student race/ethnicity has the next strongest relationship with GPA. The variation over time within racial/ethnic category is relatively small, and the total predictive power of student race/ethnicity is less than 5%. While there are slight changes across cohorts in the share of high school outcomes that are associated with student demographics, there is little evidence that the impact of student race/ethnicity on academic achievement is decreasing over time.

This decomposition was repeated for 10th grade math and reading scores. (Figures 29-30). In these cases, a significant proportion of the variation in the outcome can be explained by the variables in our analysis. That is, student demographics and test scores taken together can account for over 50% of the variation in high school test scores (leaving around 45% of the variation unexplained). In both cases, there is some evidence of math scores becoming slightly more predictive over time. Interestingly, learning disability status seems to be slightly less predictive of high school math scores as time goes on. Conversely, utilization of LEP services has become slightly more predictive of high school reading scores over time.

In sum, our analysis of academic mobility from 3rd grade through 12th grade shows that URM students experience less academic mobility compared to their White and Asian classmates.

6 Discussion and Conclusions

In this report we have documented the extent to which demographics and early measures of academic achievement predict educational progress and success in subsequent grades. Not surprisingly, academic performance of students at one point in their schooling careers is highly predictive of later performance. Indeed, it is arguably surprising just how predictive early measures are as that indicates that educational experiences and interventions appear to do little to alter the trajectories of students' educational progress. Here, however, it is important to be cautious, for we cannot know the counterfactual: we do not know, for instance, if in the absence of educational interventions, students who enter public schools in kindergarten with more limited academic skills might fall significantly further behind their peers. Nor do we know for certain that students who fall into the bottom of the 3rd grade test distribution are actually receiving the interventions they need.

Washington is one of the few states that has a mandated kindergarten assessment, and we find that this assessment strongly predicts 3rd grade outcomes, particularly 3rd grade test scores. As such, this assessment could serve the state purpose of identifying students who need support. But, as we have described above, the different assessments of readiness are differentially predictive of 3rd grade outcomes; the math skills assessment is particularly predictive of 3rd grade academic outcomes. Thus, there is more information that is embedded in assessment system than is apparent from the yes/no indicators of readiness in each domain.

The kindergarten assessments also illustrate the degree to which there are large inequities in skills when students are assessed in kindergarten. Students from historically disadvantaged groups enter kindergarten with significantly fewer readiness standards met. And our exploration of student progression from kindergarten to 3rd grade finds that URM students experience less academic mobility compared to their White and Asian classmates.

To a large extent our analysis of student academic progression from 3rd grade through high school echoes the kindergarten to 3rd grade results. The 3rd grade test assessment is strongly predictive of all high school outcomes, and we see that students eligible for the FRLP and URM students are less likely to have upward academic mobility.

That inequities across student subgroups exist in kindergarten is arguably an indicator of a need for earlier intervention. While the findings on the persistence of gains from pre-kindergarten educational interventions like Head Start are mixed, a recently released study (Horm et al., 2022) finds evidence that a particular early care and education program ("Educare" targeting children 19 months of age and younger) has impacts that persist through to 3rd grade. The findings may also suggest the need for more aggressive intervention during a students' progression through public schools. While the state has made efforts to equalize financial resources across school districts in Washington,[30] there is evidence that there are inequities in qualifications and quality of the teachers that are assigned to students that help explain the gaps in K-12 educational achievement (Goldhaber et al., 2015, 2017, 2022).[31]

In sum, we observed limited academic mobility; students who start out behind generally stay behind. In other words, in most school districts, students are often anchored to a combination of their demographics and early grade achievement (Austin et al., 2022). Thus, the large student subgroup disparities in school readiness present when students enter kindergarten persist through their K-12 schooling and manifest themselves as large disparities in a variety of test and non-test high school outcomes. These subgroup differences in high school outcomes are troubling as they foreshadow postsecondary success in college and the labor market (e.g., Backes et al., 2023; Chetty et al., 2014; Jackson et al., 2020) and hence portend future societal inequities.

[30] Most notably, Washington House Bill 2242 passed in 2017 (in response to the *McCleary* State Supreme Court decision) directed disproportionate state educational funds toward property-poor school districts.
[31] Note that many of these inequities are based on differences in teacher assignments *within* school districts and district-level equalization interventions would likely have little impact on these.

References

Allensworth, E. (2013). The Use of Ninth-Grade Early Warning Indicators to Improve Chicago Schools, *Journal of Education for Students Placed at Risk (JESPAR),* 18(1), 68-83, https://doi.org/10.1080/10824669.2013.745181

Austin, W., Figlio, D., Goldhaber, D., Hanushek, E., Kilbride, T., Koedel, C., Lee, J. S., Luo, J., Özek, U., Parsons, E., Rivkin S., Sass, T., & Strunk, K. O. (2023). Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students. CALDER Working Paper No. 227-0323-3. *National Center for Analysis of Longitudinal Data in Education Research (CALDER).*

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*(2), 129–148. https://doi.org/10.1037/1082-989X.8.2.129

Backes, B., Cowan, C., Goldhaber, D., & Theobald, R. (2023). How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes. CALDER Working Paper No. 270-0423-2

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

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114*(3), 542-551. https://doi.org/10.1037/0033-2909.114.3.542

Cameron, S. V., & Heckman J.J. (2001) The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, 109, 455-499.

Card, D., Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics.*91(11–12), 2158–2184. https://doi.org/10.1016/j.jpubeco.2007.03.006

Cawley, J., Heckman, J., & Vytlacil, E. (2001). Three observations on wages and measured 26 cognitive ability. *Labour economics*, 8(4), 419-442.

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.

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.

Coleman, J. et al. (1966). Equality of educational opportunity. *Washington D. C.: U. S. Government Printing Office.*

Cruz, L., Huxtable-Jester, K., Smentkowski, B., & Springborg, M. (2021). Place-based educational development: What center for teaching and learning spaces look like (and why that matters). *To Improve the Academy: A Journal of Educational Development, 40*(1). https://doi.org/10.3998/tia.960

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 Curtin et al 2012

Curtin, J., Hurwitch, B., & Olson, T. (2012). Development and Use of Early Warning Systems. SLDS Spotlight. *National Center for Education Statistics.*

Domina, T., Penner, A., & Penner, E. (2017). Categorical Inequality: Schools As Sorting Machines. *Annual Review of Sociology*, 43, 311-330. 10.1146/annurev-soc-060116-053354

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year. *American Sociological Review*, *69*(5), 613–635. http://www.jstor.org/stable/3593031

Dumont, H., & Ready, D. D. (2020). Do Schools Reduce or Exacerbate Inequality? How the Associations Between Student Achievement and Achievement Growth Influence Our Understanding of the Role of Schooling. *American Educational Research Journal*, *57*(2), 728–774. https://doi.org/10.3102/0002831219868182

Duncan, G. J., Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future Child*, 15(1), 35-54. 10.1353/foc.2005.0004

Easton, J.Q., Johnson, E., & Sartain, L. (2017). The predictive power of ninth-grade GPA. Chicago, IL: University of Chicago Consortium on School Research.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research, 33*(3), 464-497. https://doi.org/10.1016/j.ssresearch.2003.08.001

Fiester, L., (2010). Early Warning! Why Reading by the End of Third Grade Matters. KIDS COUNT Special Report. Baltimore, MD: Annie E. Casey Foundation.

Figlio, D. N. (2005). Names, Expectations and the Black-White Test Score Gap. NBER Working Papers 11195. *National Bureau of Economic Research (NBER).*

Finkelstein, N.D., & Fong, A.B. (2008). *Course-taking patterns and preparation for postsecondary education in California's public university systems among minority youth.* (Issues & Answers Report, REL 2008–No. 035). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs

Forte, D. (2021). *Why Parents, School Leaders, & Advocates Shouldn't Underestimate the Power of Statewide Assessments.* The Education Trust. https://edtrust.org/the-equity 27 line/why-parents-school-leaders-and-advocates-shouldnt-underestimate-the-power-ofstatewide-assessments/

Fryer R. G. Jr., Levitt S. D. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics*, 86, 447-464.

Fryer, R. G., & Levitt, S. D. (2006). The Black-White Test Score Gap Through Third Grade. *American Law and Economics Review*, *8*(2), 249–281. http://www.jstor.org/stable/42705499

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education, 60*(3), 135–155. https://doi.org/10.2307/2112271

Gamoran, A. (2001). American Schooling and Educational Inequality: A Forecast for the 21st Century. *Sociology of Education*, *74*, 135–153. https://doi.org/10.2307/2673258

Garver, K. (2020). The "why" behind kindergarten entry assessments. *National Institute for Early Education Research*, 1-48.

Goldhaber, D., & Özek, U. (2019). How Much Should We Rely on Student Test Achievement as a Measure of Success? *Educational Researcher*, *48*(7), 479 483. https://doi.org/10.3102/0013189X19874061

Goldhaber, D., Koedel, C., Özek, U., & Parsons, E. (2022). Using Longitudinal Student Mobility to Identify At-Risk Students. *AERA Open*, 8. https://doi.org/10.1177/23328584211071090

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44(5)*, 293–307.

Goldhaber, D., Quince, V., & Theobald, R. (2017). 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.

Goldhaber, D., Theobald, R., & Fumia, D. (2022). The role of teachers and schools in explaining STEM outcome gaps. *Social Science Research*, *105*, 102709.

Goldhaber, D., Theobald, R., & Fumia, D. (2022). The role of teachers and schools in explaining stem outcome gaps. *Social Science Research, 105*, 102709. https://doi.org/10.1016/j.ssresearch.2022.102709

Goldhaber, D., Wolff, M., & Daly, T. (2020). Assessing the Accuracy of Elementary School Test Scores as Predictors of Students' High School Outcomes. CALDER Working Paper 235-0520

Goodvin, R., Rashid, A., & He, L. (2020). *Early Achievers evaluation report two: Prekindergarten quality and child outcomes in kindergarten* (Document Number 20-12-2203). Olympia: Washington State Institute for Public Policy.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*(1), 363-411.Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of policy analysis and management*, *28*(3), 366-393.

Hemphill, F. C., & Vanneman, A. (2011). Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2011-459. *National Center for Education Statistics*.

Herring, W. A., Bassok, D., McGinty, A. S., Miller, L. C., & Wyckoff, J. H. (2022). Racial and Socioeconomic Disparities in the Relationship Between Children's Early Literacy Skills and Third-Grade Outcomes: Lessons From a Kindergarten Readiness Assessment. *Educational Researcher*, *51*(7), 441-450.

Horm, D.M., Jeon, S., Clavijo, M.V., & Acton, M. (2022). Kindergarten through Grade 3 Outcomes Associated with Participation in High-Quality Early Care and Education: A RCT Follow-Up Study. *Educ. Sci.* 12(12), 908. https://doi.org/10.3390/educsci12120908

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, 2(4), 491–508. https://doi.org/10.1257/aeri.20200029

Jacob, B., Dynarski, S., Frank, K., & Schneider, B. (2017). Are expectations alone enough? Estimating the effect of a mandatory college-prep curriculum in Michigan. *Educational Evaluation and Policy Analysis*, *39*(2), 333-360.

Justice, L. M., Koury, A. J., & Logan, J. A. R. (2019). Ohio's Kindergarten Readiness Assessment: Does It Forecast Third-Grade Reading Success?. *Columbus, OH: The Ohio State University*.

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual review of sociology*, *29*(1), 417-442.

Kao, G., & Tienda, M. (1998). Educational aspirations of minority youth. *American Journal of Education*, 106, 349–384.

Kelly, S. (2009). The black-white gap in mathematics course taking. *Sociology of Education*, *82*(1), 47-69.

Koretz D. (2017). *The testing charade: Pretending to make schools better*. Chicago, IL: University of Chicago Press.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity?. *Educational researcher*, 31(1), 3-12.

LiCalsi, C., Ozek, U., & Figlio, D. (2019). The uneven implementation of universal school policies: Maternal education and Florida's mandatory grade retention policy. *Education Finance and Policy*, *14*(3), 383-413.

LiCalsi, C., Osher, D., & Bailey, P. (2021). *An empirical examination of the effects of suspension and suspension severity on behavioral and academic outcomes*. American Institutes for Research, 2021-08.

Lofstrom, M. (2007). Why Are Hispanic and African-American Dropout Rates So High? IZA Working Paper No. 3265

Murnane, R. J., Willett, J. B., & Tyler, J. H. (2000). Who benefits from obtaining a GED? Evidence from high school and beyond. *Review of economics and statistics*, *82*(1), 23-37

Nitardy, C. M., Duke, N. N., Pettingell, S. L., & Borowsky, I. W. (2015). Racial and ethnic disparities in educational achievement and aspirations: Findings from a statewide survey from 1998 to 2010. *Maternal and Child Health Journal*, *19*, 58-66.

Northrop, L. (2017). Breaking the cycle: cumulative disadvantage in literacy. *Reading Research Quarterly*, *52*(4), 391-396.

Perreira, Krista, Kathleen Mullan Harris and Dohon Lee (2006) Making it in America: High School Completion by Immigrant and Native Youth. *Demography*, 43, 511-536.

Reardon, S. F. (2011). The widening socioeconomic status achievement gap: New evidence and possible explanations. *Whither opportunity*, 91-116.

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.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *Aera Open*, *2*(3), 2332858416657343.
Rothstein J., Wozny N. (2013). Permanent income and the Black-White test score gap. *Journal of Human Resources*, 48, 510-544.

Shernoff, D. J., & Schmidt, J. A. (2008). Further evidence of an engagement–achievement paradox among US high school students. *Journal of Youth and Adolescence*, *37*, 564-580.

Sorensen, L. C. (2019). "Big data" in educational administration: An application for predicting school dropout risk. *Educational Administration Quarterly*, 55(3), 404-446.
Strauss, V. (2015, March 1). The important things standardized tests don't measure. *The Washington Post*. https://www.washingtonpost.com/news/answersheet/wp/2015/03/01/the-important-things-standardized-tests-dont-measure/

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. https://doi.org/10.1086/526401

Von Hippel, P. T., & Hamrock, C. (2019). Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science*, 6, 43–80. https://doi.org/10.15195/v6.a3

Von Hippel, P. T., Workman, J., &Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of "Are Schools the Great Equalizer?" *Sociology of Education*, 91(4), 32335. https://doi.org/10.1177/0038040718801760

Weisenfeld, G. G., Garver, K., & Hodges, K. (2020). Federal and state efforts in the implementation of kindergarten entry assessments (2011-2018). *Early Education and Development*, 31(5), 632–652. https://doi.org/10.1080/10409289.2020.1720481

Washington Office of Superintendent of Public Instruction. (n.d.). *WaKIDS whole-child assessment. OSPI*. Retrieved November 27, 2023, from https://ospi.k12.wa.us/student-success/testing/state-testing/washington-kindergarten-inventory-developing-skills-wakids/wakids-whole-child-assessment

Welsh, R. O., & Little, S. (2018). The School Discipline Dilemma: A Comprehensive Review of Disparities and Alternative Approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

Figures and Tables

Figure 1: Distribution of Scores in WaKIDS assessment

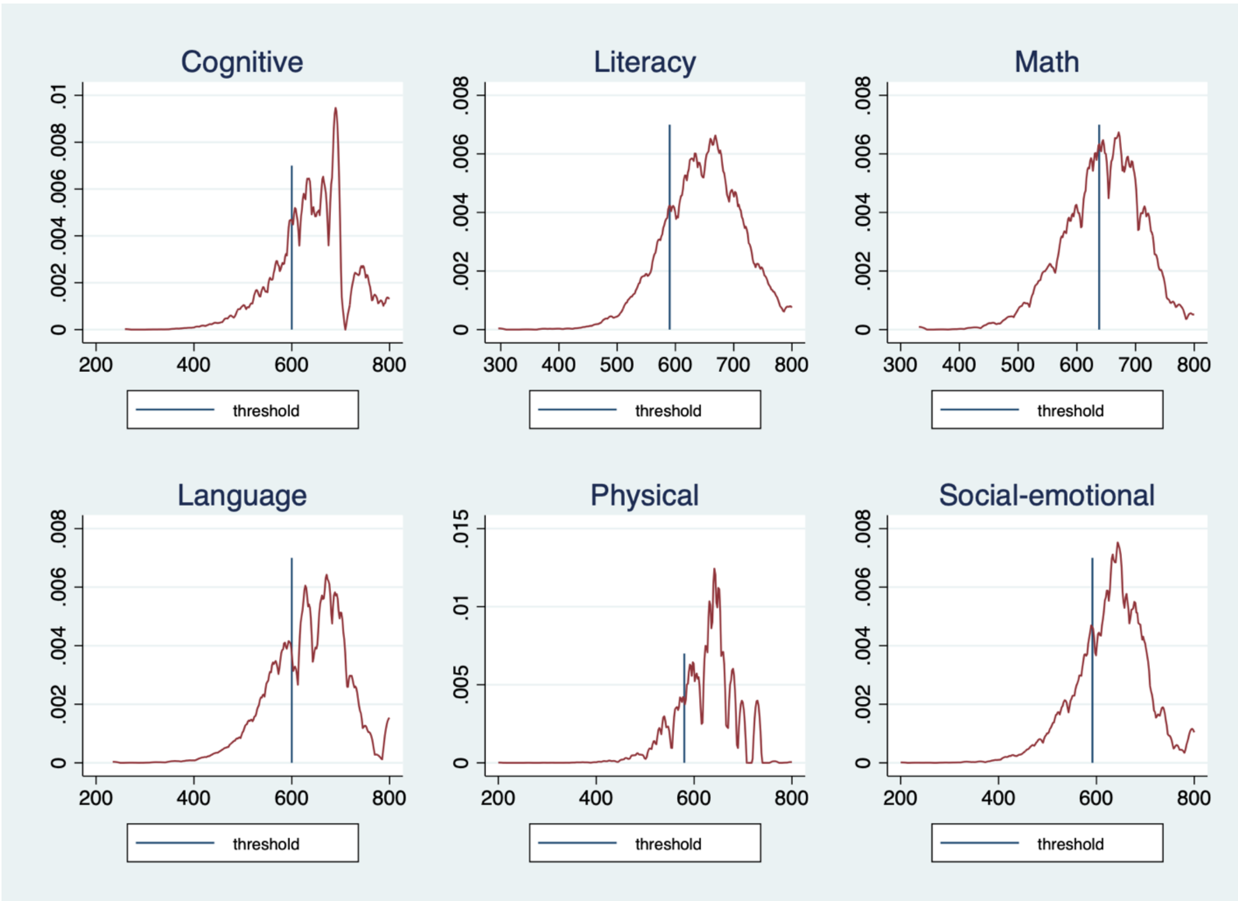
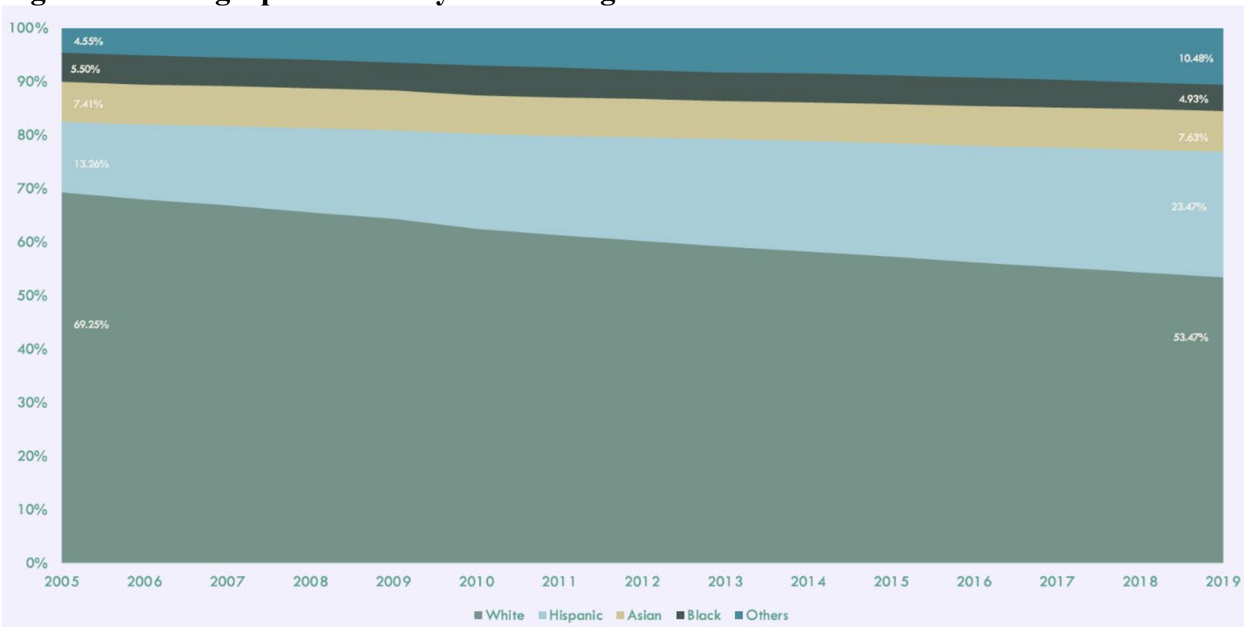


Figure 2: Data Structure Over Time

	Indicators and Outcomes / School Year	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
Indicators	Demographics and Student Classification	15 cohorts of student's data can be linked to 14 cohort of test data, 8 cohort of absences data, 7 cohorts of disciplinary incidences, 11 cohort of advanced course taking, and 10 cohorts of HS GPA and graduation														
Indicators	Kindergarten Readiness Indicators											2 cohorts of WA Kids' KRA data can be linked to 3rd Grade Scores				
Indicators / Outcomes	Test Scores		3rd grade Standardized Math and Reading test scores can be linked to 1 cohort of students' HS GPA and Graduation and Fifth grade scores can be linked to 3 cohorts of students' HS GPA and Graduation													
Outcomes	Course Taking					Number of Advanced courses and probability of taking advanced courses through middle and high school										
Outcomes	HS GPA						High School GPA on a 4.0 scale									
Outcomes	HS Graduation						Probability of HS Graduation, Graduation within 4 years, Graduation within 5 years									
Outcomes	Absences								Full time and part time excused and unexcused absences; chronic absenteeism defined as greater than 18 absent days; pathways from grade 5 to grade 10 can be linked for 2 cohort							
Outcomes	Disciplinary Incidences									Expulsion and suspensions as well as an index of total disciplinary actions; pathways from grade 5 to grade 10 can be linked to 1 cohort						

Our sample can be divided into two main groups: a kindergarten to 3rd grade sample, and a 3rd grade to 12th grade sample. The size of each of these groups varies according to the outcome data of interest and the year being studied. The sample contains 15 cohorts of students ranging from the SY 2004/05 school year to SY 2018/19 school year, with demographic data for all 15 cohorts. KRIs are utilized as explanatory, or independent variables, while 3rd grade test scores are utilized as both an outcome variable as well as an explanatory variable. Advanced courses pertain to math courses that have the following designations: Honors, AP, IB, or College-level.

Figure 3: Demographic Diversity in Washington K-12 Education Over Time

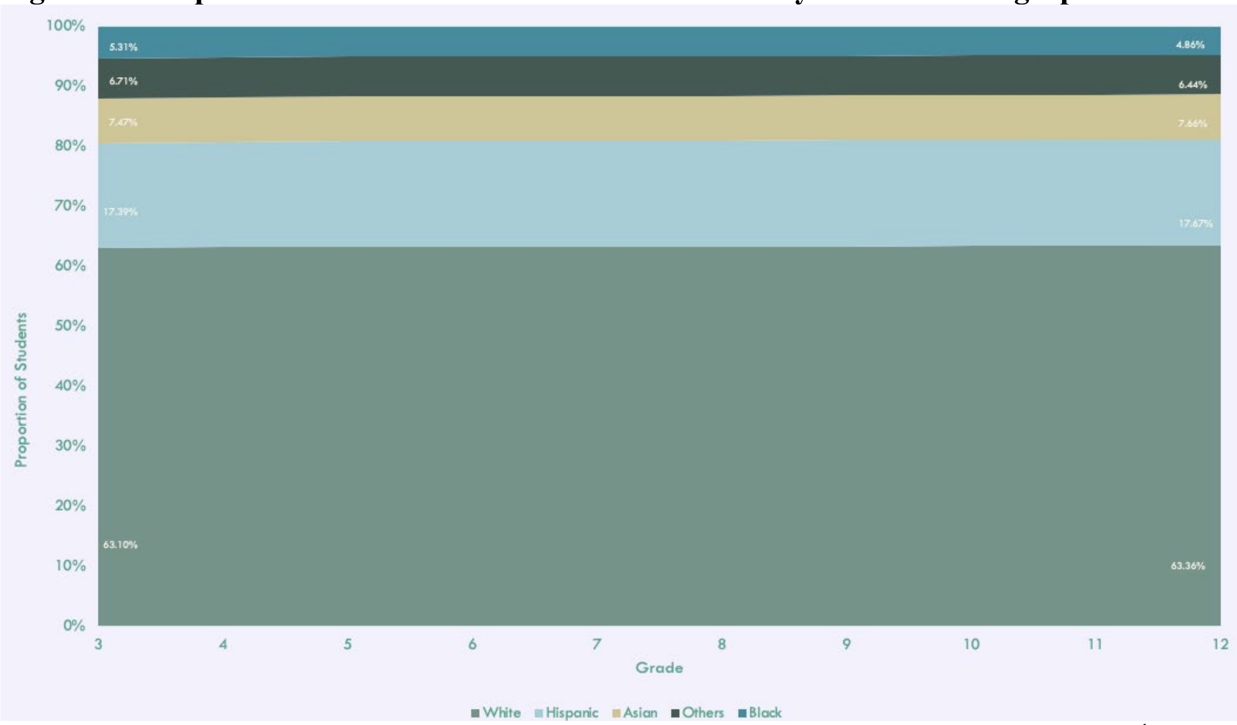


This stacked area graph shows the relative proportion of student race/ethnicity over time in the entire kindergarten to 12th grade sample. Overall, the Washington K-12 student population is becoming more diverse over time, with the proportion of Hispanic students seeing the most growth, followed by the proportion of Asian students. As of 2019, more than 40% of the student population is composed of students identifying as ethnic or racial minorities.

Figure 4: Sample Attrition from K to 3rd Grade by Student Demographic

This stacked area graph shows the relative proportion of student race/ethnicity overall as the cohorts move from kindergarten to 3rd grade. We explore this to see if there is a signficant proportion of the students who drop out of the sample as they progress to the next grade, and whether this varies by demographic. Evidence of non-random attrition is a possible sign that our sample is representative of the broader student population, which limits the external validity of our results. Students may drop out of the public school dataset for several reasons, including moving to another state or entering private school. As can be seen above, there is no significant evidence of differential attrition by student demographic in the kindergarten to 3rd grade sample.

Figure 5: Sample Attrition from 3rd Grade to 12th Grade by Student Demographic



This stacked area graph shows the relative proportion of student race/ethnicity overall as student cohorts move from 3rd grade to 12th grade. This allows us a visual inspection of attrition over time. Students may drop out of the public school dataset for several reasons, including moving to another state or entering private school. Given the stability of each demographic proportion over the grade progression, there is no significant evidence of differential attrition by student demographic from 3rd grade to 12th grade sample.

Figure 6: Number of Kindergarten Readiness Indicators Met by Student Demographic

The graphic above shows the variation of number of KRIs met by students of different racial/ethnic groups. The stacked bar graph shows the total proportions of student race/ethnicity in our sample while the pie charts show the breakdown of KRIs met (ranging from 0-6) for each group. The "number of indicators met" variables equals the number of domains in which the kindergarten student was deemed "kindergarten ready." There is significant variation in kindergarten readiness, especially when comparing White or Asian student groups to Black or Hispanic student groups.

Figure 7: Sankey Chart Mapping Kindergarten Readiness Indicators to 3rd Grade Test Score Quartiles

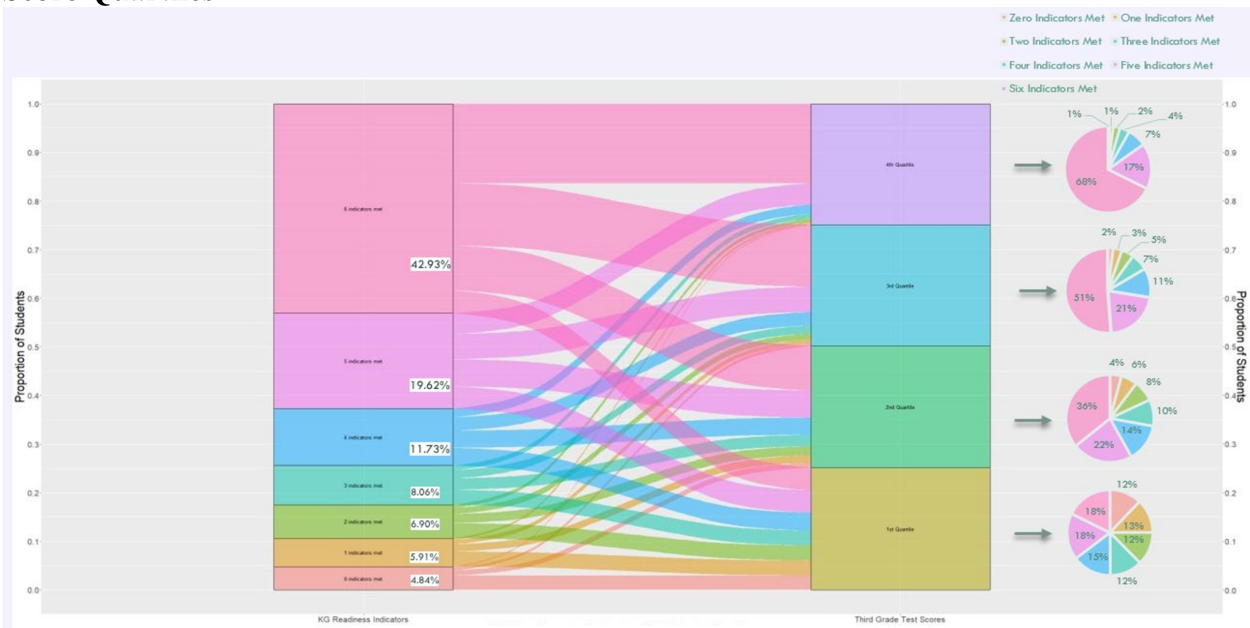


Figure 7 explores the flow of number of KRIs met to the test score distribution in the third grade. The stacked bar chart on the left displays the proportion of students in our sample by the number of indicators met. Almost 75% of the students in the sample were deemed to be kindergarten ready in 4 or more of the domains. The test score variable is calculated by averaging students' math and reading scores on the 3rd grade state test and then dividing those averaged score into quartiles. The pie charts on the right represent the proportion of indicators met in different 3rd grade test quartiles. We can see that those students who have met all six kindergarten readiness domains are most likely to end up in the highest test score quartile, while those students who met two or fewer readiness domains are more likely to fall in the lowest third grade score quartile.

Figure 8: Sankey Chart Mapping Kindergarten Readiness Indicators to 3rd Absences

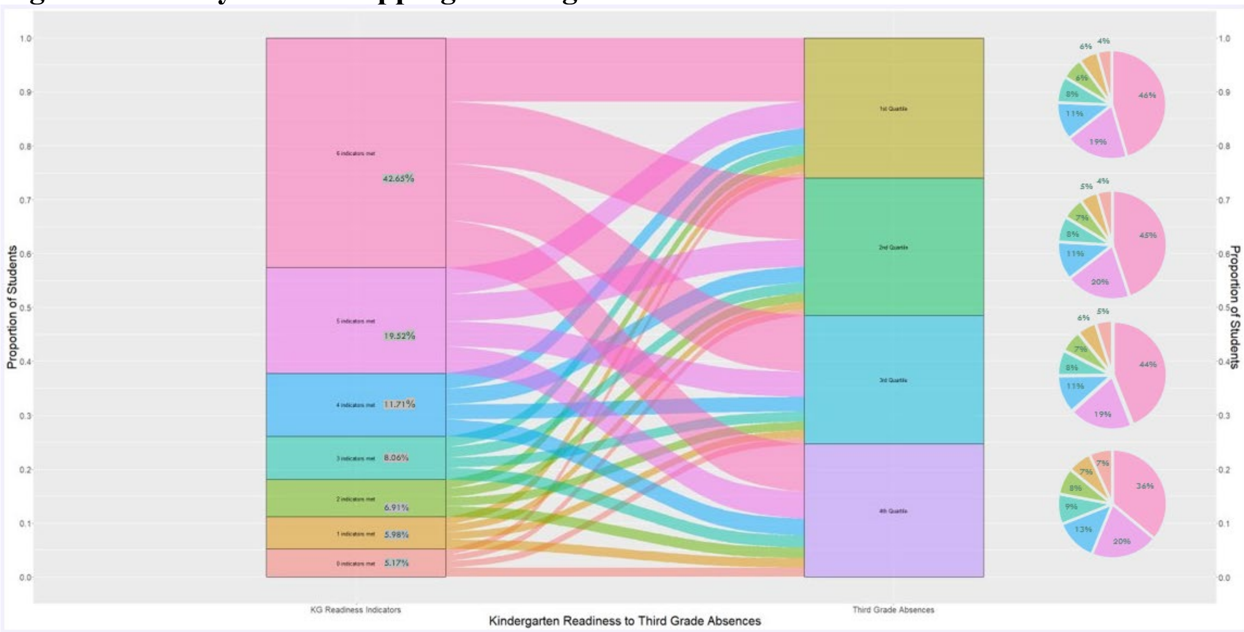


Figure 8 explores the flow of number of KRIs met to the distribution of total absences in the 3rd grade. The stacked bar chart on the left displays the proportion of students in our sample by the number of indicators met. The absence variable includes both excused and unexcused absences, with the distribution divided into quartiles. Note that students with the fewest absences are in the first quartile, while those with the most absences are in the fourth quartile. The pie charts on the right represent the proportion of indicators met in different third grade absence quartiles. Students who met all 6 KRIs are slightly more likely to have fewer absences in the third grade, though there is significantly less variation than when exploring test score outcomes.

Figure 9: Explanatory Power of Independent Variables on 3rd Grade Reading and Math Test Scores

Figure 9 explores what proportion of the variation in 3rd grade reading and math test scores can be explained by different independent variables. Explanatory power is measured by R-squared value. While race/ethnicity alone explains the smallest amount of variation in the test scores, adding in other student demographics such as Learning Disability status and use of LEP services covers more than 20% of the variation in scores. The "Number of Readiness Indicators Only" (0-6 readiness categories met) accounts for slightly under 20% of the variation. This explanatory power is increased when the six readiness domains are considered separately (0 or 1, depending on if the standard is met), as shown in the "Readiness Domains Only" bar. This suggests that considering the six readiness domains separately is more useful at predicting later success than simply the number of domains met.

Figure 10: Distribution of Math Scores From Kindergarten to 3rd Grade

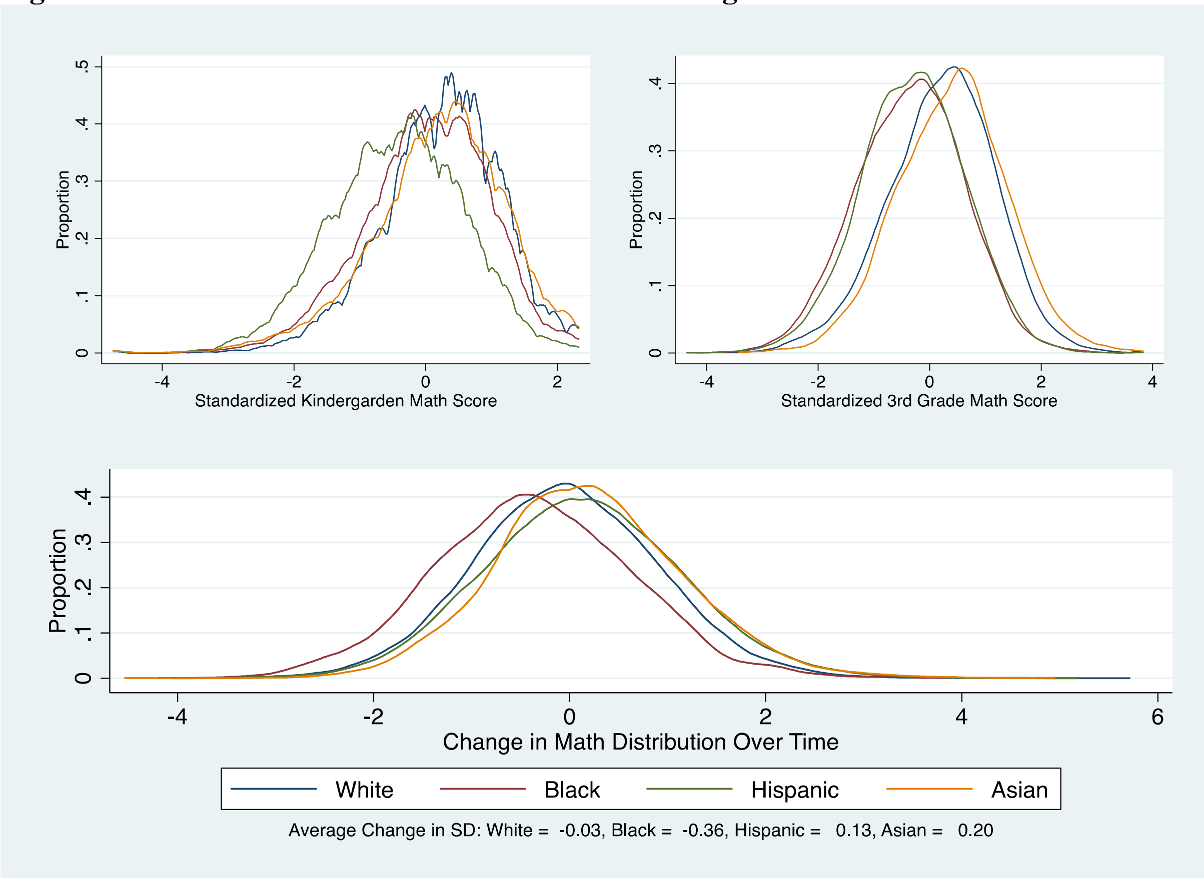
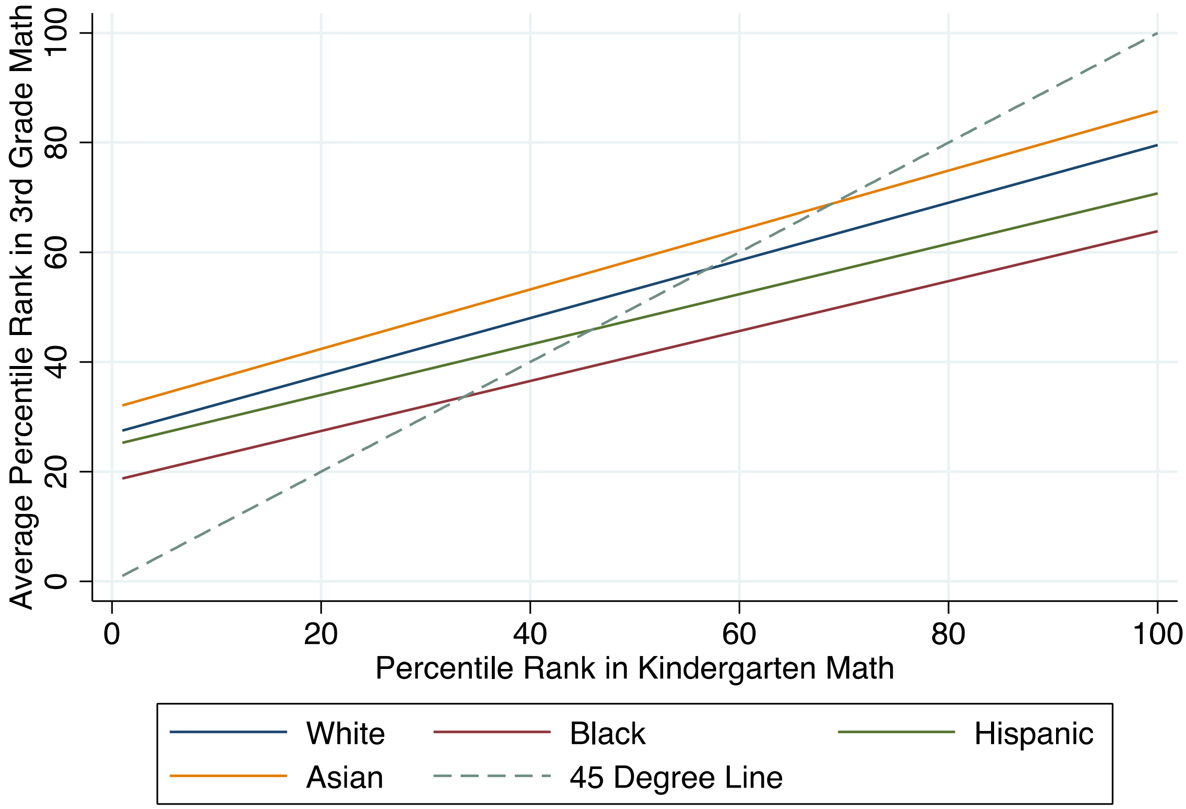


Figure 10 displays the standardized distribution of math test scores by student race/ethnicity in kindergarten, 3rd grade, and the average change over time. The math scores are standardized using the overall group mean and standard deviation. Differences in the distributions by race/ethnicity are present in kindergarten and become more distinct in third grade. The bottom panel shows the distribution of student-level differences in the standardized math score between 3rd grade and kindergarten, categorized by student race/ethnicity. For most student racial/ethnic groups, the median student had no difference in where they scored in the overall distribution. For Black students, the median student scored lower in the overall distribution in the 3rd grade than they did in kindergarten.

Figure 11: Math Score Mobility by Student Ethnicity Between Kindergarten and 3rd Grade



White slope = 0.53, Black slope = 0.46, Hispanic slope = 0.46, Asian slope = 0.54

Figure 11 displays the linear fit of average percentile rank in 3rd grade math test scores by kindergarten percentile rank bin, divided into racial/ethnicity categories. By comparing across the different student races and ethnicities, we can observe that students with the same kindergarten percentile rank in math score have varying levels of mobility by race/ethnicity. For example, Asian students have the highest degree of upwards math score mobility, and the lowest degree of downwards math score mobility. Another way to observe this is by comparing the point at which the various lines cross the 45-degree line. For Asian students, those who scored in the 70th percentile rank or below in kindergarten tend to have upwards mobility in their math score by the 3rd grade. By comparison, for Black students, only those who scored in the 35th percentile rank or below in kindergarten tend to have upwards mobility in their math score by the third grade. In general, we observe more upwards test score mobility for Asian and White students than for Black and Hispanic students. Finally, the slope of the each of the lines (as reported below the legend) indicates the strength of the relationship between kindergarten math percentile rank and Grade 3 math percentile rank.

Figure 12: Distribution of Reading Scores From Kindergarten to 3rd Grade

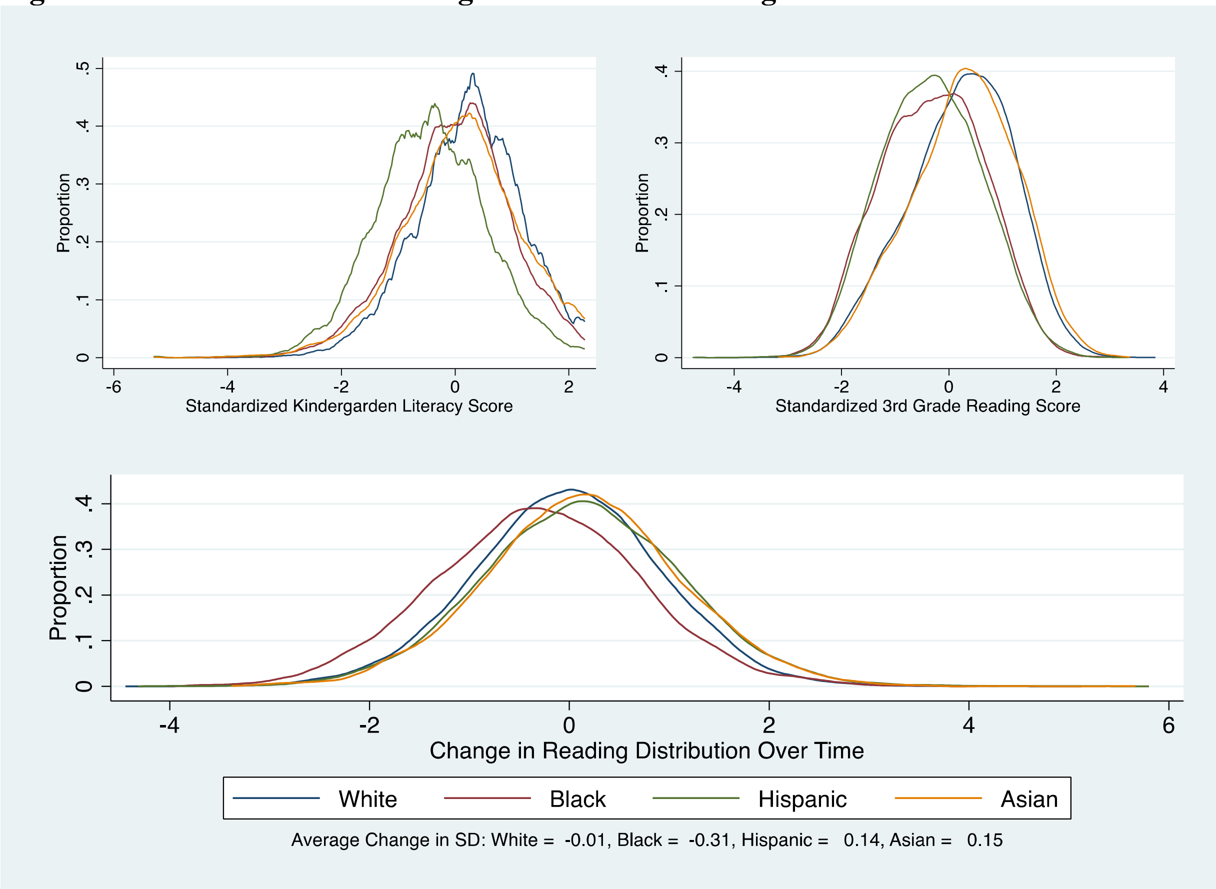


Figure 12 displays the standardized distribution of literacy and reading test scores by student race/ethnicity in kindergarten, 3rd grade, and the average change over time. The literacy and reading scores are standardized using the overall group mean and standard deviation. Differences in the distributions by race/ethnicity are present in kindergarten and become more distinct in 3rd grade. The bottom panel shows the distribution of student-level differences in the standardized literacy score between 3rd grade and kindergarten, categorized by student race/ethnicity. For most student racial/ethnic groups, the median student had no significant difference in where they scored in the overall distribution. For Black students, the median student scored lower in the overall distribution in the 3rd grade than they did in kindergarten.

Figure 13: Reading Score Mobility by Student Ethnicity Between Kindergarten and 3rd Grade

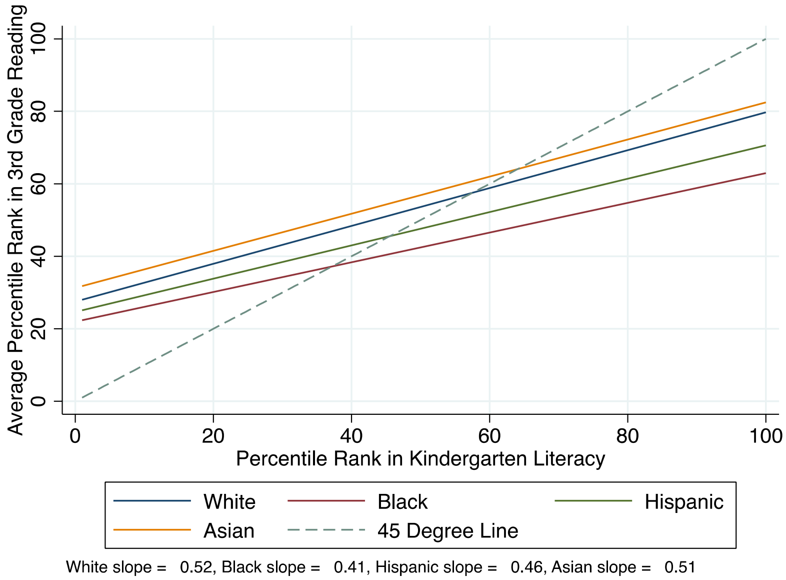


Figure 13 displays the linear fit of average percentile rank in 3rd grade reading test scores by kindergarten literacy percentile rank bin, divided into race/ethnicity categories. By comparing across the different student races/ethnicities, we can observe that students with the same kindergarten percentile rank in literacy have varying levels of mobility by race/ethnicity. For example, Asian students have the highest degree of upwards literacy score mobility, and the lowest degree of downwards literacy score mobility. Another way to observe this is by comparing the point at which the various lines cross the 45-degree line. For Asian students, those who scored in the 65th percentile rank or below in kindergarten tend to have upwards mobility in their reading score by the 3rd grade. By comparison, for Black students, only those who scored in the 38th percentile rank or below in kindergarten tend to have upwards mobility in their reading score by the 3rd grade. In general, we observe more upwards test score mobility for Asian and White students than for Black and Hispanic students. Finally, the slope of the each of the lines (as reported below the legend) indicates the strength of the relationship between kindergarten literacy percentile rank and 3rd reading percentile rank.

Figure 14: Change in R-squared using Weighted Regression

The bar chart above shows how the proportion of explained variation (the R-squared value) of test scores changes when a weighted value is used as the independent variable. The green bar shows that including the number of readiness domains met (0-6) accounts for just under 18% of the variation in 3rd grade test scores. However, some of the readiness domains (such as math and literacy) have a greater correlational relationship with 3rd grade outcomes than other domains. By weighting the six domains based on their relative correlation with the test outcome, the 'Weighted Number of Readiness Indicators' variable accounts for over 20% of the variation we see in 3rd grade test scores, and thus has greater predictive power.

Figure 15: Explanatory Power of Independent Variables on 3rd Grade Absences

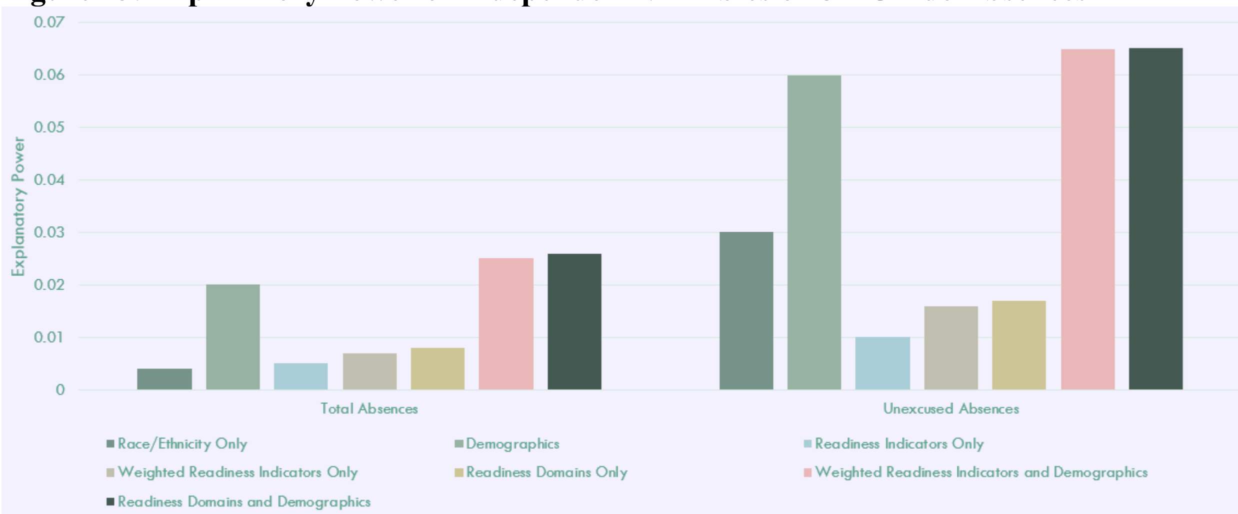
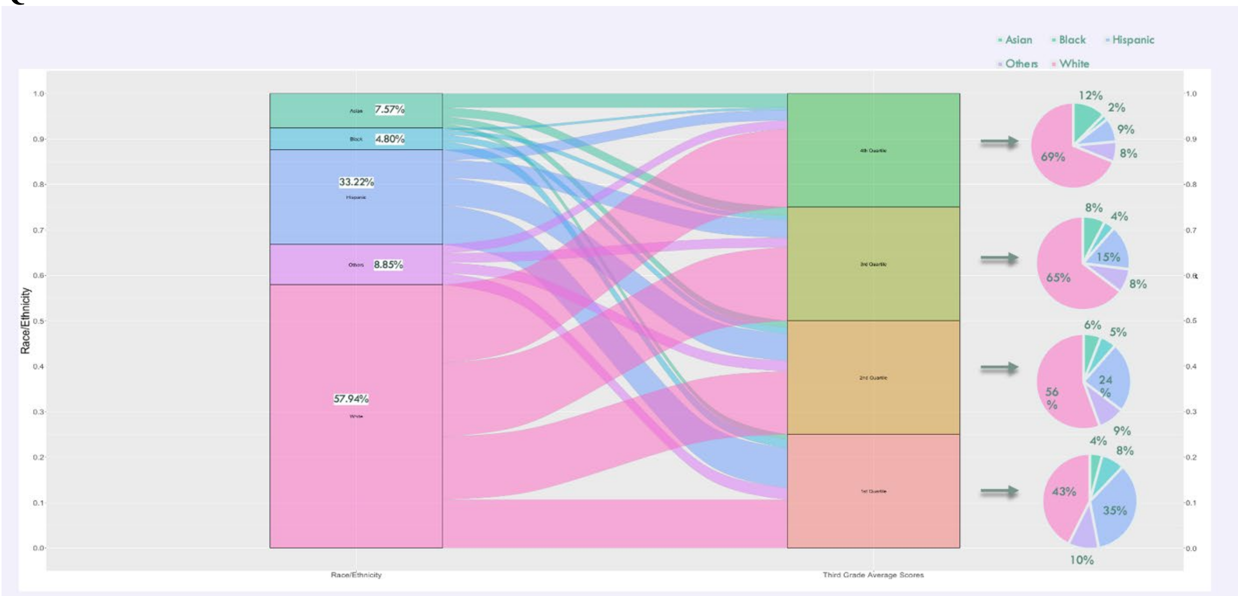


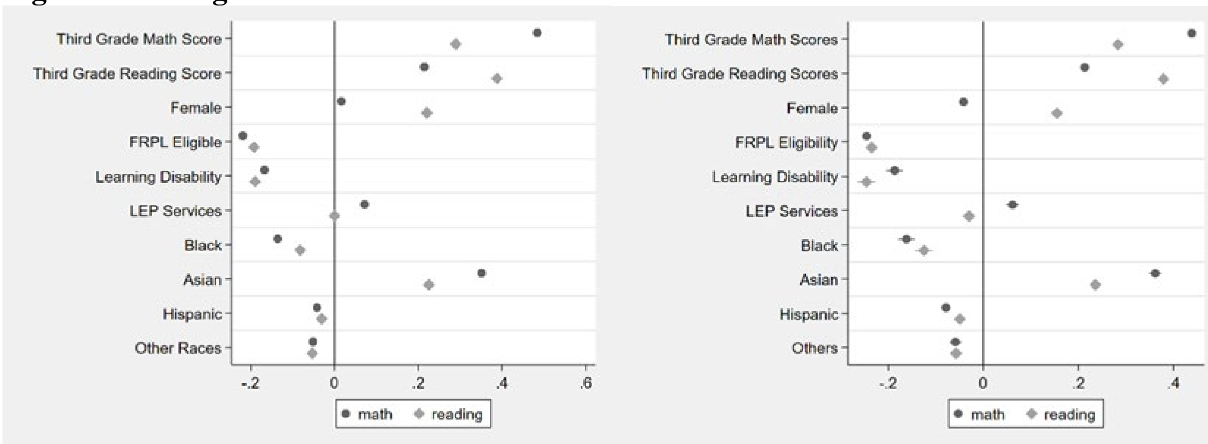
Figure 15 explores what proportion of the variation in 3rd grade total absences and unexcused absences can be explained by different independent variables. Explanatory power is measured by R-squared value. Absences are modelled as a function of demographics, readiness indicators, and/or readiness domains. We can see that the variation in absences is better explained by a combination of KRI domains and the demographic controls. However, these variables aren't as predictive of the absences as they were of of test scores. Additionally, weighting the readiness domains does little to increase the explanatory power. This is likely because there is less variation in the relative correlation of the different readiness domains and absences.

Figure 16: Sankey Chart Mapping Student Race/Ethnicity to 3rd Grade Test Score Quartiles



This Sankey chart shows the variation in student race/ethnicity by 3rd grade test scores, which is an average of math and reading scores. The pie charts represent the proportion of each racial/ethnic group in the four different quartiles. Note that the 4th quartile indicates that the average test score is in the top 25% of the sample, while the 1st quartile indicates that the average test score is in the bottom 25% of the sample. By comparing the overall percentage of the group represented in the sample (number on the left), with the percentage represented in each quartile (number on the pie chart), we can infer whether a student from a particular racial/ethnic group will be more or less likely to progress to a particular quartile compared to his or her peers in another racial/ethnic group (all else being equal).

Figure 17: Marginal Effects of 3rd Grade Scores on 8th Grade and 10th Grade Scores



Panel A: 3rd Grade to 8th Grade scores *Panel B: 3rd Grade to 10th Grade scores*

Figure 17 shows the regression results of 3rd grade math and reading scores as well as other student characteristics on 8th grade (Panel A) and 10th grade (Panel B) test scores. The x-axis is the marginal effect from the corresponding regression. The direction and the magnitude of the marginal effects are consistent across both Grade 8 and Grade 10 outcomes. Controlling for the above characteristics, 3rd grade math scores are highly positively correlated with both 8th grade and 10th grade reading and math scores. It is worthwhile to note that almost of the marginal effects present in 8th grade are persistent and maintain their magnitude through to 10th grade.

Figure 18: High School Pathways: Absences by Race/Ethnicity in 5th, 8th, and 10th Grades

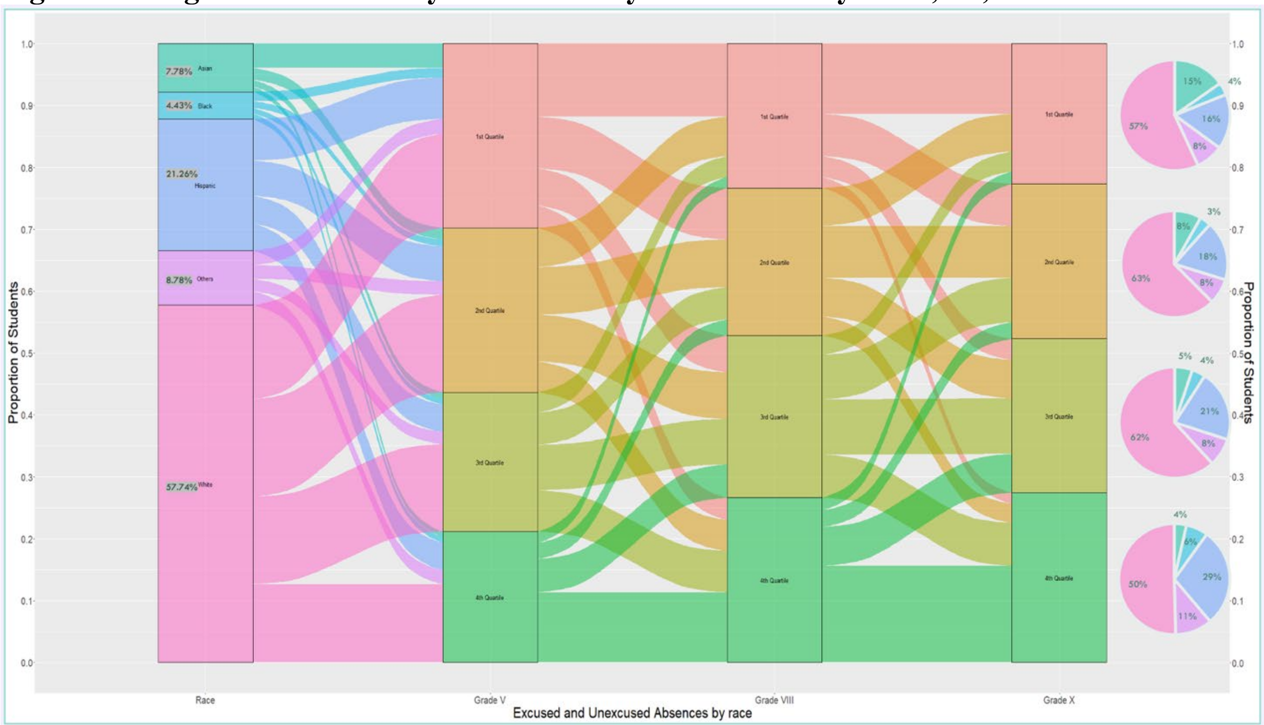


Figure 18 explores the variation in absences and student race/ethnicity in 5th, 8th, and 10th grades. We start with 5th grade here since there are very few absences in 3rd grade. The stacked bar chart on the left displays the proportion of students by race/ethnicity in our sample. The number of absences includes both excused and unexcused absences. This number has been divided into quartiles for each grade, with the 4th quartile representing the 25% of students with the most absences, and the 1st quartile representing the 25% of students with the lowest number of absences. Hispanic students are more likely to have a greater number of absences throughout these grades than their peers while Asian students are less likely to be absent from school compared to their peers. This is shown by the pie charts on the far right, which represent the proportion of students by race/ethnicity in different absences quartiles in 10th grade. Another point of interest is that the majority of students in the fourth quartile in 5th grade remain in the fourth quartile in 8th grade, and 10th grade.

Figure 19: Mapping Student Race/Ethnicity to High School GPA Quartile

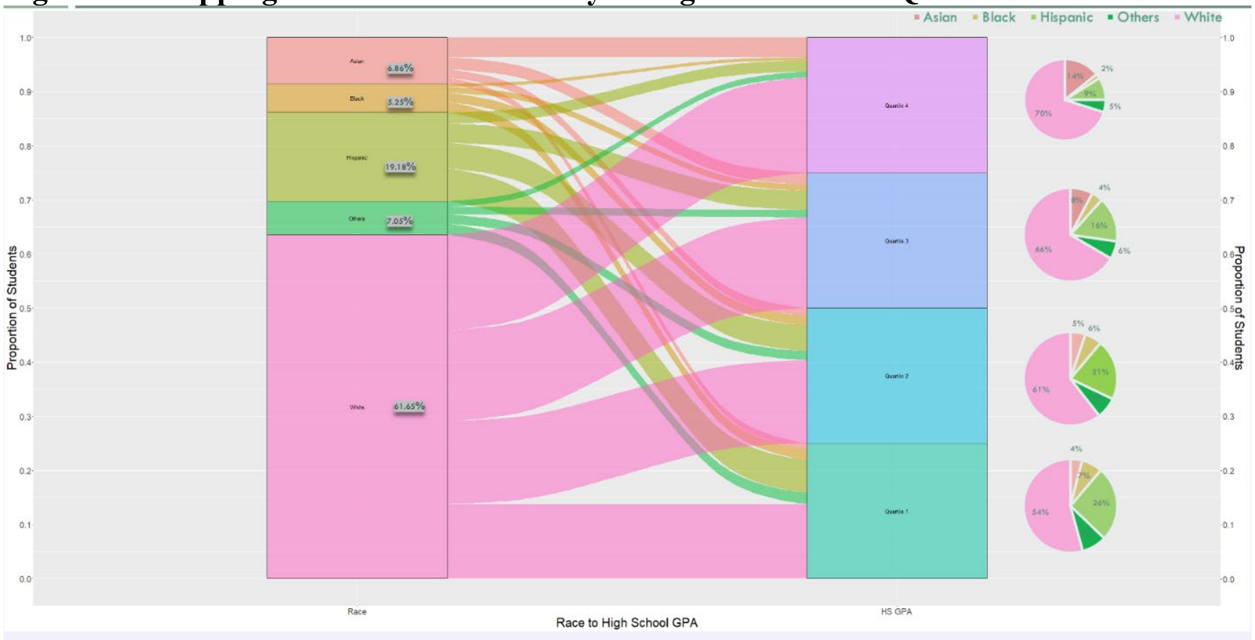


Figure 19 explores the variation in student race/ethnicity by high school Grade Point Average (GPA). The stacked bar chart on the left displays the proportion of students by race/ethnicity in our sample. The high school GPA variable has been divided into quartiles, with the 4th quartile representing the 25% of students with the highest GPAs, and the 1st quartile represents the 25% of students with the lowest GPAs. Hispanic students are more likely have a GPA in the 1st or 2nd quartile, while Asian students are more likely to have a GPA in 3rd or 4th quartile.

Figure 20: Advanced Math Course Taking by Student Race/Ethnicity

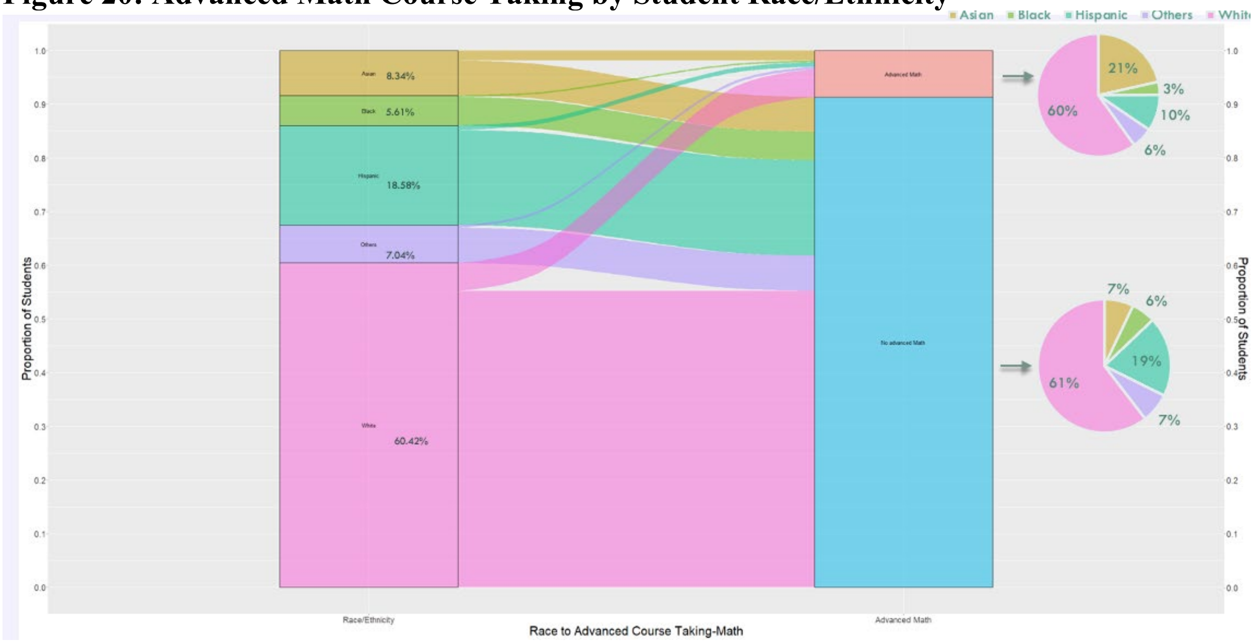


Figure 20 explores the variation in student race/ethnicity by advanced math course-taking. The stacked bar chart on the left displays the proportion of students by race/ethnicity in our sample. The stacked bar chart on the right indicates whether the student took advanced math courses at some point during their time at school or did not take any such course. Note that advanced courses are defined as those courses in the following categories: AP, IB, Honors, or College in High School. The pie charts are the far right show the proportions of student race/ethnicity within the two categories. Overall, about 90% of the students in our sample did not take advanced math courses. The figure above shows that Black and Hispanic students are underrepresented in advanced math courses while Asian students are overrepresented (8.34% in the overall sample, while making up 21% of those taking advanced math).

Figure 21: High School Graduation Rates by Student Race/Ethnicity

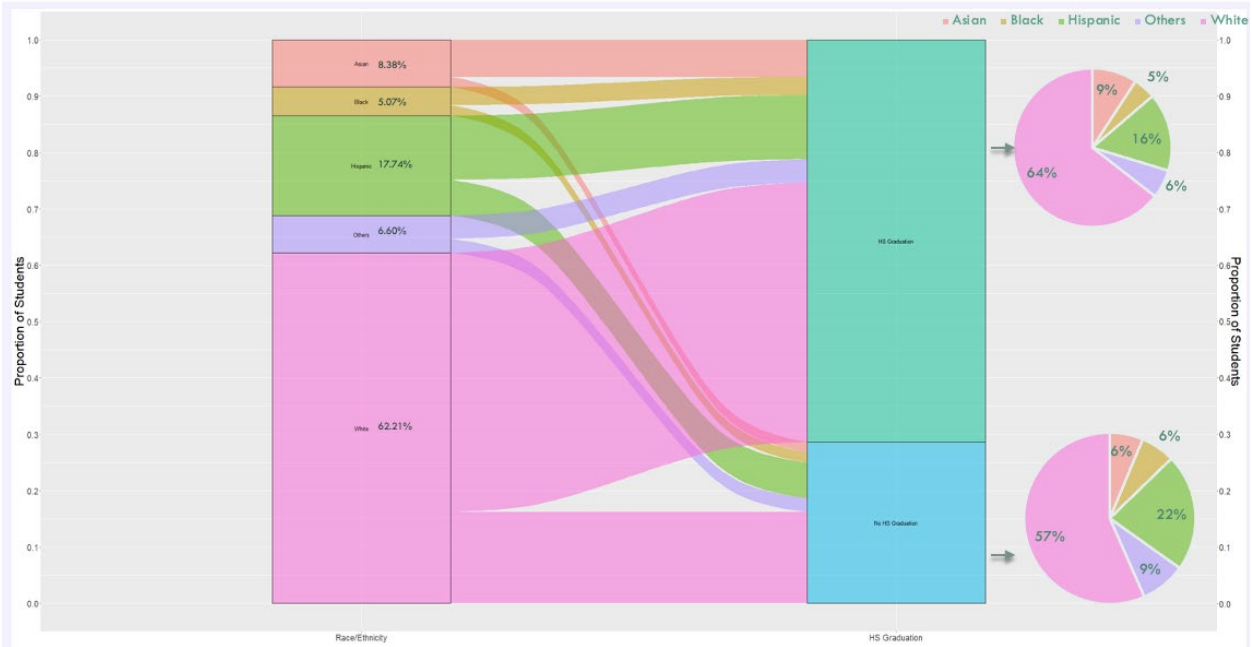


Figure 21 explores the variation in student high school graduation rates by race/ethnicity. The stacked bar chart on the left displays the proportion of students by race/ethnicity in this sample. The stacked bar chart on the right indicates whether the student graduated high school, conditional on having reached graduation age. The pie charts are the far right show the proportions of student race/ethnicity within the two categories. Overall, over 70% of the students in this sample graduated high school. The figure above shows that White and Asian students are slightly more likely to graduate compared to their peers, while Hispanic students are slightly less likely to graduate high school.

Figure 22: High School Graduation Rates by3rd Grade Test Scores

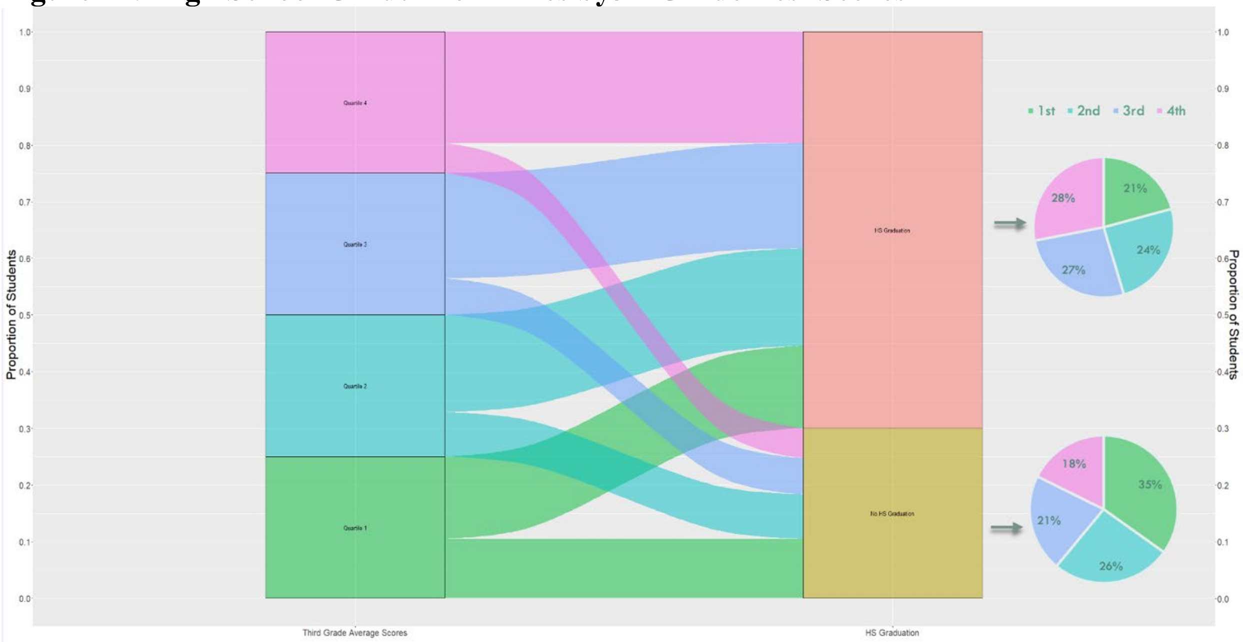


Figure 22 explores the variation in student graduation rates by 3rd grade test scores. The stacked bar chart on the left displays the 4 quartiles of 3rd grade scores. The stacked bar chart on the right indicates whether the student graduated high school, conditional on having reached graduation age. The pie charts are the far right show the proportions of student 3rd grade test score quartiles within the two categories. Overall, over 70% of the students in this sample graduated high school. The figure above shows the persistence of early grade scores as students from the bottom quartiles are more likely to not graduate.

Figure 23: GPA Mobility by Student Race/Ethnicity

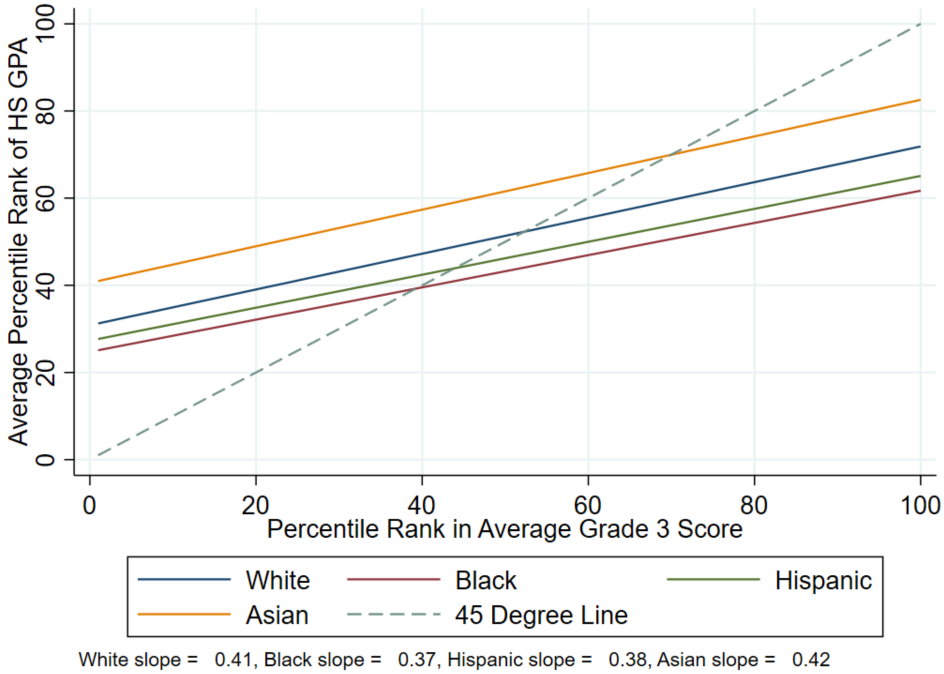


Figure 23 displays the linear fit of the average percentile rank in high school GPA by 3rd grade average math and ELA score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and high school GPA. By comparing across the different student races/ethnicities, we can observe that students with the same 3rd grade percentile rank have varying levels of GPA mobility by race/ethnicity. For example, students that scored in the 60th percentile rank in 3rd grade on average rank between about the 45th percentile and about the 65th percentile in high school GPA, depending upon student race/ethnicity. This shows that achievement gaps by race/ethnicity persist from 3rd grade through high school. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade score percentile rank and GPA percentile rank by race/ethnicity.

Figure 24: 10th Grade Math Score Mobility by Student Race/Ethnicity

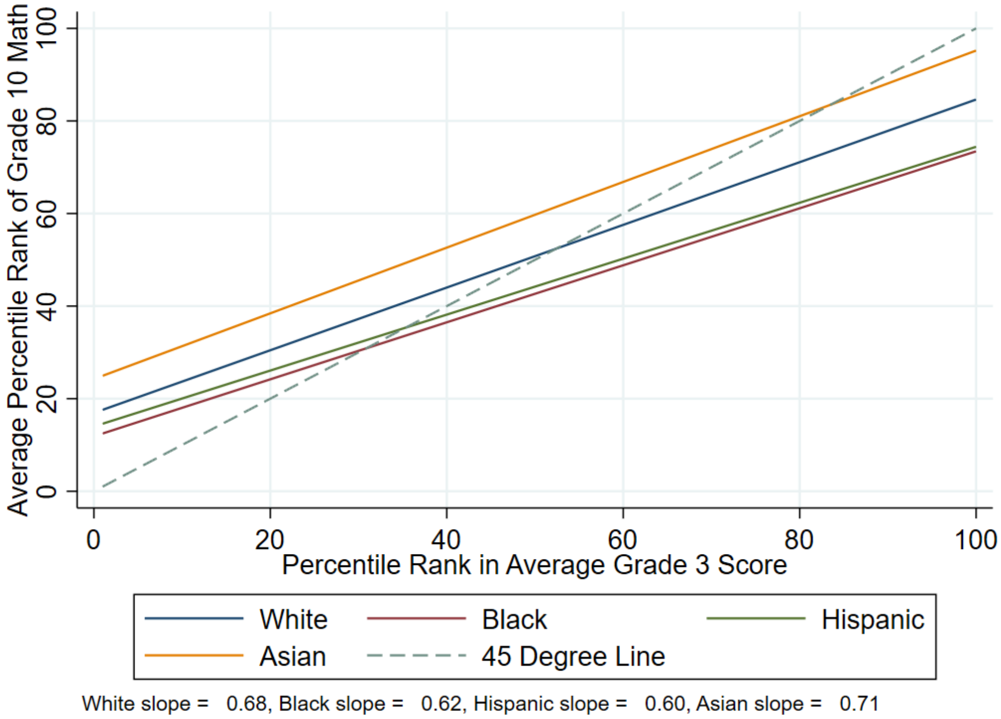


Figure 24 displays the linear fit of the average percentile rank in 10th grade math score by 3rd grade average score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and 10th grade math score. By comparing across the different student races/ethnic groups, we can observe that students with the same 3rd grade percentile rank have varying levels of 10th grade math score mobility by race/ethnicity. For example, students that all scored in the 60th percentile rank in 3rd grade on average rank roughly between the 45th percentile and the 70th percentile in 10th grade math, depending upon student race/ethnicity. This shows that achievement gaps by race/ethnicity persist from 3rd grade through high school. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade score percentile rank and 10th grade math percentile rank by student race/ethnicity.

Figure 25: 10th Reading Score Mobility by Student Race/Ethnicity

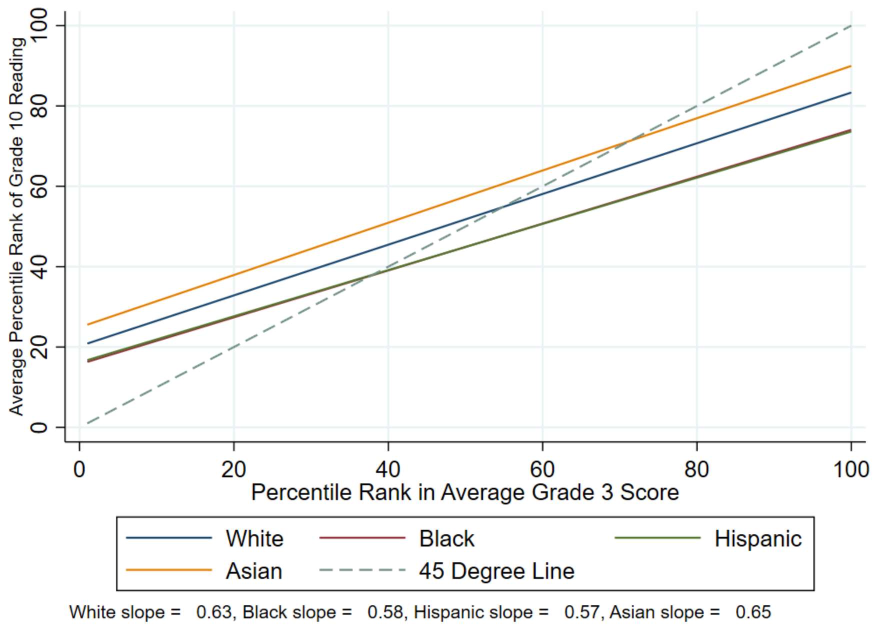
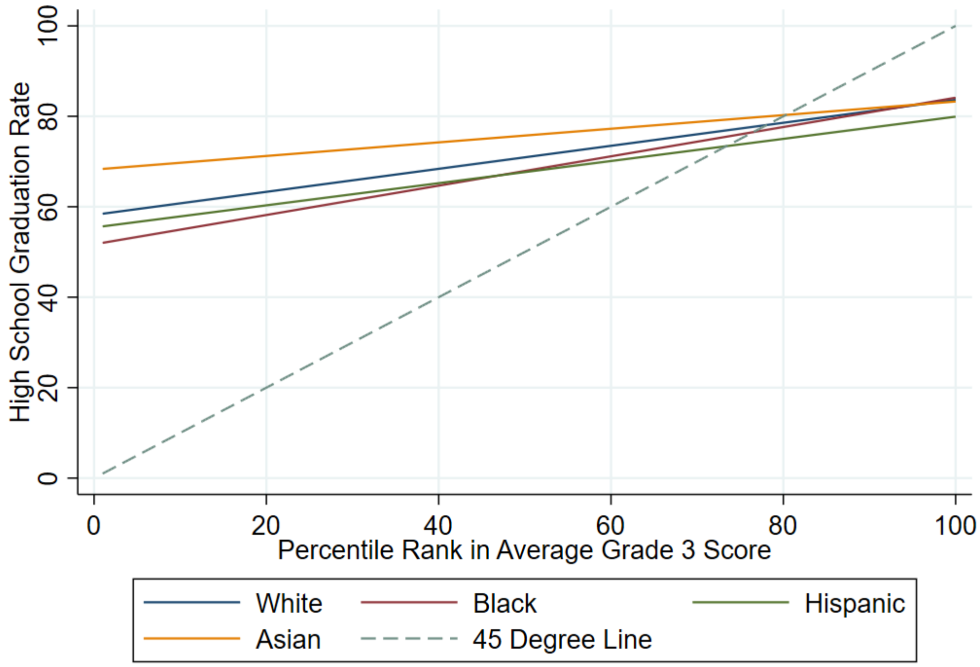


Figure 25 displays the linear fit of average percentile rank in 10th grade reading score by 3rd grade average score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and 10th grade reading score. By comparing across the different student races/ethnicities, we can observe that students with the same 3rd grade percentile rank have varying levels of 10th grade reading score mobility by race/ethnicity. For example, students that all scored in the 60th percentile rank in 3rd grade on average rank between about the 50th percentile and about the 65th percentile in 10th grade reading, depending upon student race/ethnicity. This shows that achievement gaps by race/ethnicity persist from 3rd grade through high school. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade score percentile rank and 10th grade reading percentile rank by student race/ethnicity.

Figure 26: Graduation Rate Mobility by Student Race/Ethnicity



, , ,

Figure 26 displays the linear fit of graduation rate by 3rd grade average score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and graduation rate. By comparing across the different student races/ethnic groups, we can observe that students with the same 3rd grade percentile rank have different levels of graduation rate mobility by race/ethnicity. For example, students that all scored in the 60th percentile rank in 3rd grade have roughly between 70 to 80 percent graduation rate, depending upon student race/ethnicity. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade score percentile rank and graduation rate by student race/ethnicity.

Figure 27: Absences Mobility by Student Race/Ethnicity

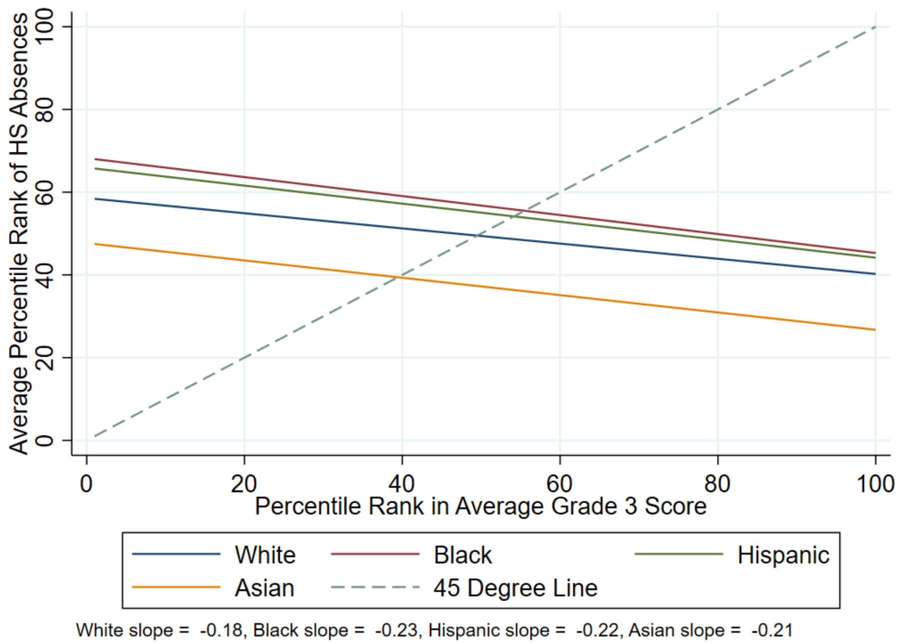


Figure 27 displays the linear of average percentile rank in absences by 3rd grade average score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and absences. By comparing across the different student races/ethnicities, we can observe that students with the same 3rd grade percentile rank have varying levels of absences mobility by race/ethnicity. For example, students that all scored in the 60th percentile rank in 3rd grade on average rank between about the 35th percentile and about the 55th percentile in absences, depending upon student race/ethnicity. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade test score percentile rank and absences percentile rank by student race/ethnicity.

Figure 28: Disciplinary Actions Mobility by Student Race/Ethnicity

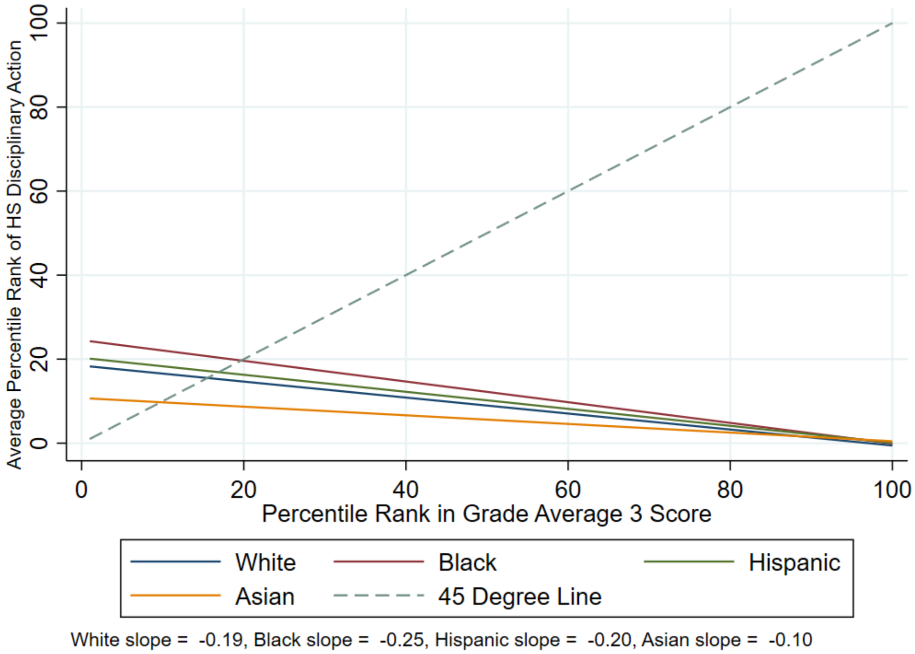
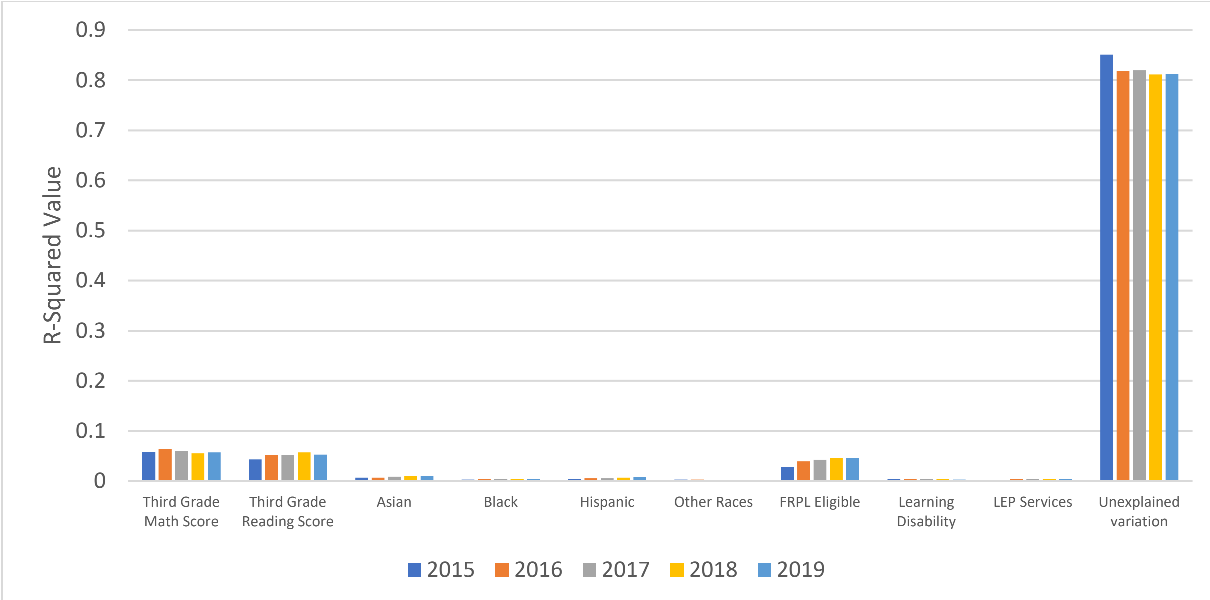


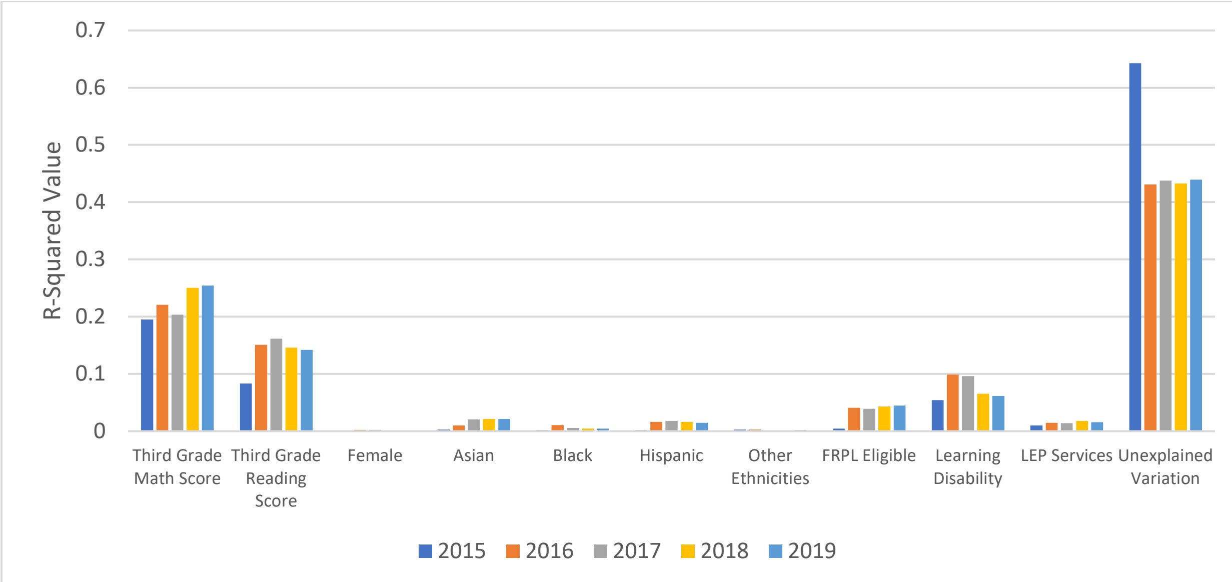
Figure 28 displays the linear fit of average percentile rank in suspensions by 3rd grade average score percentile bin, divided into race/ethnicity categories. The 45-degree line indicates a one-to-one correlation between percentile rank in 3rd grade average scores and average suspensions. By comparing across the different student races/ethnic groups, we can observe that students with the same 3rd grade percentile rank have varying levels of average suspensions mobility by race/ethnicity. For example, students that all scored in the 60th percentile rank in 3rd grade on average rank between the 5th percentile and the 10th percentile in average suspensions, depending upon student race/ethnicity. Finally, the slope of each of the lines (as reported below the legend) indicates the strength of the relationship between average 3rd grade score percentile rank and disciplinary actions percentile rank by student race/ethnicity.

Figure 29: Decomposition of High School GPA - Changes Over Time



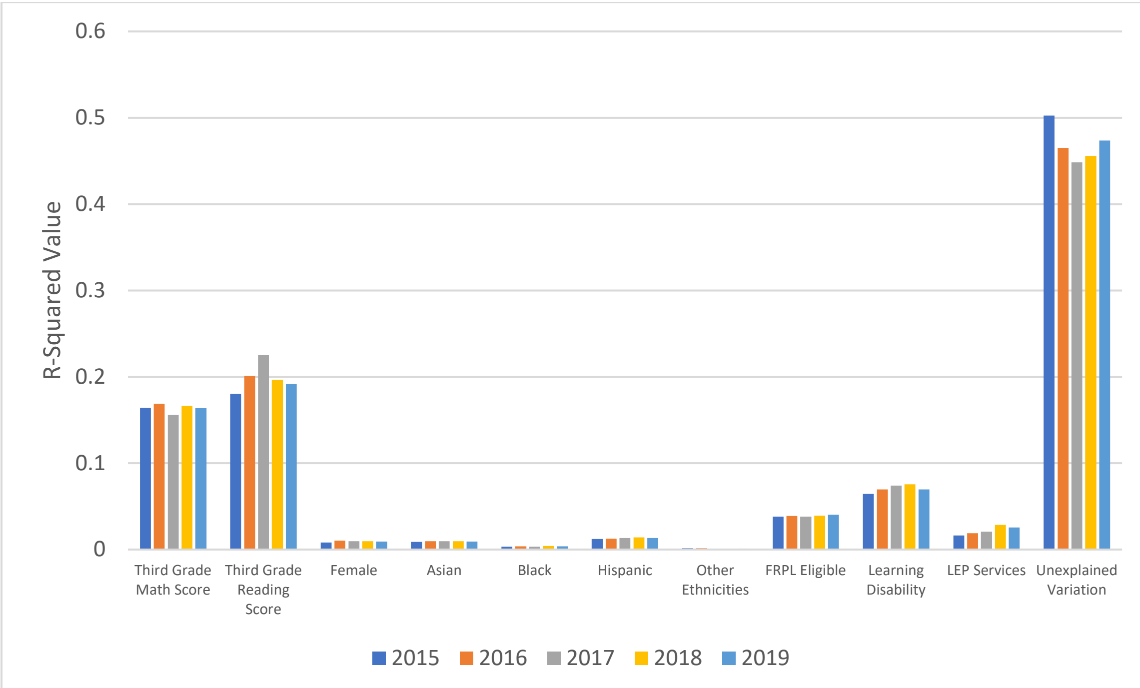
This column chart displays the R-squared value of each listed variables in a regression of student demographics and 3rd grade test scores on high school GPA. The R-squared value on the y-axis indicates the relative predictive power of each variable independently. Within each variable, the six columns show how that predictive power changes over time from 2015 to 2019. Overall about 80% of the variation in high school GPA is unexplained – that is, it cannot be accounted for by the variables and demographics included above, but this decreases over time, moving from about 85% in 2015 to about 81% in 2019. 3rd grade test scores account for about 15% of the variation in GPA, with this number being relatively stable over time. After this, being free/reduced lunch eligible predicts about 5% of the variation in GPA and has become slightly more predictive over time.

Figure 30: Decomposition of 10th Grade Math Scores - Changes Over Time



This column chart displays the R-squared value of each listed variables in a regression of student demographics and 3rd grade test scores on 10th grade math scores. The R-squared value on the y-axis indicates the relative predictive power of each variable independently. Within each variable, the six columns show how that predictive power changes over time from 2015 to 2019. Overall, with the exception of 2015, about 45% of the variation in 10th grade math scores is unexplained – that it, it cannot be accounted for by the variables and demographics included above. The variable with the highest predictive power is the 3rd grade math score, with a slight increase of importance from 2015 to 2019. In contrast, having a learning disability is generally becoming less predictive of math scores in the 10th grade.

Figure 31: Decomposition of 10th Grade Reading Scores - Changes Over Time



This column chart displays the R-squared value of each listed variables in a regression of student demographics and 3rd grade test scores on 10th grade reading scores. The R-squared value on the y-axis indicates the relative predictive power of each variable independently. Within each variable, the six columns show how that predictive power changes over time from 2015 to 2019. Overall, about 45- 50% of the variation in 10th grade reading scores is unexplained – that it, it cannot be accounted for by the variables and demographics included above. The variable with the highest predictive power is the 3rd grade reading score, followed by the 3rd grade math score, which relatively stable over time.

Table 1: Outcome variables definition and construction

Outcomes	Elements	Definition	Notes
Test Scores	Math Score	Math test scores standardized (to mean 0 and 1 standard deviation) by grade and year	Given the changes in tests, we use standardized test scores
	Reading Score	Reading test scores standardized (to mean 0 and 1 standard deviation) by grade and year	
Absences	Unexcused absences	Number of unexcused absences in a year	We also conducted PCA to generate a "non-test" outcome including these variables. Most of the analysis is reported separately but all the analysis has also been conducted for this overall non-test measure.
	Total absences	Total of excused and unexcused absences in a year	
Disciplinary incidences	Suspensions	Sum of emergency expulsion, expulsion, in-school suspension, long term suspension, short term suspension, and classroom exclusion	
	All disciplinary incidences	Sum of suspensions (as defined above) and non-suspension disciplinary incidences	
Advanced course taking	Advanced math	Probability of having taken any of the following advanced math courses from grade 8-12: Advanced placement; College in HS; International baccalaureate; Running start; Cambridge scholars program	In some cases, we also look at the number of advanced courses taken
	Advanced reading (ELA)	Probability of having taken any of the following advanced reading courses from grade 8-12: Advanced placement; College in HS; International baccalaureate; Running start; Cambridge scholars program	
HS GPA	HS GPA	GPA in a 4.0 scale	
HS Graduation	Graduation	Whether or not there is a record of the student graduating HS	

Table 2: Sample Statistics for Kindergarten Readiness Sample

	Panel A: Full Sample		Panel B: Analytical Sample	
	Mean	SD	Mean	SD
Race				
Asian	0.05	0.22	0.05	0.21
Black	0.05	0.23	0.05	0.21
Hispanic	0.30	0.46	0.31	0.46
White	0.49	0.50	0.48	0.50
Other	0.03	0.17	0.03	0.17
Female	0.49	0.50	0.49	0.50
FRLP Eligible	0.61	0.49	0.61	0.49
With learning disability	0.00	0.02	0.00	0.02
LEP services	0.24	0.43	0.25	0.43
Number of domains met	4.24	2.09	4.27	2.07
Sample Size	98718	98718	87117	87117

Table 3: Sample Statistics for High School Sample

	Panel A: Full sample		Panel B: Analytical sample	
	Mean	SD	Mean	SD
Race				
Asian	0.07	0.26	0.08	0.27
Black	0.05	0.22	0.05	0.21
Hispanic	0.17	0.38	0.17	0.37
White	0.63	0.48	0.64	0.48
Others	0.07	0.25	0.06	0.24
Female	0.49	0.50	0.50	0.50
FRPL Eligible	0.44	0.50	0.41	0.49
With learning disability	0.05	0.22	0.04	0.19
LEP Services	0.09	0.28	0.09	0.29
3rd Grade ELA score	0.00	1.00	0.09	0.94
3rd Grade Math score	0.01	0.99	0.10	0.93
Sample Size	369143	369143	212951	212951

Table 4: Sample Statistics by Ethnicity

	All Students	Asian	Black	Hispanic	Others	White
WA Kids Dataset						
Cognitive Score	644.9	645.5	635.0	625.0	646.5	657.7
	(74.61)	(76.08)	(76.24)	(74.24)	(73.90)	(71.85)
Literacy Score	648.6	657.2	649.4	619.7	652.7	664.4
	(66.34)	(68.35)	(63.91)	(64.74)	(63.96)	(61.91)
Math Score	646.9	659.6	646.8	619.2	651.1	661.7
	(65.61)	(67.22)	(63.69)	(66.08)	(63.76)	(60.11)
Language Score	636.9	625.0	633.7	611.3	642.7	652.7
	(76.66)	(84.65)	(76.19)	(77.95)	(73.61)	(71.17)
Physical Score	624.6	629.0	620.0	619.1	626.2	627.7
	(56.35)	(55.84)	(60.35)	(56.52)	(56.17)	(55.62)
Socio-emotional Score	632.0	634.1	622.7	624.8	631.8	637.2
	(71.55)	(71.16)	(73.16)	(70.53)	(71.55)	(71.58)
Number of Domains	4.470	4.604	4.374	3.881	4.576	4.805
	(1.832)	(1.802)	(1.858)	(1.974)	(1.788)	(1.652)
Third Grade Outcomes						
Standardized Math Score	-0.135	0.250	-0.474	-0.419	-0.216	0.0657
	(0.958)	(0.944)	(0.923)	(0.903)	(0.963)	(0.933)
Standardized Reading Score	-0.150	0.116	-0.435	-0.463	-0.203	0.0694
	(0.969)	(0.949)	(0.922)	(0.912)	(0.979)	(0.944)
Unexcused Absences	1.576	1.004	2.989	1.800	2.541	1.117
	(4.205)	(2.891)	(6.449)	(4.256)	(5.944)	(3.325)
Total Absences	10.21	7.570	10.42	10.07	12.14	10.08
	(9.100)	(7.724)	(10.58)	(8.957)	(11.05)	(8.547)
Suspension	1.122	0.434	1.111	0.940	1.225	1.214
	(1.791)	(1.111)	(1.895)	(1.523)	(1.945)	(1.849)
Other Action	1.705	2.007	2.474	1.468	1.942	1.559
	(4.150)	(3.769)	(5.845)	(3.062)	(6.199)	(3.286)
High School Outcomes						
High School GPA	2.588	2.984	2.271	2.276	2.369	2.678
	(0.956)	(0.909)	(0.905)	(0.919)	(0.964)	(0.940)
Graduation Within 4 years	0.693	0.778	0.605	0.626	0.602	0.719
	(0.461)	(0.416)	(0.489)	(0.484)	(0.489)	(0.449)
Graduation Within 5 Years	0.724	0.796	0.649	0.664	0.638	0.748
	(0.447)	(0.403)	(0.477)	(0.472)	(0.481)	(0.434)
Unexcused Absences	0.0958	0.0635	0.208	0.139	0.115	0.0751
	(0.992)	(0.796)	(1.576)	(1.285)	(1.109)	(0.820)
Total Absences	19.33	12.65	26.15	24.19	23.08	17.64
	(17.45)	(14.28)	(22.27)	(20.35)	(20.39)	(15.28)
Suspensions	0.104	0.0308	0.224	0.144	0.151	0.0854

	(0.413)	(0.197)	(0.615)	(0.499)	(0.520)	(0.360)
Other Action	0.0151	0.00471	0.0266	0.0185	0.0162	0.0143
	(0.176)	(0.0836)	(0.265)	(0.224)	(0.162)	(0.159)

Table 5: Regression of WaKIDS Assessments by Domain on Student Demographics

VARIABLES	(1) Number of Domains	(2) Cognitive	(3) Literacy	(4) Math	(5) Language	(6) Physical	(7) Social Emotional
Black	-0.0308	-0.116***	0.0278*	0.0141	-0.0393**	-0.0446***	-0.0705***
	(0.0317)	(0.0162)	(0.0151)	(0.0152)	(0.0155)	(0.0173)	(0.0165)
Asian	0.0993***	-0.0516***	0.0470***	0.119***	-0.204***	0.0537***	0.0186
	(0.0300)	(0.0158)	(0.0152)	(0.0152)	(0.0164)	(0.0160)	(0.0158)
Hispanic	-0.342***	-0.120***	-0.241***	-0.231***	-0.148***	-0.00654	0.0374***
	(0.0177)	(0.00861)	(0.00834)	(0.00833)	(0.00853)	(0.00897)	(0.00873)
Others	-0.0529**	-0.0654***	-0.0626***	-0.0531***	-0.0443***	0.0150	-0.0133
	(0.0206)	(0.0107)	(0.0101)	(0.0101)	(0.0102)	(0.0110)	(0.0109)
Female	0.321***	0.160***	0.118***	-0.0300***	0.177***	0.235***	0.317***
	(0.0128)	(0.00637)	(0.00607)	(0.00602)	(0.00624)	(0.00665)	(0.00653)
FRPL	-0.633***	-0.319***	-0.446***	-0.403***	-0.315***	-0.194***	-0.256***
	(0.0137)	(0.00706)	(0.00674)	(0.00663)	(0.00685)	(0.00736)	(0.00726)
LEP Services	-1.335***	-0.478***	-0.653***	-0.661***	-0.680***	-0.153***	-0.268***
	(0.0217)	(0.00988)	(0.00952)	(0.00971)	(0.0103)	(0.0103)	(0.00982)
Learning	-1.324***	-0.640***	-0.730***	-0.764***	-0.597***	-0.467***	-0.542***
Disability	(0.0311)	(0.0146)	(0.0134)	(0.0138)	(0.0142)	(0.0163)	(0.0152)
Constant	4.919***	0.290***	0.436***	0.492***	0.313***	0.0558***	0.0707***
	(0.0118)	(0.00637)	(0.00600)	(0.00587)	(0.00609)	(0.00663)	(0.00660)
Observations	87,385	86,281	84,117	86,101	85,239	86,603	86,875
R-squared	0.172	0.127	0.228	0.221	0.173	0.045	0.078

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Regressions of 3rd Grade Test Outcomes on Student Demographics and Number of Readiness Indicators

VARIABLES	(1) Math	(2) Math	(3) Reading	(4) Reading	(5) Absences	(6) Absences	(7) Suspensions	(8) Suspensions
0 indicators met	-1.302***	-0.919***	-1.291***	-0.829***	2.954***	2.622***	0.0851***	0.0780***
	(0.0154)	(0.0156)	(0.0139)	(0.0142)	(0.172)	(0.180)	(0.0103)	(0.0108)
1 indicator met	-1.086***	-0.765***	-1.108***	-0.720***	2.069***	1.811***	0.0660***	0.0623***
	(0.0137)	(0.0134)	(0.0131)	(0.0130)	(0.151)	(0.155)	(0.00788)	(0.00824)
2 indicators met	-0.906***	-0.639***	-0.939***	-0.615***	1.747***	1.491***	0.0541***	0.0500***
	(0.0128)	(0.0122)	(0.0125)	(0.0120)	(0.133)	(0.136)	(0.00732)	(0.00768)
3 indicators met	-0.764***	-0.546***	-0.783***	-0.520***	1.514***	1.256***	0.0605***	0.0564***
	(0.0118)	(0.0111)	(0.0117)	(0.0111)	(0.124)	(0.126)	(0.00783)	(0.00808)
4 indicators met	-0.607***	-0.435***	-0.629***	-0.422***	1.316***	1.052***	0.0476***	0.0430***
	(0.00989)	(0.00934)	(0.0101)	(0.00950)	(0.104)	(0.104)	(0.00581)	(0.00595)
5 indicators met	-0.407***	-0.301***	-0.412***	-0.290***	0.992***	0.779***	0.0364***	0.0337***
	(0.00807)	(0.00755)	(0.00833)	(0.00778)	(0.0824)	(0.0825)	(0.00430)	(0.00437)
Black		-0.298***		-0.246***		-0.473***		0.0562***
		(0.0138)		(0.0139)		(0.166)		(0.0111)
Asian		0.291***		0.186***		-2.239***		-0.0444***
		(0.0133)		(0.0129)		(0.129)		(0.00420)
Hispanic		-0.0550***		-0.0461***		-0.608***		-0.0297***
		(0.00748)		(0.00753)		(0.0838)		(0.00446)
Others		-0.167***		-0.152***		1.519***		0.00965
		(0.00941)		(0.00961)		(0.115)		(0.00620)
Female		-0.129***		0.105***		0.00105		-0.0876***
		(0.00559)		(0.00564)		(0.0613)		(0.00316)
FRPL		-0.314***		-0.332***		2.441***		0.0491***
		(0.00633)		(0.00643)		(0.0661)		(0.00376)
LEP Services		-0.336***		-0.474***		-1.575***		-0.0482***
		(0.00837)		(0.00820)		(0.0991)		(0.00477)
Learning Disability		-0.877***		-0.792***		0.919***		-0.0196**
		(0.0125)		(0.0116)		(0.148)		(0.00845)
Constant	0.278***	0.557***	0.273***	0.443***	9.336***	8.353***	0.0346***	0.0663***
	(0.00444)	(0.00612)	(0.00456)	(0.00617)	(0.0421)	(0.0584)	(0.00180)	(0.00297)

Observations	82,881	82,881	82,959	82,959	85,970	85,970	85,995	85,995
R-squared	0.184	0.297	0.187	0.304	0.009	0.035	0.003	0.017

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Regressions of 3rd Grade Test Outcomes on Student Demographics and Readiness Domains

VARIABLES	(1) Math	(2) Math	(3) Reading	(4) Reading	(5) Absences	(6) Absences	(7) Suspensions	(8) Suspensions
Cognitive	0.0414***	0.0438***	0.0252***	0.0354***	-0.0729	-0.0400	0.00785**	0.00786**
	(0.00528)	(0.00501)	(0.00532)	(0.00503)	(0.0574)	(0.0568)	(0.00335)	(0.00333)
Literacy	0.223***	0.173***	0.296***	0.223***	-0.771***	-0.670***	-0.00341	0.00276
	(0.00530)	(0.00511)	(0.00535)	(0.00514)	(0.0575)	(0.0582)	(0.00315)	(0.00323)
Math	0.274***	0.208***	0.188***	0.143***	-0.383***	-0.330***	-0.00113	-0.00979***
	(0.00502)	(0.00484)	(0.00508)	(0.00490)	(0.0558)	(0.0564)	(0.00282)	(0.00293)
Language	-0.0163***	-0.0461***	0.0438***	-0.00574	0.668***	0.454***	0.0247***	0.0210***
	(0.00495)	(0.00478)	(0.00493)	(0.00475)	(0.0519)	(0.0526)	(0.00277)	(0.00276)
Physical	-0.0377***	-0.0188***	-0.0659***	-0.0477***	-0.313***	-0.262***	0.00467*	0.00840***
	(0.00387)	(0.00370)	(0.00387)	(0.00370)	(0.0428)	(0.0428)	(0.00242)	(0.00243)
Social Emotional	0.0673***	0.0770***	0.0661***	0.0624***	-0.359***	-0.254***	-0.0691***	-0.0598***
	(0.00431)	(0.00415)	(0.00429)	(0.00413)	(0.0476)	(0.0476)	(0.00347)	(0.00337)
Black		-0.308***		-0.258***		-0.432***		0.0568***
		(0.0133)		(0.0135)		(0.168)		(0.0113)
Asian		0.260***		0.169***		-2.114***		-0.0394***
		(0.0128)		(0.0126)		(0.130)		(0.00431)
Hispanic		-0.0127*		-1.96e-05		-0.670***		-0.0218***
		(0.00733)		(0.00740)		(0.0846)		(0.00450)
Others		-0.153***		-0.137***		1.495***		0.00984
		(0.00905)		(0.00925)		(0.116)		(0.00621)
Female		-0.109***		0.119***		-0.00484		-0.0806***
		(0.00551)		(0.00557)		(0.0628)		(0.00312)
FRPL		-0.242***		-0.257***		2.288***		0.0513***
		(0.00615)		(0.00625)		(0.0674)		(0.00389)
LEP Services		-0.253***		-0.374***		-1.625***		-0.0347***
		(0.00853)		(0.00845)		(0.102)		(0.00476)
Learning Disability		-0.771***		-0.686***		0.769***		-0.0186**
		(0.0125)		(0.0117)		(0.150)		(0.00840)

Constant	-0.131***	0.178***	-0.142***	0.0755***	10.20***	9.197***	0.0643***	0.0840***
	(0.00284)	(0.00576)	(0.00287)	(0.00580)	(0.0310)	(0.0588)	(0.00165)	(0.00345)
Observations	81,649	81,649	81,726	81,726	84,663	84,663	84,691	84,691
R-squared	0.281	0.352	0.284	0.356	0.016	0.040	0.012	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Dominance Analysis Weights by Kindergarten Readiness Domain

Outcome Variable	Cognitive	Literary	Math	Language	Physical	Social-Emotional	R-squared (weights)	R-squared (original)
3rd Grade Math	0.1334	0.214	0.4287	0.111	0.0339	0.0789	0.302	0.287
3rd Grade Reading	0.1339	0.2288	0.3886	0.1383	0.0296	0.0807	0.306	0.295
Total Absences	0.0979	0.1868	0.4058	0.0721	0.099	0.1383	0.025	0.024
Total Unexcused Absences	0.1267	0.2171	0.5008	0.0598	0.0327	0.063	0.064	0.063
All Disciplinary Incidences	0.0978	0.0448	0.074	0.0409	0.1115	0.631	0.06	0.05

Table 9: Variation in 3rd Grade Outcomes by Demographics

VARIABLES	(1) Math	(2) Reading	(3) Absences	(4) Suspensions
Black	-0.308***	-0.257***	-0.449***	0.0571***
	(0.0146)	(0.0145)	(0.167)	(0.0112)
Asian	0.301***	0.196***	-2.267***	-0.0455***
	(0.0140)	(0.0135)	(0.129)	(0.00420)
Hispanic	-0.0991***	-0.0876***	-0.503***	-0.0260***
	(0.00787)	(0.00786)	(0.0838)	(0.00446)
Others	-0.176***	-0.161***	1.545***	0.0104*
	(0.0100)	(0.0102)	(0.116)	(0.00621)
Female	-0.0780***	0.152***	-0.136**	-0.0921***
	(0.00587)	(0.00587)	(0.0611)	(0.00324)
FRPL	-0.419***	-0.431***	2.704***	0.0586***
	(0.00661)	(0.00663)	(0.0654)	(0.00371)
LEP Services	-0.506***	-0.633***	-1.151***	-0.0340***
	(0.00842)	(0.00818)	(0.0967)	(0.00445)
Learning Disability	-1.093***	-0.992***	1.472***	-0.00192
	(0.0126)	(0.0115)	(0.144)	(0.00809)
Constant	0.363***	0.258***	8.850***	0.0856***
	(0.00598)	(0.00599)	(0.0539)	(0.00323)
Observations	82,881	82,959	85,970	85,995
R-squared	0.221	0.238	0.029	0.014

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Regressions of High School Grade Outcomes on Student Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
VARIABLES		High School GPA			Grade 10 Math Score			Grade 10 ELA Score			High School Graduation Rate			High School Absences			High School Disciplinary Action	
3^{rd} Grade	0.00995***		-0.000254**	0.0204***		0.00462***	0.0211***		0.00581***	0.00306***		0.000701***	-0.0960***		0.0283***	-0.00147***		0.000262***
Percentile	(9.75e-05)		(0.000126)	(9.06e-05)		(0.000102)	(8.73e-05)		(9.83e-05)	(6.00e-05)		(8.24e-05)	(0.00193)		(0.00262)	(4.02e-05)		(5.51e-05)
8^{th} Grade		0.0146***	0.0147***		0.0260***	0.0228***		0.0261***	0.0221***		0.00389***	0.00340***		-0.160***	-0.180***		-0.00233***	-0.00251***
Percentile		(9.13e-05)	(0.000126)		(7.47e-05)	(0.000102)		(7.22e-05)	(9.82e-05)		(5.95e-05)	(8.23e-05)		(0.00189)	(0.00261)		(3.98e-05)	(5.50e-05)
Female	0.276***	0.226***	0.225***	-0.0680***	-0.154***	-0.146***	0.156***	0.0697***	0.0801***	0.0132***	0.000301	0.00156	1.447***	2.011***	2.062***	-0.0631***	-0.0550***	-0.0545***
	(0.00477)	(0.00449)	(0.00449)	(0.00443)	(0.00367)	(0.00364)	(0.00427)	(0.00355)	(0.00349)	(0.00294)	(0.00292)	(0.00293)	(0.0947)	(0.0928)	(0.0929)	(0.00197)	(0.00195)	(0.00196)
Asian	0.251***	0.111***	0.110***	0.362***	0.129***	0.145***	0.234***	0.00407	0.0237***	0.0901***	0.0554***	0.0577***	-4.481***	-2.868***	-2.772***	-0.0341***	-0.0112***	-0.0103***
	(0.00864)	(0.00818)	(0.00819)	(0.00803)	(0.00669)	(0.00663)	(0.00774)	(0.00647)	(0.00637)	(0.00532)	(0.00533)	(0.00534)	(0.171)	(0.169)	(0.169)	(0.00356)	(0.00356)	(0.00357)
Black	-0.113***	-0.0617***	-0.0627***	-0.211***	-0.151***	-0.133***	-0.152***	-0.0994***	-0.0762***	0.0357***	0.0445***	0.0473***	4.126***	3.400***	3.513***	0.0776***	0.0680***	0.0691***
	(0.0126)	(0.0118)	(0.0118)	(0.0117)	(0.00962)	(0.00954)	(0.0112)	(0.00931)	(0.00915)	(0.00773)	(0.00767)	(0.00767)	(0.249)	(0.243)	(0.244)	(0.00518)	(0.00512)	(0.00513)
Hispanic	-0.0579***	-0.0316***	-0.0321***	-0.120***	-0.0898***	-0.0798***	-0.0666***	-0.0406***	-0.0280***	0.0569***	0.0613***	0.0628***	2.051***	1.675***	1.736***	0.000229	-0.00472	-0.00416
	(0.00724)	(0.00679)	(0.00680)	(0.00674)	(0.00555)	(0.00550)	(0.00648)	(0.00537)	(0.00528)	(0.00446)	(0.00442)	(0.00443)	(0.144)	(0.141)	(0.141)	(0.00299)	(0.00296)	(0.00296)
Other Race	-0.0733***	-0.0555***	-0.0559***	-0.0514***	-0.0320***	-0.0245***	-0.0276***	-0.0110*	-0.00166	-0.0103*	-0.00743	-0.00631	2.269***	2.012***	2.057***	0.0150***	0.0117***	0.0121***
	(0.00892)	(0.00836)	(0.00836)	(0.00829)	(0.00684)	(0.00677)	(0.00798)	(0.00661)	(0.00650)	(0.00549)	(0.00545)	(0.00545)	(0.177)	(0.173)	(0.173)	(0.00368)	(0.00364)	(0.00364)
FRPL	-0.297***	-0.203***	-0.203***	-0.219***	-0.0887***	-0.0749***	-0.221***	-0.0990***	-0.0817***	-0.0783***	-0.0588***	-0.0567***	6.106***	4.883***	4.968***	0.0580***	0.0413***	0.0421***
	(0.00545)	(0.00516)	(0.00517)	(0.00507)	(0.00422)	(0.00419)	(0.00488)	(0.00408)	(0.00402)	(0.00336)	(0.00336)	(0.00337)	(0.108)	(0.107)	(0.107)	(0.00225)	(0.00225)	(0.00225)
Special ED	0.00180	0.0712***	0.0687***	-0.190***	-0.132***	-0.0866***	-0.318***	-0.275***	-0.218***	0.0230***	0.0316***	0.0384***	0.503***	-0.590***	-0.312**	0.0281***	0.0141***	0.0167***
	(0.00744)	(0.00689)	(0.00700)	(0.00692)	(0.00564)	(0.00567)	(0.00666)	(0.00545)	(0.00544)	(0.00459)	(0.00449)	(0.00456)	(0.148)	(0.143)	(0.145)	(0.00307)	(0.00300)	(0.00305)
LEP Services	0.0570***	0.0687***	0.0662***	0.0979***	0.0672***	0.112***	-0.0414***	-0.0841***	-0.0276***	-0.0933***	-0.0980***	-0.0912***	-0.782***	-1.170***	-0.894***	-0.0307***	-0.0348***	-0.0323***
	(0.00913)	(0.00848)	(0.00857)	(0.00849)	(0.00693)	(0.00693)	(0.00818)	(0.00670)	(0.00666)	(0.00563)	(0.00552)	(0.00558)	(0.181)	(0.175)	(0.177)	(0.00377)	(0.00369)	(0.00373)
Constant	2.410***	2.155***	2.160***	-0.880***	-1.170***	-1.267***	-0.876***	-1.128***	-1.250***	0.203***	0.160***	0.146***	19.71***	23.36***	22.76***	0.161***	0.209***	0.204***
	(0.00748)	(0.00683)	(0.00733)	(0.00695)	(0.00559)	(0.00594)	(0.00669)	(0.00540)	(0.00570)	(0.00460)	(0.00445)	(0.00477)	(0.148)	(0.141)	(0.152)	(0.00308)	(0.00297)	(0.00319)
Observations	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391	99,391
R-squared	0.241	0.332	0.332	0.500	0.660	0.666	0.549	0.691	0.701	0.061	0.077	0.077	0.128	0.167	0.168	0.056	0.075	0.075

Note: While not included, all models also include controls for student's gender, race/ethnicity, a Limited English Proficiency flag, an economically disadvantaged flag and participation in special education services. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Appendix

A1: WaKIDS Assessment Domains: Objectives

GOLD® Objectives and Dimensions (WaKIDS)

Social–Emotional

1. Regulates own emotions and behaviors
   a. Manages feelings
   b. Follows limits and expectations
   c. Takes care of own needs appropriately
2. Establishes and sustains positive relationships
   c. Interacts with peers
3. Participates cooperatively and constructively in group situations
   a. Balances needs and rights of self and others
   b. Solves social problems

Physical

4. Demonstrates traveling skills
5. Demonstrates balancing skills
7. Demonstrates fine-motor strength and coordination
   a. Uses fingers and hands
   b. Uses writing and drawing tools

Language

8. Listens to and understands increasingly complex language
   a. Comprehends language
   b. Follows directions
9. Uses language to express thoughts and needs
   b. Speaks clearly
10. Uses appropriate conversational and other communication skills
   a. Engages in conversations

Note: These 20 objectives are a subset of the 38 objectives for development and learning appearing in *Teaching Strategies GOLD® Objectives for Development & Learning, Birth Through Kindergarten*, © 2010 by Teaching Strategies, LLC, Bethesda, MD. The number associated with the objective corresponds with the *GOLD*® objective; numbers are missing when the associated *GOLD*® objective is not part of WaKIDS.

Cognitive

11. Demonstrates positive approaches to learning
   a. Attends and engages
   b. Persists
   c. Solves problems
12. Remembers and connects experiences
   a. Recognizes and recalls
13. Uses classification skills
14. Uses symbols and images to represent something not present
   a. Thinks symbolically

Literacy

15. Demonstrates phonological awareness
   a. Notices and discriminates rhyme
   c. Notices and discriminates smaller and smaller units of sound
16. Demonstrates knowledge of the alphabet
   a. Identifies and names letters
   b. Uses letter–sound knowledge
17. Demonstrates knowledge of print and its uses
   b. Uses print concepts
18. Comprehends and responds to books and other texts
   b. Uses emergent reading skills
19. Demonstrates emergent writing skills
   a. Writes name

Mathematics

20. Uses number concepts and operations
   a. Counts
   b. Quantifies
   c. Connects numerals with their quantities
21. Explores and describes spatial relationships and shapes
   b. Understands shapes

A2: Linkages Across Cohorts

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2003	K													
2004	1	K												
2005	2	1	K											
2006	3	2	1	K										
2007	4	3	2	1	K									
2008	5	4	3	2	1	K								
2009	6	5	4	3	2	1	K							
2010	7	6	5	4	3	2	1	K						
2011	8	7	6	5	4	3	2	1	K					
2012	9	8	7	6	5	4	3	2	1	K				
2013	10	9	8	7	6	5	4	3	2	1	K			
2014	11	10	9	8	7	6	5	4	3	2	1	K		
2015	12	11	10	9	8	7	6	5	4	3	2	1	K	
2016	PS1	12	11	10	9	8	7	6	5	4	3	2	1	K
2017	PS2	PS1	12	11	10	9	8	7	6	5	4	3	2	1
2018	PS3	PS2	PS1	12	11	10	9	8	7	6	5	4	3	2
2019	PS4	PS3	PS2	PS1	12	11	10	9	8	7	6	5	4	3

HS Outcomes Sample

Kindergarten Sample

Our sample can be divided into two main groups: a kindergarten to 3rd grade sample, and a 3rd grade to 12th grade sample. The kindergarten readiness assessment program (WaKIDS) was rolled out and started reporting data starting in 2015, and we are able to track these students to third grade. The High School Outcomes Sample includes data from students who started 3rd grade between 2006 and 2016, and we can track these students to 12th grade.

A3: Proportion of Kindergarten Ready Students by Domain and Year

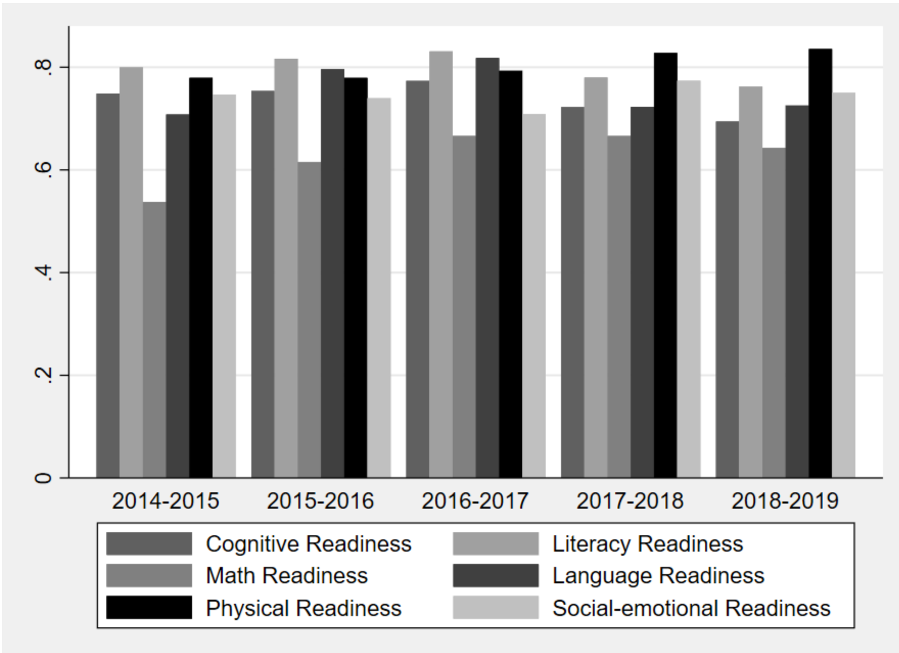


Figure A3 displays the proportion of students deemed to meet the kindergarten readiness threshold for each readiness domain each school year. Overall, all the domains have at least 50% of the students meeting the readiness threshold every year. The math readiness domain consistently has the lowest proportion deemed kindergarten ready, while an increasing proportion of students are deemed ready in the physical domain. None of the other domains follow a clear trend over time.

A4: High School Pathways: Average Scores by Free and Reduced Price Lunch Eligibility in 3rd, 8th, and 10th Grades

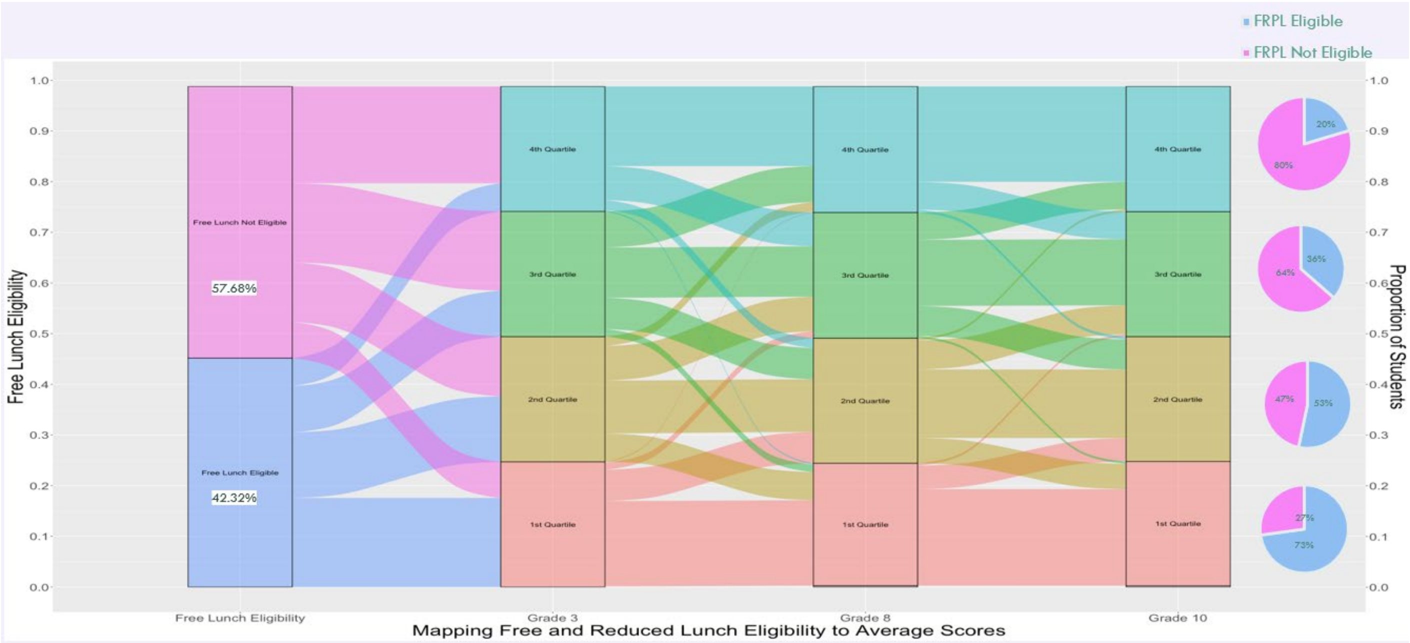


Figure A4 explores the variation in FRLP Eligibility and average test scores in 3rd, 8th, and 10th grades. The stacked bar chart on the far left displays the proportion of students who are eligible for FRLP overall in the sample. Math and reading test scores for each grade have been averaged and divided into quartiles, with the 4th quartile representing the highest 25% of scores, and the 1st quartile representing the lowest 25% of scores. Students who are eligible for FRLP are more likely to perform in lower on reading and math tests in the 3rd grade. This discrepancy persists through to 10th grade, with this group of students making up the majority of students in the lower end of the score distribution. This is shown by the pie charts on the far right, which represent the proportion of students by eligibility in different 10th grade average score quartiles.

A5: High School Pathways: Math Scores by Gender in 3rd, 8th, and 10th Grades

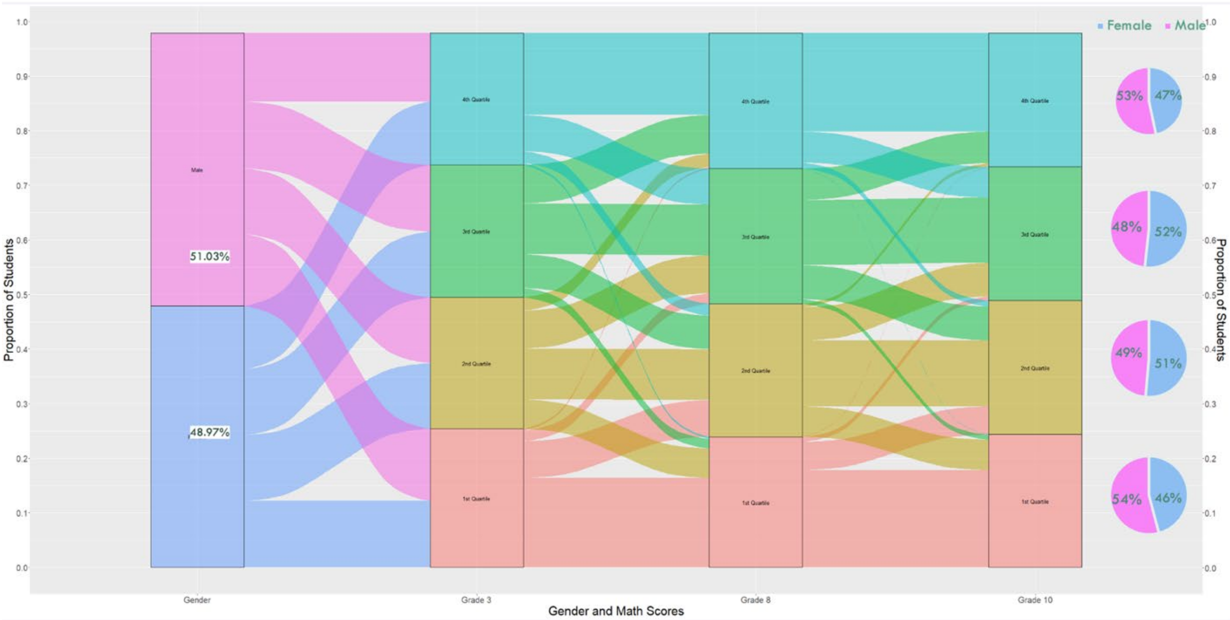


Figure A5 explores the variation in math test scores and gender in 3rd, 8th, and 10th grades. The stacked bar chart on the left displays the proportion of students by gender in our sample. This math test score has been divided into quartiles for each grade, with the 4th quartile representing the 25% of students with the highest scores, and the 1st quartile representing the 25% of students with the lowest scores. There is relatively little variation by gender across the quartiles, though female students are slightly more likely to be in the middle of the distribution, while male students are slightly more likely to be on the outer ends of the distribution (the 1st and 4th quartiles). This is shown by the pie charts on the far right, which represent the proportion of students by gender in different grade 10 math test quartiles. Another point of interest is that over 50% of students in the 4th quartile in grade 3 remain in the 4th quartile in 8th grade, with a similar trend for 10th grade. This persistence of performance over time can also be seen in the 1st quartile. Within the 2nd and 3rd quartile, we observe more mobility over time.

A6: High School Pathways: Reading Scores by Gender in 3rd, 8th, and 10th Grades

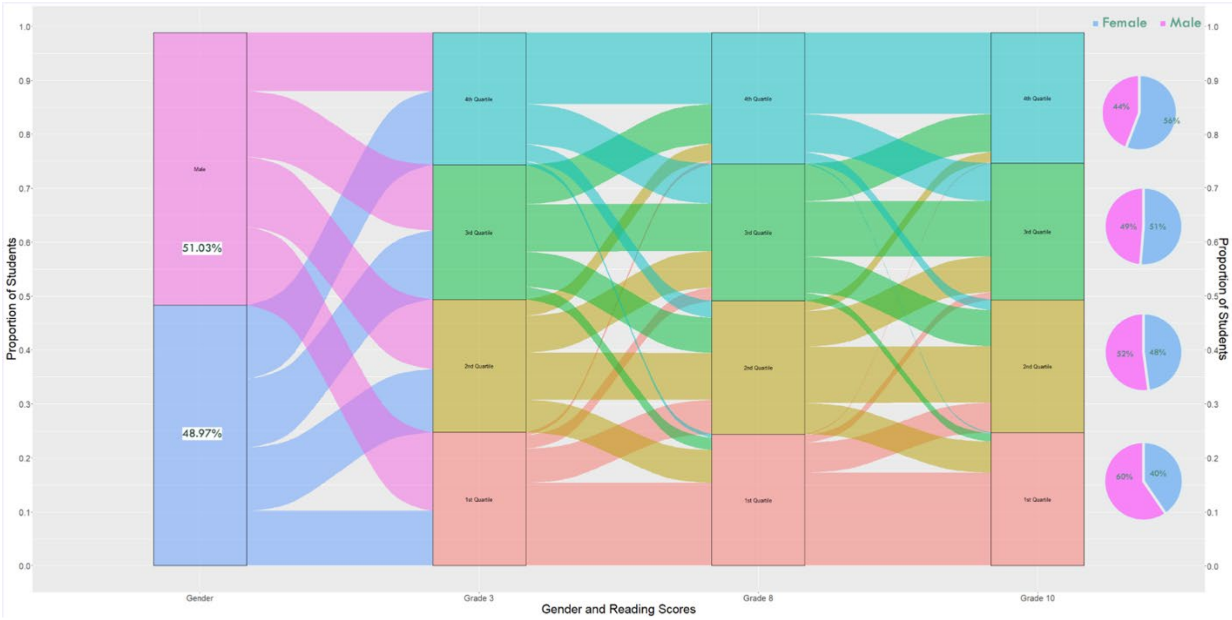


Figure A6 explores the variation in reading test scores and gender in 3rd, 8th, and 10th grade. The stacked bar chart on the left displays the proportion of students by gender in our sample. This reading test score has been divided into quartiles for each grade, with the 4th quartile representing the 25% of students with the highest scores, and the 1st quartile representing the 25% of students with the lowest scores. There is relatively little variation by gender across the quartiles, though female students are slightly more likely to test in the higher end of the distribution compared to their male peers. This is shown by the pie charts on the far right, which represent the proportion of students by gender in different 10th grade reading test quartiles. Another point of interest is that the majority of students in the lowest quartile in 3rd grade remain in the lowest quartile in 8th grade, with a similar trend for 10th grade. Within the 2nd and 3rd quartile, we observe more mobility of reading test scores over time.

A7: High School Pathways: Average Scores by LEP Service Status in 3rd, 8th, and 10th Grades

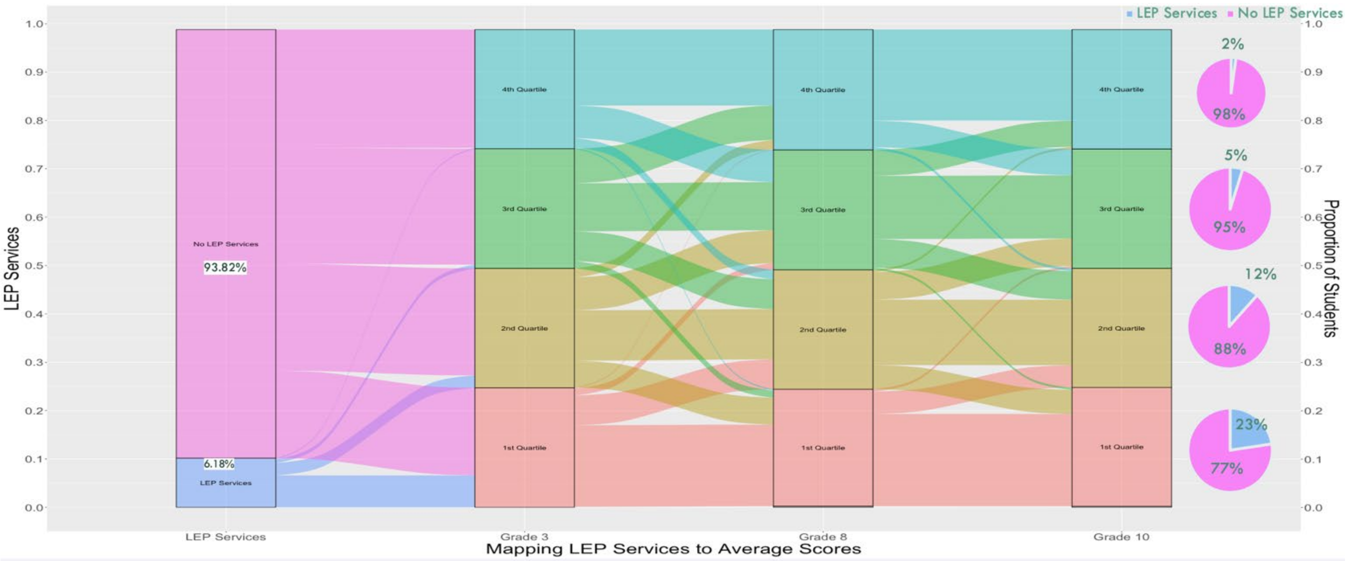


Figure A7 explores the variation in average test scores and students who use LEP services in 3rd, 8th, and 10th grades. The stacked bar chart on the left displays the proportion of students utilizing LEP services, (Limited-English Proficient services), as determined in 3rd grade. Math and reading test scores have been averaged and divided into quartiles for each grade, with the 4th quartile representing the 25% of students with the highest scores, and the 1st quartile representing the 25% of students with the lowest scores. Students in the LEP services category are more likely to transition to the bottom quartile of the test score distribution and this effect is persistent from 3rd grade to 10th grade. This is shown by the pie charts on the far right, which represent the proportion of students in the LEP services category in different 10th grade average test quartiles.

A8: High School Pathways: Average Scores by Disability Status in 3rd, 8th, and 10th Grades

Figure A8 explores the variation in average test scores and Learning Disability status in 3rd, 8th, and 10th grades. The stacked bar chart on the left displays the proportion of students identified as having a Learning Disability in 3rd grade. Math and reading test scores have been averaged and divided into quartiles for each grade, with the 4th quartile representing the 25% of students with the highest scores, and the 1st quartile representing the 25% of students with the lowest scores. Students who are identified as having a Learning Disability are more likely to transition to the bottom quartile of the test score distribution, and this effect is persistent to 10th grade. This is shown by the pie charts on the far right, which represent the proportion of students by Learning Disability status in different 10the grade average test quartiles.

A9: Unexcused Absences Analysis

VARIABLES	(1) 3rd Grade unexcused absences	(2) 3rd Grade unexcused absences	(3) HS unexcused absences	(4) HS unexcused absences
Cognitive	-0.0644**	-0.0392		
	(0.0268)	(0.0266)		
Literacy	-0.345***	-0.253***		
	(0.0261)	(0.0264)		
Math	-0.164***	-0.110***		
	(0.0260)	(0.0263)		
Language	0.161***	0.149***		
	(0.0240)	(0.0241)		
Physical	-0.00661	-0.0242		
	(0.0190)	(0.0190)		
Social-Emotional	-0.0643***	-0.0562**		
	(0.0224)	(0.0222)		
Black		1.531***		7.009***
		(0.101)		(0.0841)
Asian		-0.0189		-0.577***
		(0.0497)		(0.0677)
Hispanic		0.339***		3.386***
		(0.0385)		(0.0531)
Other Races		1.169***		3.302***
		(0.0597)		(0.0672)
Female		-0.00343		-0.541***
		(0.0289)		(0.0350)
FRPL Eligible		0.991***		5.355***
		(0.0288)		(0.0392)
LEP		-0.467***		0.172***
		(0.0484)		(0.0667)
Learning Disability		0.235***		
		(0.0721)		
3rd grade math			-2.098***	-1.320***
			(0.0252)	(0.0250)
3rd grade reading			-1.831***	-1.100***
			(0.0251)	(0.0256)
Constant	1.577***	0.734***	8.708***	5.400***
	(0.0143)	(0.0243)	(0.0178)	(0.0327)
Observations	84,663	84,663	530,183	525,677
R-squared	0.013	0.040	0.071	0.142

A10: Coefficients of Regressions on Non test PCA vars
Early Grade Non-Test Outcomes on Later Grade Non-Test Outcomes

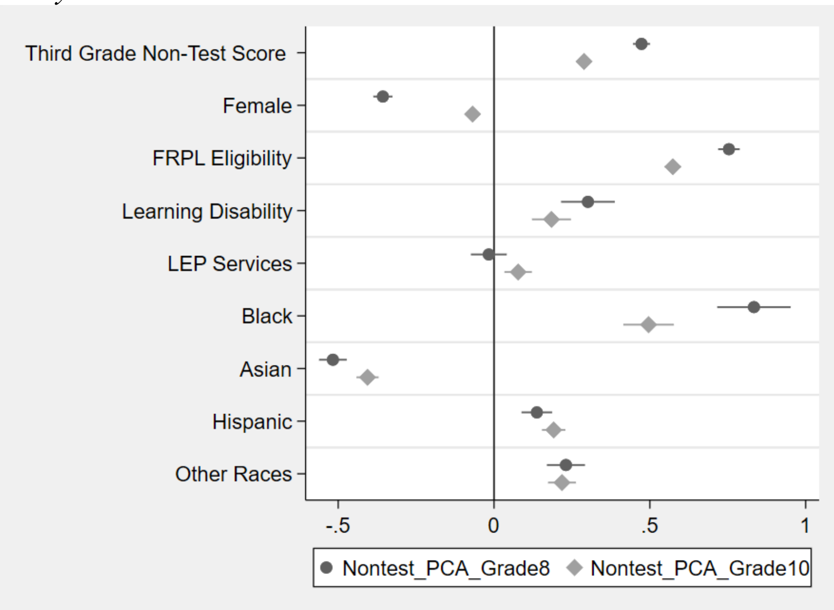


Figure A10 shows the regression results from running 3rd grade non-test scores as well as other student characteristics on 8th grade and 10th grade non-test scores, respectively. The non-test score variable is an index composed of information related to absences, disciplinary incidents, and suspension, formed using principal-component analysis. The x-axis shows the magnitude of the marginal effect measured from the regression analysis, with the horizontal lines on each point indicating a 95% confidence interval for the point estimate. In general, the sign of the correlations remains consistent from 8th grade to 10th grade, while the magnitude of the relationships seems to decrease over time. Students who are Black and students who are eligible for FRLP are more likely to have absences, disciplinary incidents, etc than their peers in both 8th grade and 10th grade. In addition, students who have more absences and disciplinary incidents in 3rd grade are more likely to have these incidents in later grades.

A11: Advanced Algebra Course Taking by Student Race/Ethnicity

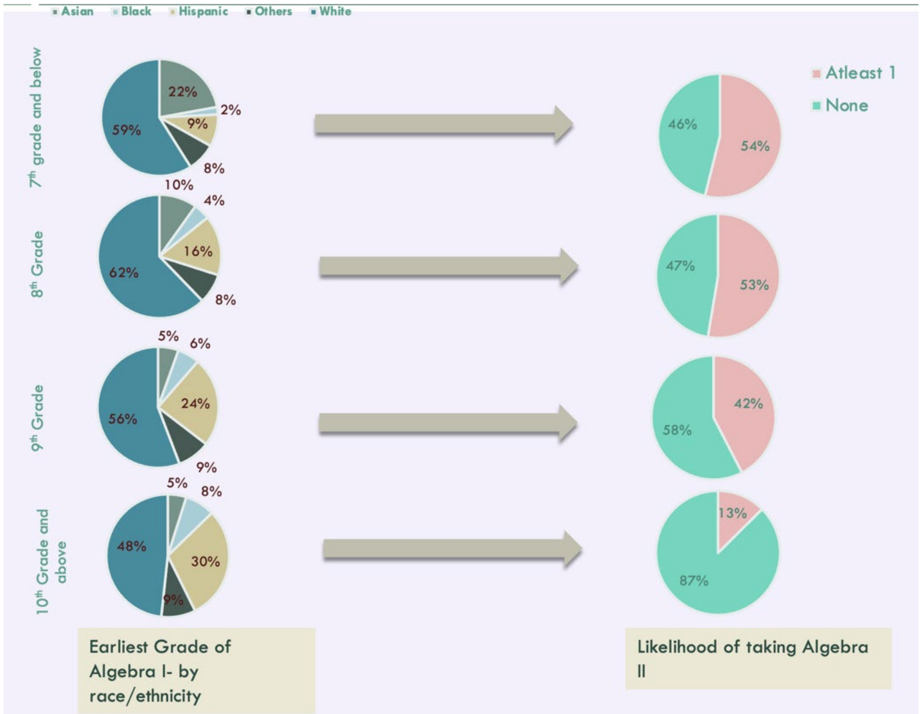
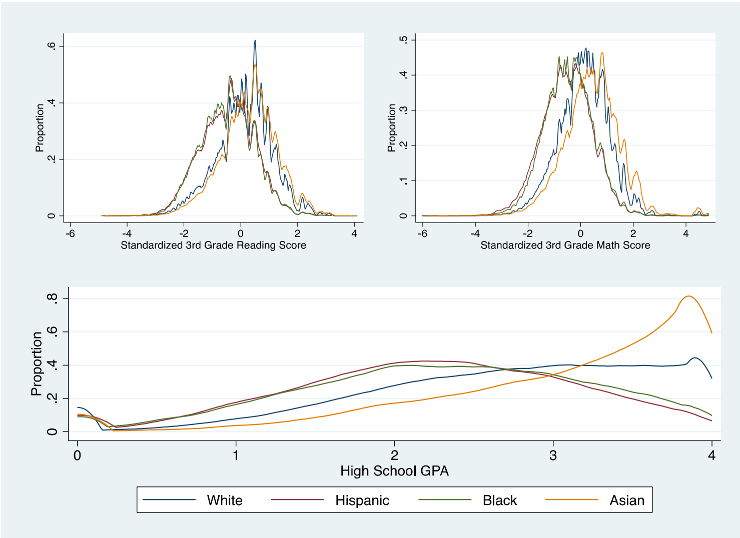


Figure A11 shows the transition probability of taking Algebra 2 conditional on taking Algebra 1, with variation by race/ethnicity and grade level. The pie charts on the left indicate the earliest grade that the student was enrolled in Algebra 1, and the race/ethnicity proportions within each cohort. Students who are White and Asian are more likely to take Algebra 1 in 8th grade or earlier while Hispanic and Black students are more likely to take the course in 9th grade or later. By following each cohort to the pie chart on the right, we can see what proportion go on to enroll in Algebra 2 courses. Of those who took Algebra 1 in 8th grade or earlier, the majority go on to take advanced Algebra courses (53% and 54% in the pie charts above, respectively). However, only 13% of those who take Algebra 1 in 10th grade or later end up enrolling in Algebra 2.

A12: Distribution of 3rd Grade Reading, 3rd Grade Math, High School GPA



The top two graphs of Figure A12 display the standardized distribution of reading and math test scores by student race/ethnicity in 3rd grade. The math and reading scores are standardized using the overall group mean and standard deviation. Differences in the distributions by race/ethnicity are present in 3rd grade, with White and Asian students on average scoring higher than Black and Hispanic students. The bottom panel shows the distribution of high school Grade Point Average (GPA) by student race/ethnicity, with the Hispanic and Black student distributions as bell-shaped curves and a modal GPA between 2.0 and 2.5. In comparison, the White and Asian student distributions have a modal GPA between 3.7 and 3.9. GPA is measured on a 4.0 scale.

A13: Gender Pathways in 3rd, 8th, and 10th Grade Reading Scores

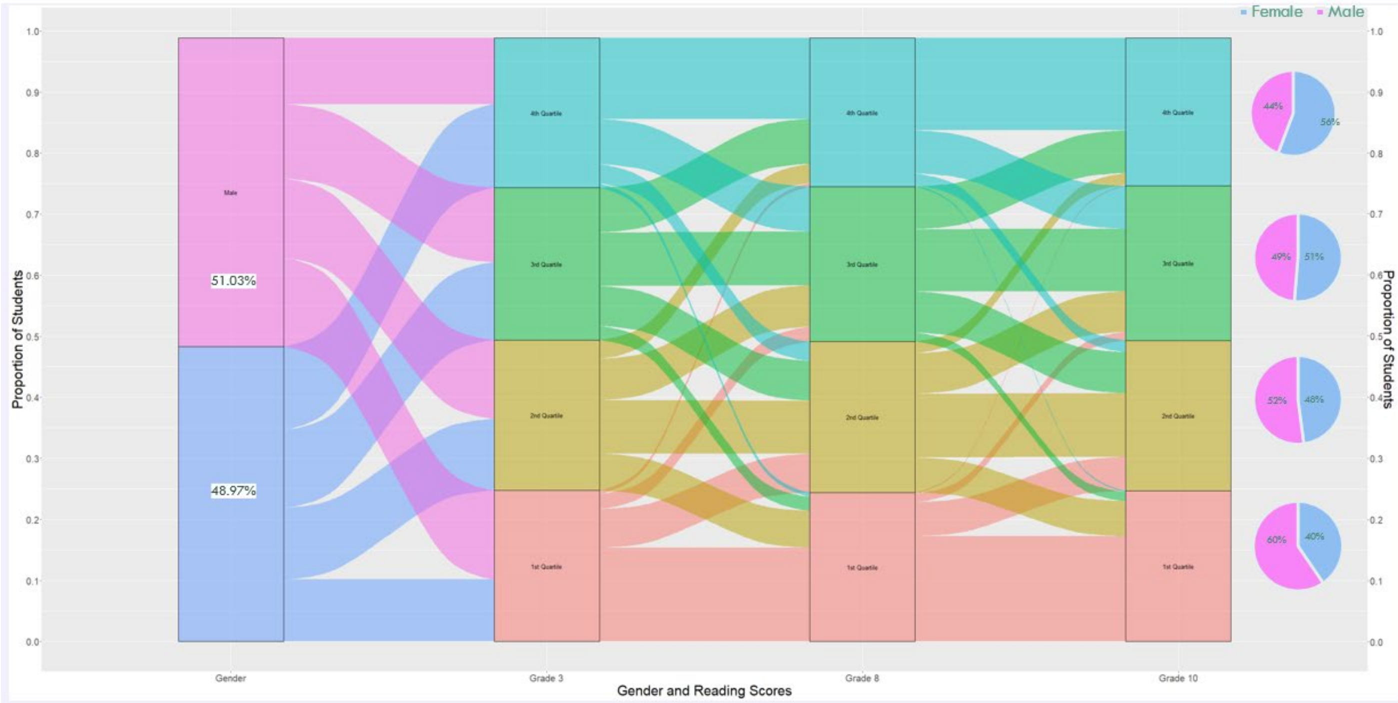


Figure A13 explores the variation in gender and reading scores in 3rd, 8th, and 10th grades. The stacked bar chart on the far left displays the proportion of students identifying as male or female overall in the sample. The reading scores for each grade have been divided into quartiles, with the 4th quartile representing the highest 25% of scores, and the 1st quartile representing the lowest 25% of scores. As can be seen above, male and female students are present in each of the 3rd grade quartiles in roughly equal proportions. By 10th grade, however, female students are slightly more likely to perform higher in reading than their male peers. This is shown by the pie charts on the far right, which represent the proportion of proportion of students by gender in different 10th grade reading quartiles.