Skills That Pay: Unpacking Subject-Specific Competencies to Predict Economic Wellbeing*

Cameron Conrad† Nolan G. Pope‡ George W. Zuo§

Abstract

This paper studies how students' K-12 academic performance relates to their long-term educational attainment and earnings. Using linked administrative data from Maryland, we estimate both descriptive correlations and causal effects via a teacher value-added framework. Both approaches show that math and English state standardized test scores are similarly predictive of broad educational attainment, but math scores are substantially more predictive of STEM degrees and earnings. Mediation analysis reveals that most of the English-earnings relationship is explained by educational attainment, while nearly half of the math-earnings relationship remains unexplained. Heterogeneity analysis confirms that math scores are more predictive of earnings across all student subgroups. However, the strength of this relationship is weaker for historically disadvantaged and lower-achieving students, while English scores show a stronger association with earnings for these same groups. These findings suggest that policies aimed at fostering economic mobility should consider differences in the strength of relationships between subject-specific skills and long-term outcomes across student groups.

*We thank Matthew Chingos, Maggie Reeves, and Victoria Nelson as well as conference participants at the Student Upward Mobility Initiative (SUMI) mini conference for helpful comments. We are grateful for financial support from SUMI at the Urban Institute, a sponsored project of Rockefeller Philanthropy Advisors. Funding for SUMI is provided by the Walton Family Foundation, the Bill & Melinda Gates Foundation, the Joyce Foundation, and the Overdeck Family Foundation. Conrad's contribution is part of the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2236417. This research was supported by the Maryland Longitudinal Data System (MLDS) Center. We are grateful for the assistance provided by the MLDS Center. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations, the Urban Institute and its trustees, or the MLDS Center and its partner agencies. All remaining errors are our own. This study was determined to be exempt by the University of Maryland IRB (Project #2205386).

†PhD Candidate, Department of Economics, University of Maryland, College Park (cconrad3@umd.edu).

‡Associate Professor, Department of Economics, University of Maryland, College Park; Research Fellow, NBER (npope@umd.edu).

§Economist, RAND Corporation (gzuo@rand.org).

1 Introduction

Students develop a wide range of skills in school, but not all of them carry equal weight in shaping economic outcomes. There is ongoing debate about which specific skills matter most, for whom they matter, and whether schools emphasize the competencies most relevant to labor market success. Standardized test scores are generally treated as proxies for student skill or ability and play a prominent role in high-stakes decisions, from teacher evaluation to school accountability, particularly since the passage of the landmark No Child Left Behind Act in 2001 (U.S. Congress, 2001). College admissions decisions also rely heavily on standardized test scores and students' grade point averages (GPAs). Given the central role of these proxies in educational and institutional decision-making, understanding how they relate to long-term outcomes is critically important.

Most prior research examining the relationship between cognitive skills and labor market outcomes relies on longitudinal survey data. Several studies find that general cognitive ability, as measured by the Armed Forces Qualifying Test (AFQT), is strongly associated with earnings.[1] Other research distinguishes between math and verbal skills, finding that math skills are more strongly linked to earnings, and that the math-earnings relationship has become stronger over time.[2] However, these studies typically rely on small survey samples, self-reported outcomes, and lack quasi-random variation to support causal inference.

A notable exception is Chetty et al. (2014b), which uses administrative data and quasi-experimental variation in teacher value-added for test scores to estimate causal effects on college attainment and earnings. Yet even this work pools skills across subjects for most outcomes.[3] There remains limited research that uses administrative data to examine long-term outcomes, applies quasi-experimental designs to identify causal effects, and directly

[1] Neal and Johnson (1996); Heckman et al. (2006); Lin et al. (2018) use the AFQT from the National Longitudinal Survey of Youth (NLSY), which combines components focused on both math and verbal skills.

[2] Murnane et al. (1995, 2000) use U.S. Department of Education longitudinal surveys. Grogger and Eide (1995); Dougherty (2003); Deming (2017) use the NLSY. Hanushek et al. (2015) use international survey data from the Programme for the International Assessment of Adult Competencies (PIAAC).

[3] Their primary estimates pool math and English teacher value-added. They separately estimate subject-specific value-added effects only for college "quality", finding somewhat larger effects for English.

compares which academic skills are most predictive of labor market success.

The purpose of this paper is to use linked administrative data to examine how proxies for skill, such as test scores and GPAs, predict and influence long-term outcomes. This study contributes to the literature and policy debate in three key ways. First, by leveraging linked administrative data for the full population of students in Maryland, we overcome several limitations of survey-based studies. The use of population-level data and objectively measured outcomes allows for greater external validity and more granular analysis of subgroup differences. Second, we provide new evidence on which proxies for academic skill are most predictive of educational attainment, degree field, and earnings by subject area. Our findings reveal differences in how subject-specific skills are associated with long-term outcomes and highlight implications for policies aimed at promoting economic mobility. Third, we go beyond descriptive correlations by employing a teacher value-added approach that uses quasi-random variation in teacher assignment to estimate the causal effects of skill acquisition on long-run outcomes. By combining comprehensive administrative data with both descriptive and quasi-experimental methods, this study offers some of the most fine-grained evidence to date on how school-acquired skills shape students' economic trajectories.

We use data from the Maryland Longitudinal Data System (MLDS) Center, which links students' test scores and GPAs from K-12 education to postsecondary and earnings records. Our empirical strategy combines two complementary approaches. First, we estimate descriptive regressions to examine how test scores and GPAs predict long-term outcomes, including educational attainment, degree field, and earnings. Second, we exploit quasi-random variation in teacher assignment to implement a value-added approach, estimating the causal effect of being assigned to a higher or lower value-added teacher.[4] We also conduct mediation analysis to assess the extent to which educational attainment and degree field explain the relationship between test scores and earnings, and we estimate heterogeneous effects across demographic and achievement subgroups.

[4]Prior research has shown teacher value-added estimates represent unbiased estimates of teachers' causal effects on student outcomes (Chetty et al., 2014a; Jackson, 2018; Petek and Pope, 2023).

The results of this analysis show that both math and English Language Arts (ELA) test scores are predictive of broad educational attainment outcomes, including college enrollment and bachelor of arts (BA) degree receipt. However, math scores are substantially more predictive of earning a BA in a STEM field and of higher earnings in adulthood. A one standard deviation (SD) increase in test scores is associated with a 7.7 percentage point (p.p.) (74 percent) increase in STEM BA attainment for math and a 2.5 p.p. (24 percent) increase for ELA. The same increase in scores is associated with an increase in average annual earnings at ages 29-32 of approximately $8,100 (17 percent) for math and $2,200 (4.5 percent) for ELA.

The estimated relationship between test scores and earnings attenuates moderately when GPAs are included in the regression model. Notably, GPAs in ELA and social studies (SS) show stronger associations with earnings than those in math and science. These findings suggest that GPA, particularly in verbal subjects, captures skill dimensions relevant to later labor market performance that are not fully reflected in test scores. This may be because grades partly reflect non-cognitive skills such as effort, motivation, and organization, which are less directly measured by standardized tests.[5]

We also examine the relationship between test scores and a broader set of BA degree fields. Estimates reveal that associations are strongest when the subject of the test score closely aligns with the degree field. When test scores for all four core subjects are included in the model,[6] math scores most strongly predict attainment in math, engineering, and business fields; ELA scores are most predictive of humanities degrees; science scores of science degrees; and SS scores of social science degrees. These patterns are consistent with students selecting into fields in which they hold a comparative academic advantage.

Mediation analysis indicates that most of the relationship between ELA test scores and earnings is explained by educational attainment and degree field. In contrast, a much larger

[5] The larger coefficients for ELA and SS GPAs relative to math and science GPAs may also reflect differences in course difficulty or grading standards across subjects.

[6] We define the four core subjects as math, ELA, science, and SS.

share of the math-earnings relationship remains unexplained by these mediators. Even after controlling for these educational factors, a one SD increase in math scores is associated with an earnings increase of $3,700 (7.7 percent), compared to just $700 (1.5 percent) for ELA. This persistent gap underscores the particularly strong link between math scores and earnings.

Heterogeneity analysis confirms that the math-earnings relationship is stronger than the ELA-earnings relationship across all demographic subgroups. However, the analysis also reveals meaningful variation in the strength of these associations. The relationship between math scores and earnings is substantially weaker for students from disadvantaged groups, including those eligible for Free and Reduced-price Meals (FARMS), Black students, or students with lower achievement, while the ELA-earnings relationship is relatively stronger for these groups relative to their peers. We also find that test scores measured at higher grade levels are more predictive of later outcomes, and that test score-earnings relationships are stronger for earnings measured at older ages.

Finally, results from our teacher test score value-added analysis align with patterns observed in the descriptive estimates. Increases in teacher value-added for math and ELA both have similarly sized impacts on broad educational attainment. However, math teacher value-added has substantially larger effects on STEM degree attainment and earnings than ELA value-added. A one SD increase in teacher value-added raises on-time BA attainment by approximately 0.5 p.p. for both subjects. However, the same increase in math teacher value-added increases STEM degree receipt by 1.3 p.p. and earnings by roughly $184 (0.8 percent), with the latter effect marginally significant.

This paper proceeds as follows. We discuss policy background in Section 2. Section 3 describes the data while Section 4 details the methodology for our descriptive correlational analysis (4.1) and teacher value-added analysis (4.2). Section 5 presents main results (5.1), heterogeneity results (5.2), teacher value-added results (5.3), and robustness results (5.4). We conclude the paper in Section 6.

2 Policy Background

Policy Overview. The publication of *A Nation at Risk* in 1983 ushered in an era of education reform that emphasized rigorous curricula and increased use of standardized testing for accountability. This increased focus on conceptual reasoning and application of knowledge over rote memorization led to three critical waves of policy change affecting state standardized testing in Maryland. First, Maryland introduced the Maryland School Performance Assessment Program (MSPAP) for elementary and middle school students in 1990 and first administered end-of-course high school assessments (HSA) in 2000. Second, the passage of the federal No Child Left Behind (NCLB) Act in 2001 required Maryland to develop the Maryland School Assessments (MSA) so that achievement data could be disaggregated by student subgroups. Third, the Maryland State Board of Education adopted the Common Core State Standards (CCSS) in 2010, leading to the implementation of the CCSS-aligned tests from the Partnership for Assessment of Readiness for College and Careers (PARCC) (MSDE, 2015). Similarly, the state adopted the Next Generation Science Standards (NGSS) in 2013, leading to the implementation of the NGSS-aligned Maryland Integrated Science Assessment (MISA) in 2018 (MSDE, 2025).

MSPAP and HSA. Maryland's focus on rigorous learning standards and corresponding assessment dates to at least 1990 with the introduction of the Maryland School Performance Assessment Program (MSPAP). This test was given to 3rd, 5th, and 8th grade students to assess competency in reading, language, writing, math, science, and social studies. To complement the MSPAP, the state also established the high school assessments (HSA) in the 1990s with the first tests administered in 2000. Students were assessed via high-stakes exams upon completion of their Algebra I, 10th grade English, Biology, and U.S. Government courses.[7] The HSA tests continued to be administered through the mid- to late-2010s

[7]Data collection from HSA exams includes the overall composite scores and proficiency levels as well as subscores focused on particular skill areas. Appendix Table A1 shows subscores for each of these exams. The Algebra I exam has 4 subscores (e.g. Analyzing Patterns and Functions), the English 10 exam has 4 subscores (e.g. Reading and Literature: Comprehension and Interpretation), the Biology exam has 6 subscores (e.g. Skills and Processes of Biology), and the U.S. Government exam has 5 subscores (e.g. U.S. Government

depending on the subject, while the MSPAP was phased out with the introduction of the MSA to satisfy the requirements of NCLB (MSDE, 2015).

NCLB and MSA. The passage of the landmark federal No Child Left Behind (NCLB) legislation in 2001 placed greater importance on assessment, data, and accountability in schools nationwide. This legislation required states to adopt state-level academic standards and develop aligned tests. Schools had to assess students annually in reading and math in grades 3-8 and once in high school. The law also required states to disaggregate test data by student subgroups to provide greater transparency about educational inequities. This emphasis on assessment and data was paired with accountability mechanisms. Schools were required to demonstrate Adequate Yearly Progress (AYP) towards all students achieving proficient levels on their tests, and schools failing to to hit AYP goals faced accountability measures such as school improvement plans or restructuring (U.S. Congress, 2001).

In response to the requirements of NCLB, Maryland developed the Maryland School Assessments (MSA) in reading and math. These tests allowed the state to report testing results by student subgroups. Students in grades 3, 5, and 8 were first assessed in 2003, while those in grades 4, 6, and 7 were first assessed in 2004. The MSA science tests was first administered to students in grades 5 and 8 in 2007.[8] The HSA assessments first administered in 2000 were also part of the NCLB-era testing regime given the law's requirement to test all high school students at least once (MSDE, 2015).

CCSS, PARCC, and NGSS. In 2010, a multi-state effort the Common Core State Standards (CCSS) Initiative began with the aim of strengthening the rigor and consistency of state academic standard nationwide. These standards focused on the knowledge and skills that students should possess in math and ELA by the end of each grade. Maryland's State Board of Education adopted CCSS in 2010 and the state joined the Partnership for Assess-

Structure, Functions, and Principles) (MSDE, 2012).

[8]Data collection from MSA tests includes the overall composite scores and proficiency levels as well as subscores focused on particular skill areas. Appendix Table A2 shows subscores for each of these tests. The math test has 5 subscores (e.g. Algebra, Patterns, or Functions), the ELA test has 3 subscores (e.g. General Reading Processes), and the science test has 6 subscores (e.g. Skills and Processes) (MSDE, 2003, 2016).

ment of Readiness for College and Careers (PARCC) consortium to develop and administer the standards-aligned tests. PARCC assessments in math and ELA were first given to students in grades 3-8 and high school in 2015 and continued through 2019 (MSDE, 2015).[9]

Continuing the trend of implementing more rigorous standards and tests, the Maryland State Board of Education also adopted the Next Genereation Science Standards (NGSS) in 2013. Similar to CCSS, these standard focus more on depth over breadth, conceptual reasoning, and knowledge application. NGSS focuses on three broad areas that students are expected to master: i) Science and Engineering Practices, ii) Crosscutting Concepts, and iii) Disciplinary Core Ideas. Following the adoption of NGSS, Maryland established the NGSS standards-aligned Maryland Integrated Science Assessments (MISA). MISA was first administered to 5th and 8th grade students in 2018 and replaced the MSA science test while the HS MISA test was first administered to HS students in 2019 as an end-of-course Biology exam that replaced the HSA Biology test MSDE (2025).

Together, these waves of policy reform reshaped Maryland's K-12 assessment system over nearly three decades, expanding both the scope and stakes of standardized testing. The resulting data, spanning multiple subjects, grade levels, and assessment regimes, form the foundation for our analysis. We discuss Maryland's assessment data in more detail in the next Section (3).

3 Data

Data Overview. We use data from the Maryland Longitudinal Data System (MLDS) Center from 2008 to 2024. The MLDS Center provides a centralized repository of linked student and workforce data stemming from a partnership among numerous state and non-profit entities. There are three key sets of data for our analysis: i) K-12 data from the Maryland State Department of Education (MSDE), ii) postsecondary data from the Maryland Higher Education Commission (MHEC) and the National Student Clearinghouse (NSC),

[9]The PARCC tests for high school students were for Algebra I and English 10.

and iii) workforce data from the Maryland Department of Labor Division of Unemployment Insurance (UI).

The K-12 data from MSDE include information on all students enrolled in Maryland public schools, including attendance and demographics, assessments, course enrollments, school information, and high school graduation.[10] For postsecondary data, MHEC provides information on all students enrolled in Maryland colleges, while NSC captures out-of-state enrollment for Maryland public high school graduates. The two sources substantially overlap and we have nearly complete coverage for our postsecondary outcomes for the cohorts we are able to observe those outcomes for.[11] The earnings data from the Maryland Division of UI cover wages for individuals who work for Maryland employers that report to the state UI. Consequently, the earnings data exclude wages for the following types of employees: federal, military, self-employed, private contractor, or out-of-state.[12]

Construction of Samples. We use three different samples with varying structures depending on students' grade and the analytic task. The three samples include: i) elementary and middle school (EMS) student sample, ii) high school (HS) student sample, and iii) teacher value-added (TVA) student sample. All samples include students enrolled in Maryland regular public schools.[13] The EMS and HS samples also exclude outgoing transfer students who transfer from a Maryland public high school to a private or out-of-state high school.[14]

The construction of the three different samples is as follows. First, the EMS sample is student-academic year level data for all Maryland 3rd through 8th grade students from 2008-2019 with non-missing math and English Language Arts (ELA) test scores. Math and

[10] Course enrollment data begin in 2013.

[11] We are missing information on bachelor's degree field for less than 0.2 percent of observations, as shown in Table 1.

[12] In 2023, 5.9 percent of the Maryland workforce was employed in the federal government,(Maryland State Archives, 2025) compared to 1.8 percent of the U.S. population (BLS, 2025a,b).

[13] We include students enrolled in traditional, charter, and career and technical education schools and exclude students enrolled in special education, alternative, and other special program schools.

[14] We identify outgoing transfers using exit codes (MSDE, 2020). We include these students in the TVA sample to construct value-added measures although we are not able to observe long-term outcomes for outgoing transfer students.

ELA were assessed annually for these grades and years.[15]

Second, the HS sample is student-level data for all Maryland students from 2008-2019 with non-missing Algebra I, English 10, Biology, and U.S. Government end-of-course HS test scores. These tests are intended to be administered once although some students may take a test more than once if they did not pass on their first attempt. In cases of multiple assessment attempts, we retain the score from the first attempt at the assessment. Although these are HS-level assessments, some students first take them in middle school, particularly the Algebra I assessment.[16]

Third, the TVA sample used to construct TVA measures is student-year-subject level data for all Maryland 4th through 8th grade students from 2013-2019 with non-missing math or ELA test scores. The MLDS course data can be used to link students to their math and ELA classroom teachers each year. We link our student-year EMS data to course data to construct student-year-subject data for the relevant grades and years.[17] We impose five sample restrictions that are common practice in the TVA literature (Chetty et al., 2014a; Jackson, 2018; Petek and Pope, 2023). First, each observation must have non-missing data for the necessary observable covariates to compute TVA measures, in particular lagged own subject test score. Second, we omit observations when there is more than one subject area course observed for a student in an academic year. Third, we drop classrooms with over 40 percent of students identifying as special education (SPED). Fourth, we drop classrooms in which the assigned teacher is linked to 200 or more students in an academic year. Fifth, we drop classrooms with fewer than 7 students or 40 or more students. All of these sample restrictions are essential for the computation of TVA estimates or to reduce measurement error in the estimates. Once computed, TVA measures are linked to the TVA sample, which

[15] The math and ELA assessments from 2008-2014 were MSA and the assessments from 2015-2019 were PARCC. 5th and 8th graders were also assessed in science in 2008-2015 (MSA) and 2018-2019 (MISA).

[16] The Algebra I and English 10 assessments from 2008-2014 were HSA and the assessments from 2015-2019 were PARCC. The Biology assessment from 2008-17 was HSA and for 2019 was HS MISA. The U.S. Government exam from 2008-2019 was HSA.

[17] We focus our TVA analysis on students in the EMS sample, where annual assessments in both math and ELA allow us to control for prior test scores, a key requirement for the TVA model. In contrast, HS assessments are administered only once, limiting our ability to implement the same approach.

is reshaped to the student-year level with separate TVA variables for math and English, respectively.

Key Explanatory Variables. The key explanatory variables are standardized test scores and standardized GPAs. For the EMS sample, the two main explanatory variables are math and ELA test scores, which are standardized to have a mean of zero and a standard deviation of one. Other explanatory variables for this sample also include annual GPA measures in math, ELA, science, and social studies (SS), which are also standardized to have a mean of zero and a standard deviation of one. For the HS sample, the four main explanatory variables are standardized test scores for Algebra I, English 10, Biology, and U.S. Government. This sample also includes standardized GPA measures in the four core subjects, but these GPA measures are cumulative based on grades earned throughout HS since this sample contains student-level data.

Student Outcomes. Our main outcome variables include measures of educational attainment, degree field, and earnings. First, we create binary indicators for educational attainment, including high school graduation within four years, on-time college enrollment, associate of arts (AA) degree receipt within four years, and bachelor of arts (BA) degree receipt within 6 years.[18] We classify degree fields using the STEM-designated degree list maintained by the Department of Homeland Security (DHS, 2023), along with the Classification of Instructional Programs (CIP) codes from the National Center for Education Statistics (NCES, 2020). Our degree field outcomes include indicators for BA completion within six years in the following categories: i) STEM; ii) mathematics, statistics, or computer science; iii) engineering or architecture; iv) life, physical, and environmental science; v) social science; vi) humanities; vii) health professions; viii) business; and ix) education.[19]

[18]The timing of educational attainment outcomes are measured relative to the first year of observed enrollment in 9th grade in Maryland public schools. Thus high school graduation is within 4 years of 9th grade enrollment, on-time college enrollment is within 5 years of 9th grade enrollment, etc.

[19]There is substantial overlap in i) STEM majors and CIP categories ii)-iv). However, some DHS-designated STEM fields are outside the field categories determined using CIP codes in ii)-iv). Additionally, not all fields in ii)-iv) are DHS-designated STEM fields. Field categories defined using CIP codes ii)-ix) are defined to be mutually exclusive categories, although a small percentage of students may earn multiple degrees from different fields. The CIP major categories are not collectively exhaustive, but do comprise the

For workforce outcomes, our main outcomes are conditional average annual earnings at ages 22-24, 25-28, and 29-32. We create these measures in four steps: i) sum quarterly earnings for all employers over the calendar year, ii) convert to 2024 dollars using the CPI-U (Minneapolis Fed, 2025), iii) winsorize at the 99th percentile, and iv) take the average across the ages within each respective age range. Other workforce outcomes are employment and unconditional earnings. Employment is defined using a binary indicator for having positive earnings observed in Maryland UI data in each respective age range. Unconditional earnings measures impute zeros for those who are missing earnings.[20]

Descriptive Statistics. Table 1 provides summary statistics for both the EMS and HS samples. Overall both samples are relatively similar and there is slight positive selection in the samples due to the exclusion of specialized schools and restriction to non-missing test scores. One of the main differences in the samples is that almost 42 percent of observations in the EMS sample are Free and Reduced-price Meals (FARMS) eligible while only about 36 percent of the HS sample is FARMS eligible. Standardized test score means are all above zero. This suggests that conditioning the EMS sample on non-missing math and ELA test scores and the HS sample on non-missing test scores in the four core subjects induces some small positive selection in the samples. This sample restriction also reduces the variance of the test score and GPA variables. There is less evidence of positive selection on the GPA variables, which may be driven by students with non-missing test scores sorting into classrooms in which they are graded relative to a more positively selected group of peers. High school graduation rates are higher in the HS sample, which is likely driven by the fact that there is less high school dropout among students with non-missing high school test scores relative to the EMS sample which is conditional on non-missing test scores in earlier grades. However, BA receipt, STEM BA receipt, and earnings from ages 25-28 and 29-32

most common majors.

[20] The largest sources of missingness in earnings data are non-employment and out-of-state migration. During the time period when earnings are measured in our study, the national prime-age employment to population ratio ranged from 75 to 81 percent (BLS, 2025c). This suggests that up to 25 percent of observations that are missing earnings are real zeros from non-employment.

are higher in the EMS sample.

4 Empirical Strategies

We use two complementary empirical strategies in our analysis. First, we conduct descriptive analysis to estimate which composite test scores and GPAs are most predictive of educational attainment and earnings. Second, our causal analysis estimates impacts of teacher value-added (TVA) for students' test scores on long-term outcomes.

4.1 Relationships Between Student Performance and Outcomes

For our descriptive analysis, we follow the approach in Bettinger et al. (2013) and estimate the additional predictive power that each composite test score provides conditional on other composite test scores and demographic covariates. Our main multivariate composite test score specification is shown in the following equation:

$$Y_i = \alpha + \sum_{k=1}^{K} \beta^k S_{it}^k + \gamma C_{it} + \epsilon_{it}, \tag{1}$$

where Y_i is an outcome (e.g. BA receipt or earnings) for student i; S_{it}^k is a composite test score variable for subject k in academic year t, standardized in the population to have mean zero and standard deviation one; C_{it} are demographic covariates including indicators for gender, race, gender-race interactions, FARMS, English-language learner (ELL), special education (SPED), and imputed values for demographic variables; and ϵ_{it} is an error term clustered at the student level. β^k estimates the relationship between a one standard deviation increase in composite test S_{it}^k conditional on other subject tests S_{it}^{k-1} and demographic covariates C_{it}. Equation (1) can be generalized to include only one composite test score ($K = 1$) in univariate models or to include standardized GPA variables in models that also estimate the conditional relationship between standardized GPA and outcomes.

This specification is useful for assessing which subject test scores are most strongly associated with educational attainment and earnings, but threats from omitted variable bias suggest we are not able to make causal inferences. For example, although we include a crude proxy for socioeconomic status with our FARMS covariate and may partially account for non-cognitive skills in our model that also includes GPAs, our β^k coefficients are still likely biased upwards due to the omission of more precise measures of family socioeconomic status and more comprehensive measures of non-cognitive skills. We turn next to our TVA strategy in an effort to identify causal effects.

4.2 Teacher Value-Added

We build on prior research showing that TVA measures yield unbiased estimates of teachers' causal effects on student outcomes (Chetty et al., 2014a; Jackson, 2018; Petek and Pope, 2023). The TVA approach involves residualizing current-year student test scores on a rich set of covariates, including prior test scores and demographics. A teacher's value-added is estimated by predicting the average residual score of their students in a given year using residuals from students taught in other years. These TVA estimates have been widely used to study the causal impacts of teachers on a range of outcomes. The main identifying assumption is that TVA measures and student unobservables that also affect outcomes are uncorrelated. Key threats to identification include non-random student sorting and peer effects. However, the aforementioned prior research using TVA methods demonstrates that controlling for prior achievement and peer characteristics substantially mitigates these concerns.

There are three main steps necessary to implement our TVA empirical strategy. First, we construct test score TVA measures following the methodology in Chetty et al. (2014a). Second, we estimate the effects of test score TVA on on educational attainment and earnings. Third, we assess the validity of our research design by estimating forecast bias.

Estimating Test Score TVA. There are two main steps to estimate test score TVA:

i) residualize student test scores on our vector of controls and ii) predict residual test score TVA in the current year with teacher test score residuals in surrounding years. First, we residualize test score S^k_{ijt} by regressing it on a vector of controls X_{ijt} for student i in teacher j's class in academic year t for subject k. The vector controls for lagged student achievement, demographic characteristics, and the classroom environment. Specifically, the controls include: i) lags of a cubic polynomial in students' math and ELA test scores; ii) lags of a cubic polynomial in class- and grade-level means of those test scores; iii) the same current demographic controls used in our descriptive test score model (Equation 1): indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and imputed values for missing demographics; and iv) class- and grade-level means of those demographic variables. All these covariates are interacted with grade fixed effects, and we include a control for class size.[21] Residualized test scores ν^k_{ijt} are computed using the following equation:

$$\nu^k_{ijt} = S^k_{ijt} - \hat{\Gamma} X_{ijt}. \tag{2}$$

This residualization purges S^k_{ijt} of measures of prior achievement and demographics for each student, each student's class, and each student's grade.

Second, we compute the mean of the residuals across students by year for each teacher j. This is estimated as follows:

$$\bar{\nu}^k_{jt} = \frac{1}{N} \sum_{i=1}^{N} \nu^k_{ijt}. \tag{3}$$

We then predict test score TVA in year t with mean teacher test score residuals in surrounding years:

$$\hat{\nu}^k_{jt} = \sum_{s=t-a}^{s=t+a} \hat{\psi}_s \bar{\nu}^k_{js} \mathbb{1}[s \neq t]. \tag{4}$$

[21] When lagged scores in the other subject are missing, we set the other subject lagged score to zero and include an indicator for missing data in the other subject interacted with controls for lagged own-subject test scores.

This approach uses data from surrounding years because unobservables in year t that affect both the mean teacher test score residual $\bar{\nu}^k_{jt}$ and test scores in year t could bias estimates of teacher impacts on test scores.[22] This yields the test score TVA measure $\hat{\nu}^k_{jt}$ for teacher j in year t. The weight on the value-added measure $\hat{\psi}_s$ varies by the number of years before or after year t. The weights are estimated by minimizing the mean-squared error of the difference between $\bar{\nu}^k_{jt}$ and the predictions of $\bar{\nu}^k_{jt}$ made with the teacher test score residual measure the years before and after t. The minimization problem is:

$$\Psi = \underset{\{\psi_{t-a},\ldots,\psi_{t+a}\}}{\arg\min} \sum_{j=1}^{J} (\bar{\nu}^k_{jt} - \sum_{s=t-a}^{s=t+a} \hat{\psi}_s \bar{\nu}^k_{js} \mathbb{1}[s \neq t]) \tag{5}$$

This yields leave-year-out jackknife test score TVA predictions which allow teacher quality to drift over time and shrinks TVA predictions to the mean through Bayesian shrinkage Chetty et al. (2014a).

Estimating Effects of Test Score TVA on Outcomes. Our primary aim in the TVA analysis is to use the leave-year-out estimates of test score TVA $\hat{\nu}^k_{jt}$ constructed in the previous step to estimate teachers' causal effects on educational attainment and earnings. Our main model to estimate these effects is a multivariate TVA specification including TVA from multiple subjects as follows:

$$Y_i = \alpha + \sum_{k=1}^{K} \delta^k \hat{\nu}^k_{jt} + \Gamma X_{ijt} + \eta_{ijt}, \tag{6}$$

where Y_i is an outcome, $\hat{\nu}^k_{jt}$ are TVA measures for teacher j in academic year t in subject k, X_{ijt} is our vector of controls from Equation (2),[23] and η_{ijt} is the error term clustered at

[22] In Equation (4), a equals 6 years.

[23] The controls include: i) lags of a cubic polynomial in students' math and ELA test scores; ii) lags of a cubic polynomial in class- and grade-level means of those test scores; iii) the same current demographic controls used in our descriptive test score model (Equation 1): indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and imputed values for missing demographics; and iv) class- and grade-level means of those demographic variables. All these covariates are interacted with grade fixed effects, and we include a control for class size. When lagged scores in the other subject are missing, we set the other subject lagged score to zero and include an indicator for missing data in the other subject interacted with controls for lagged own-subject test scores.

the school level. δ^k is the coefficient of interest and represents the impact of having a one standard deviation higher test score TVA teacher in subject k on outcomes conditional on TVA $\hat{\nu}_{jt}^{k-1}$ in other subjects. In practice, our main multivariate specification includes just two TVA measures: math TVA $\hat{\nu}_{jt}^{m}$ and ELA TVA $\hat{\nu}_{jt}^{e}$. Equation (6) can also be generalized to include only one test score TVA measure (K=1) in univariate models.

The main identifying assumption is that the residualized leave-year-out predicted TVA measures in Equation (6) and student unobservables that affect outcomes Y_i are uncorrelated. If we let tildes denote the residualized TVA measures, then we represent $\tilde{\nu}_{jt}^{m} = \hat{\nu}_{jt}^{m} - \hat{\Gamma} X_{ijt} - \hat{\delta}^e \hat{\nu}_{jt}^{e}$. Then the main identifying assumption can be stated as follows:

$$Cov(\tilde{\nu}_{jt}^{m}, \eta_{ijt}) = Cov(\tilde{\nu}_{jt}^{e}, \eta_{ijt}) = 0 \tag{7}$$

The key threat to identifying the impact of TVA on long-term outcomes is if students with relatively higher- or lower-than-average unobserved likelihood of attaining an outcome (e.g. BA receipt, higher earnings) systematically sort to teachers with relatively higher or lower test score TVA. Prior research has shown there is little bias in test score TVA estimates (Chetty et al., 2014a; Petek and Pope, 2023; Jackson, 2018), but we explore this threat further in our estimation of forecast bias.

Estimating Forecast Bias. To estimate forecast bias, we first estimate the relationship between residualized test scores ν_{ijt}^{k} and estimated test score TVA $\hat{\nu}_{jt}^{k}$:

$$\nu_{ijt}^{k} = \alpha + \lambda \hat{\nu}_{jt}^{k} + \xi_{ijt}. \tag{8}$$

Under random assignment of teachers, regressing test score residuals ν_{ijt}^{k} on $\hat{\nu}_{jt}^{k}$ yields a λ coefficient of 1 because $\hat{\nu}_{jt}^{k}$ is the best linear predictor of ν_{ijt}^{k}. The amount of forecast bias in test score TVA $\hat{\nu}_{jt}^{k}$ is $B(\hat{\nu}_{jt}^{k}) = 1 - \lambda$. The main idea is that the degree of forecast bias can be quantified by the extent to which students are sorted to teachers on unobservables.

Using our TVA sample, we present estimates of the relationship between estimated test score TVA and residualized test scores in Appendix Table A3. For math, we find that a one unit increase in current year test score TVA increases current year test scores by 1.12 standard deviations, with the confidence interval for the coefficient ranging from 1.06 to 1.17. For ELA, we find that a one unit increase in current year test score TVA increases current year test scores by 1.1 standard deviations, with the confidence interval for the coefficient ranging from 1.01 to 1.19. Plugging in our estimates for λ into the forecast bias equation, we find estimates of forecast bias in the range of -17 to -6 percent for math and -19 to -1 percent for ELA. This implies that our TVA estimates are downward biased and understate the true variance in teacher quality. The attenuation of our TVA estimates relative to the true values also implies that using our TVA measures to estimate impacts on long-term outcomes will lead to bias away from zero in the range of 6 to 17 percent for math and 1 to 19 percent for ELA. Given this, the magnitudes of the estimates for TVA impacts on long-term outcomes should be interpreted with caution.

5 Results

Our discussion of results begins by examining the relationship between test scores and GPAs and long-term outcomes in Section 5.1. Next, we explore heterogeneity by FARMS status, race, achievement, grade, and age when earnings are measured in Section 5.2. Third, we discuss TVA effects on long-term outcomes in Section 5.3. Finally, we conduct robustness checks to assess the relationship between test scores and employment, unconditional earnings, and conditional earnings in Section 5.4.

5.1 Relationships Between Test Scores and GPAs and Long-Term Outcomes

Table 2 shows estimates of univariate relationships between test scores and outcomes. For the EMS sample, we find that for educational attainment outcomes the magnitude of the coefficients for math test scores are somewhat larger than estimates for ELA test scores, but overall the estimates are relatively similar. A one standard deviation (SD) increase in test scores is associated with increases high school graduation within four years by 6 percentage points (p.p.) for math and 5.2 p.p. for ELA. The results also show about a 14 p.p. increase in on-time college enrollment. For AA attainment within four years, the coefficient for math test scores is twice as large as the coefficient for ELA tests scores: a 0.6 p.p. (8.8 percent) increase compared to a 0.3 p.p. (4.4 percent) increase. The coefficients for BA receipt within 6 years and STEM BA receipt within 6 years are also somewhat larger for math test scores. A one SD increase in math test scores is linked to increases in BA receipt by about 17 p.p. (53 percent) and STEM BA receipt by about 9.3 p.p. (90 percent) while the coefficients for ELA for these same respective outcomes are 15 p.p. (47 percent) and 7.3 p.p. (70 percent). However, the earnings estimates show substantially larger coefficients for math scores than for ELA scores. We find that a one SD increase in math test scores corresponds to increases in earnings from ages 29-32 by $9,500 (20 percent) while the estimate for ELA test scores is about $7,000 (14 percent).

For the HS sample, most of the estimates are similar and show similar patterns. We also estimate univariate relationships for science (Biology) and social studies (U.S. Government) test scores in addition to math (Algebra I) and ELA (10th grade English) for the HS sample. One difference between our EMS and HS sample estimates is that our estimates for the relationship with AA receipt within 4 years are considerably larger in the HS sample. All test score coefficients show that a one SD increase in scores is related to increases in AA receipt by about 1.5 p.p. (20 percent). For age 29-32 earnings, the coefficients for ELA, science, and social studies are relatively similar in magnitude, all showing at least a $6,200 increase in earnings for a one SD increase in test scores. The math coefficient is larger, but smaller than the math coefficient in the EMS sample: a one SD increase in math test scores is associated with increases in earnings by about $7,800.

Next, we investigate the multivariate relationship between test scores and long-term outcomes in Table 3. For EMS students, we find that the inclusion of both test scores in the same model attenuates coefficients substantially, although we still find strong relationships between test scores and outcomes. We find somewhat larger coefficients for math relative to ELA for most educational attainment outcomes, although the coefficient for math for STEM BA receipt is about three times larger than the coefficient for ELA for this outcome, consistent with math skills being particularly predictive of STEM persistence. We find that a one SD increase in test scores corresponds to increases in on-time college enrollment by 9.2 (14 percent) p.p. for math and 7.9 p.p. (12 percent) for ELA. The coefficients for BA receipt are 12 p.p. (37 percent) for math and 7.7 p.p. (24 percent) for ELA while the estimates for STEM BA receipt are 7.7 p.p. (74 percent) for math and 2.5 p.p. (24 percent) for ELA. For ages 29-32 earnings, relative to the univariate estimates, the multivariate estimate for math declines only modestly while the multivariate estimate for ELA drops substantially. Consequently, the coefficient for math is nearly four times as large as the coefficient for ELA: a one SD increase in test scores is linked to increases in earnings from ages 29-32 by about $8,100 for math (17 percent) and $2,200 for ELA (4.5 percent).

For the HS sample, we find somewhat larger coefficients for general educational attainment for ELA and social studies (SS) test scores, while the coefficients for science test scores are the smallest. We find that BA receipt increases by 5.9 p.p. for math, 6.4 p.p. for ELA, 3.6 p.p. for science, and 6.9 p.p. for social studies. However, this pattern reverses for STEM BA receipt: a one SD increase in test scores increases STEM BA receipt by 3 p.p. for math, 2 p.p. for ELA, 3.7 p.p. for science, and 2 p.p. for SS. The pattern for earnings estimates in the HS sample is similar to the pattern in the EMS sample. Relative to the univariate relationships, the coefficients for math attenuate the least while the coefficients for ELA decline the most. For earnings from ages 29-32, we find that a one SD increase in test scores is related to increases in earnings by about $4,700 for math, $800 for ELA, $1,500 for science, and $2,800 for SS.

Now we examine the degree to which adding GPAs to the econometric model attenuates the estimates for test scores in Table 4. For the EMS sample, we find that the inclusion of GPAs in the model attenuates estimates considerably, although most coefficients still show a strong relationship between test scores and outcomes. The test score coefficients for high school graduation shrink by nearly an order of magnitude, which may be driven by the fact that GPAs may be particularly important determinants of high school graduation. We also find that for most educational attainment outcomes, the coefficients are often larger for ELA and SS GPAs relative to math and science GPAs.[24]

For BA receipt, the test score estimates diminish moderately: we find that conditional on GPAs, a one SD increase in test scores corresponds to increases in BA attainment by 7.7 p.p. for math and 3.9 p.p. for ELA. We also find that conditional on test scores, a one SD in GPAs is associated with increases in BA attainment by 2.3 p.p. for math, 6 p.p. for ELA, 4.1 p.p. for science, and 5.8 p.p. for SS. However, the estimated relationship between math test scores and STEM BA receipt is nearly identical to the estimate in the model excluding GPAs. We find a one SD increase in math scores is associated with an increase in STEM BA attainment by about 7.7 p.p. while the coefficient for ELA drops to 1.4 p.p. For STEM degree receipt, the coefficient for math GPAs is the largest while the coefficient for SS GPAs is the smallest.

Finally, for earnings from ages 25-28, we find moderate attenuation of test score estimates with the inclusion of GPAs in the model.[25] The math coefficient attenuates from \$4,900 to \$3,500, while the ELA coefficient declines from \$800 to –\$900 after controlling for GPAs. Similar to the educational attainment outcomes, we also find stronger relationships between ELA and SS GPAs and earnings than for math and science GPAs. A one SD increase in GPAs is linked to increases in earnings by about \$800 for math, \$1,800 for ELA, \$1,100 for science, and \$1,400 for SS.

The results for the HS sample show similar patterns. For BA attainment, the coefficients

[24]The larger coefficients for ELA and SS GPAs relative to math and science GPAs may partially be driven by differences in course difficulty and grading standards between ELA and SS compared to math and science.

[25]We are unable to observe earnings from ages 29-32 in our EMS sample with the inclusion of GPAs in the model because the course data to compute GPAs first begin in 2013.

for math, ELA, and SS test scores all suggest at least a 2.7 p.p. increase in BA receipt, while the coefficient for science is slightly negative. The coefficients for ELA and SS GPAs are substantially larger than the coefficients for math and science GPAs. A one SD increase in GPA is related to increases in BA attainment by 11 p.p. for ELA, 7.1 p.p. for SS, 4.1 p.p. for science, and 3.6 p.p. for math. For STEM BA receipt, the coefficients are larger for math and science than for ELA and SS, both for test scores and GPAs. For earnings from ages 29-32, we again find moderate attenuation of test score estimates. The coefficient attenuates from \$4,700 to \$2,900 for math, from \$800 to \$90 for ELA, from \$1,500 to \$1,100 for science, and from \$2,800 to \$2,400 for SS. The coefficients for GPAs are relatively similar in magnitude, with all GPA coefficients showing at least a \$1,500 increase in earnings.

Table 5 shows estimates of the multivariate relationship between test scores and BA degree field within 6 years. Overall, we find that that there are stronger positive associations between test scores and degree field when the test score subject area and degree field are closely related. For the EMS sample, the difference between math and ELA coefficients is largest for math/statistics/computer science and engineering/architecture fields while the difference between ELA and math coefficients is largest for social sciences and humanities fields. We also find that the math coefficient is considerably larger than the ELA coefficient for science and business fields. Specifically, all the estimates for the aforementioned math, engineering, science, and business fields show that a one SD increase in math test scores is associated with increases in degree attainment in the respective field by at least 1.7 p.p. (business) and as much as 2.7 p.p. (engineering/architecture). Among these fields, the ELA coefficient of 1.4 p.p. is largest for science fields, a category that includes life, physical, and environmental sciences, which may reflect greater complementarity between these fields and ELA skills compared to more strictly quantitative fields. A one SD increase in ELA scores is associated with a 2.8 percentage point increase in social science BA attainment and a 1.9 percentage point increase in humanities BA attainment. In contrast, the same increase in math scores corresponds to smaller gains of 1.1 and 0.5 percentage points, respectively.

Our results for the HS sample are similar, but provide even more convincing evidence about the link between test scores and degree fields because we estimate coefficients for all four core subjects. We find that among the four test scores, the coefficient for math is largest for math, engineering, and business fields; the coefficient for ELA is largest for humanities fields; the coefficient for science is largest for science fields; and the coefficient for SS is largest for social science fields. Further, the results show that conditional on other test scores, there is a negative relationship between science scores and social science, business, and education fields. Overall, these findings provide strong evidence of students earning degrees in fields in which they have a comparative advantage in skills.

Next, we examine the extent to which educational attainment and degree field mediate the relationship between test scores and earnings from ages 29-32 (Table 6). Column (1) replicates our main specification,[26] while columns (2) through (4) sequentially add controls for on-time college enrollment, BA attainment within six years, and major field of study (from Table 5).

Adding these controls substantially attenuates the test score coefficients, especially for ELA. In the EMS sample, relative to the main estimates, controlling for college enrollment reduces the math and ELA coefficients by 16 percent and 33 percent, respectively; adding BA attainment reduces them by 41 percent (math) and 84 percent (ELA); and including degree field controls leads to total attenuation of 54 percent for math and 67 percent for ELA. Even with all controls included, a one SD increase in math test scores is associated with an earnings increase of approximately $3,700, compared to $700 for ELA.

These patterns suggest that much of the ELA-earnings relationship is mediated by educational pathways, while nearly half of the math-earnings relationship remains unexplained by college enrollment, degree attainment, or field of study. The results for the HS sample show similar patterns. With all the controls included, the coefficients attenuate from $4,700 to $3,000 for math (36 percent), $830 to -$60 for ELA (107 percent), $1,500 to $550 for sci-

[26] The estimates in Table 6, column (1) are nearly identical to those in Table 3, column (8). Minor differences reflect sample restrictions due to missing degree field data in Table 6.

ence (64 percent), and \$2,800 to \$1,200 for SS (56 percent). The estimates in the HS sample suggest that the ELA-earnings relationship is completely explained by educational pathways, while most of the math-earnings relationship is not explained by these mechanisms.

5.2 Heterogeneity Analysis

Heterogeneity by Socioeconomic Status, Race, and Achievement. We begin our heterogeneity discussion by exploring how estimates of the relationship between test scores and outcomes vary by FARMS status, race subgroups, and student achievement terciles. The outcomes of interest are on-time college enrollment, BA attainment within six years, and earnings from ages 29-32. Table 7 presents these estimates for the EMS sample.

The results show that for college enrollment, the test score coefficients are larger for historically disadvantaged subgroups including students who are FARMS, Black, Hispanic, or in the bottom two terciles of achievement. The achievement results are most striking: a one SD increase in math test scores corresponds to at least a 10 p.p. increase in college enrollment in the bottom two achievement terciles while this same increase in math scores only relates to a 3.5 p.p. increase in enrollment in the top achievement tercile. The coefficients for ELA are at least 9.4 p.p. for bottom and middle achievement tercile students, but only 2.5 p.p. for top tercile students.

However, this pattern reverses for BA attainment: the coefficients are typically smaller for FARMS, Black, Hispanic, and bottom achievement tercile students relative to their peers. The coefficients are 4.9 p.p. for math and 2.7 p.p. for ELA for bottom achievement tercile students, 14.9 p.p. and 10 p.p. for middle achievement tercile students, and 10.2 p.p. and 6 p.p. for top achievement tercile students.

For earnings from ages 29-32, the coefficients for math test scores follow the same pattern as the BA attainment coefficients with smaller coefficients for disadvantaged groups, while the coefficients for ELA test scores follow the same pattern as the college enrollment coefficients, with larger coefficients for disadvantaged groups. For math, the coefficients for the achievement terciles from lowest to highest are are \$6,000, \$9,200, and \$8,500, respectively. For ELA, the coefficients for the achievement terciles from lowest to highest are are \$2,400, \$2,700, and \$1,000, respectively.

For all subgroups, the heterogeneity results for earnings have the same pattern as our main results of substantially larger coefficients for math relative to ELA. However, the heterogeneity results provide additional nuance: math test scores are more strongly associated with earnings for higher-achieving students, while ELA test scores show stronger associations for lower-achieving students. The heterogeneity estimates for the HS sample, presented in Table 8, follow a similar pattern of results. Notably, the coefficients are relatively larger for the top two achievement terciles for math, science, and SS test scores, while the coefficients are larger for the bottom two achievement terciles for ELA test scores.

Heterogeneity by Grade and Age of Earnings. Figure 1 shows estimates of the relationship between test scores and outcomes by grade level. There are two broad takeaways from these figures. First, the test score coefficients are typically larger at higher grade levels, suggesting that test scores in later grades are more predictive of long-term outcomes.[27] Second, while the math coefficients are typically only modestly larger than the ELA coefficients for high school graduation, college enrollment, and AA receipt, the math estimates are considerably larger than the ELA coefficients for BA receipt, STEM BA attainment, and earnings from ages 25-28. This discrepancy in coefficients for attainment and earnings outcomes is especially pronounced for 8th graders. A one SD increase in 8th grade math test scores is associated with a 14 p.p. increase in BA attainment, a 10 p.p. increase in STEM BA receipt, and an earnings increase of over \$5,000. In contrast, a one SD increase in 8th grade ELA scores is associated with a 7.5 p.p. increase in BA attainment, a 1.5 p.p. increase in STEM BA receipt, and an earnings increase of about \$1,000.

[27]One exception to the finding that coefficients are larger for higher grade levels is that the coefficient for math in HS (Algebra I) is often smaller than the coefficient for math in 8th grade and typically more comparable to the coefficients for elementary grades. This discrepancy may be driven by the fact that the Algebra I test covers different content and has a different structure than the math assessments in previous grades or that variance in the grade level when the Algebra I test is first taken contributes to a weaker correlation with long-term outcomes.

Next, we examine the relationship between GPAs and outcomes by grade level in Figure 2. Although there is some evidence that the GPA coefficients are larger at higher grade levels, this pattern is less clear relative to the test score coefficients. However, there are few other interesting patterns. First, for general educational attainment outcomes, the math coefficient is the smallest at all grade levels. For 6th-8th grades, ELA, science, and SS GPAs are similarly predictive, while ELA and SS coefficients are much larger than the science coefficient in HS. For example, for the BA receipt outcome, a one SD increase in cumulative high school GPA is associated with a 4.5 p.p. increase for math, 12 p.p. for ELA, 5 p.p. for science, and 9 p.p. for SS. The smaller coefficients for math and science likely reflect differences in course difficulty and grading standards. Students tracked into more rigorous math and science courses are often graded relative to higher-achieving peers, which may weaken the correlation between GPAs in these subjects and general educational attainment outcomes. However, there is a different pattern for the STEM BA and earnings at age 24 outcomes. For STEM BA attainment, the coefficients are larger for math and science GPAs relative to ELA and SS GPAs. For cumulative HS GPA, a one SD increase corresponds to almost a 6 p.p. increase in STEM BA receipt, while the coefficients for the other subjects are about half as large. For earnings at age 24, the GPA coefficients across all subjects are relatively similar in grades 7-8. However, in HS, a one SD increase in cumulative GPA is associated with an earnings increase of nearly $2,000 for math and ELA, compared to only about $1,000 for science and SS.

Figure 3 presents estimates of the relationship between test scores and earnings by age at which earnings are measured. Two clear patterns are apparent in these results. First, the test score coefficients are larger for earnings measured at older ages. This pattern is most pronounced for untransformed earnings but is also evident for log earnings. Second, there is a large gap between math coefficients and the coefficients for other subjects at almost all ages when earnings are measured.

For our EMS sample, we find that a one SD increase in math test scores corresponds to over a \$2,000 (approximately 10 percent) increase in earnings at age 23 and about a \$9,000 (approximately 20 percent) increase in earnings at age 30. In contrast, a one SD increase in ELA scores is associated with about a \$500 (approximately 3 percent) decrease in earnings at age 23 and almost a \$2,500 (approximately 6 percent) increase in earnings at age 30. For our HS sample, we observe similar patterns, although the coefficients are somewhat smaller, likely in part because the multivariate model includes test scores for science and SS in addition to math and ELA. Notably, the coefficients for science and SS are larger than those for ELA, suggesting that these variables may absorb variation that would otherwise be attributed to ELA in a more limited model.

Finally, our estimates of the relationship between GPAs and earnings by age are shown in Figure 4. We focus on the results for our HS sample since we are able to observe earnings measured at older ages for these students. The results for the EMS sample are similar, but the patterns are somewhat less clear due to the lack of availability of earnings outcomes at older ages. Similar to the test score results, we find that the GPA coefficients are larger for earnings measured at older ages, although the increase in magnitude is less pronounced than in the test score estimates. We also find that the math and ELA coefficients are modestly larger than the science and SS coefficients at almost all ages. The results show that a one SD increase in math and ELA GPAs corresponds to about a \$1,000 (approximately 4 percent) increase in earnings at age 23 and about a \$3,000 (approximately 7 percent) increase in earnings at age 28. In contrast, a one SD increase in science and SS GPAs is associated with about a \$500 (approximately 2 percent) increase in earnings at age 23 and about a \$2,000 (approximately 5 percent) increase in earnings at age 28.

5.3 Teacher Value-Added Effects on Long-Term Outcomes

We complement our descriptive analyses by using test score TVA measures to estimate causal effects on outcomes. Table 9 presents these estimates for our TVA sample: Panel A reports results from univariate models that include only one TVA measure (math or ELA), while Panel B shows results from multivariate models that include both math and ELA TVA simultaneously. For the univariate results, we find limited evidence that test score TVA has significant or substantive effects on HS graduation, college enrollment, or AA degree receipt.[28] However, we find that a one SD increase in test score TVA increases BA receipt within four years by about 0.6 p.p. for both math and ELA. These estimates are no longer significant for BA receipt within 6 years; the math TVA coefficient attenuates substantially while the ELA TVA estimate is nearly identical but less precisely estimated. The results also show that a one SD increase in math TVA boosts STEM BA receipt within six years by 1.3 p.p. and leads to a marginally significant increase in earnings from ages 22-24 of about $160 (0.7 percent). In contrast, the estimated impact of ELA TVA on STEM BA attainment is nearly zero while the earnings effect is insignificantly negative.

The estimates for the multivariate model show the same pattern, likely due in part to the fact that math and ELA TVA are weakly correlated. The estimates for on-time BA receipt attenuate slightly: a one SD increase in TVA increases BA receipt within four years by about 0.56 p.p. for math and 0.4 p.p. for ELA. The estimates for BA attainment within six years are similar to the estimates from the univariate model. For STEM BA receipt, the math TVA estimate is nearly identical to the univariate estimate while the ELA TVA estimate becomes insignificantly negative. Finally, the earnings estimates are also similar although the math TVA estimate increases to $184 while the ELA TVA coefficient becomes slightly more negative.

Overall, we find some alignment between our descriptive and TVA-based analyses. The estimated effects on educational attainment, such as BA receipt within four years, are similar for math and ELA, while the effects on STEM BA attainment and earnings are larger for math. This consistency lends credibility to the descriptive patterns when examined through a causal lens. However, two caveats remain. First, the TVA estimates should be interpreted with caution due to modest forecast bias, as shown in Appendix Table A3. Second, with the

[28]There is a small, negative, marginally significant coefficient for math TVA on HS graduation, but this result may be due to chance and is suggestive of only a 0.1 p.p. decrease.

exception of the estimate for STEM BA attainment associated with math TVA, the relatively small magnitudes of the TVA estimates suggest that improving students' test scores may have more limited causal impacts than descriptive estimates imply.

5.4 Robustness: Employment and Earnings Outcomes

Thus far, our analyses have focused on conditional earnings as the primary labor market outcome. To assess robustness, we examine the relationship between test scores and three labor market measures: (i) employment, (ii) unconditional earnings, and (iii) conditional earnings. We define employment as having positive earnings reported in the Maryland UI data. Unconditional earnings are calculated by imputing zero earnings for individuals who are missing UI earnings data, while conditional earnings include only those with observed positive earnings.

Table A4 presents estimates for each outcome. Since the patterns are similar across outcomes measured at ages 22-24, 25-28, and 29-32, we focus on the results for ages 29-32. In the EMS sample, a one standard deviation increase in either math or ELA test scores is associated with a roughly 3 p.p. decrease in employment. This negative association likely reflects the fact that higher-achieving students are more likely to migrate out-of-state. These higher-achieving students would likely have higher earnings on average if we could observe their earnings, so imputing zeros for missing earnings in the unconditional earnings measure introduces negative bias relative to conditional earnings. For example, a one SD increase in math scores corresponds to a $2,800 increase in unconditional earnings but an $8,100 increase in conditional earnings. For ELA, there is also a gap in the estimates: a one SD increase predicts a $700 decrease in unconditional earnings but a $2,200 increase in conditional earnings. Estimates from the HS sample, including coefficients for science and social studies scores, show similar patterns, with negative associations between test scores and employment and downward bias in unconditional earnings estimates.

Overall, these findings highlight the importance of conditioning on observed earnings when analyzing labor market outcomes. Unconditional measures tend to substantially understate the strength of the relationship between test scores and earnings, while conditional earnings provide a more accurate picture of this correlation. Recent research suggests that even our conditional earnings estimates are likely biased toward zero, as individuals with higher test scores are less likely to be observed in Maryland UI data, and, if observed, would likely have higher earnings on average (Foote and Stange, 2022).

6 Conclusion

A growing body of research has examined how cognitive skills developed in school shape students' long-term economic outcomes.[29] Much of this literature relies on longitudinal survey data and finds that both general and subject-specific skills, particularly math, are predictive of earnings. However, prior research often uses small samples from survey data, self-reported outcomes, and limited causal identification strategies. This paper builds on and extends this literature by using population-level administrative data and quasi-experimental methods to provide new evidence on the link between specific academic skills and long-term economic outcomes.

Our findings show that both math and ELA test scores predict college attainment, but math scores are substantially more predictive of STEM degree completion and post-college earnings. We find that a one SD increase in test scores is associated with an increase in average annual earnings at ages 29-32 of approximately 17 percent for math and 4.5 percent for ELA. GPA also predicts earnings, potentially reflecting non-cognitive skills like effort and motivation that are not fully captured by test scores. Patterns across degree fields suggest that students tend to pursue majors aligned with their strongest academic subjects, consistent with comparative advantage. Mediation analysis reveals that most of

[29] Neal and Johnson (1996); Heckman et al. (2006); Lin et al. (2018); Murnane et al. (1995, 2000); Grogger and Eide (1995); Dougherty (2003); Deming (2017); Hanushek et al. (2015) use longitudinal surveys to study the relationship between cognitive skills and earnings. Chetty et al. (2014b) uses administrative data and quasi-experimental teacher value-added methods to estimate impacts on postsecondary and earnings outcomes.

the ELA test-earnings relationship operates through educational attainment and degree field, while a substantial portion of the math test-earnings link remains unexplained. Importantly, heterogeneity analyses show that math scores are more predictive of earnings than ELA scores across all student subgroups. However, compared to their more advantaged peers, historically disadvantaged students exhibit a weaker association between math scores and earnings and a somewhat stronger association between ELA scores and earnings. Finally, teacher value-added results corroborate the descriptive findings: math teacher value-added has larger impacts on both STEM attainment and earnings than ELA value-added.

Taken together, these findings offer novel insights into how different types of academic skills relate to long-run economic success. They suggest that policies aiming to promote economic mobility may have differential effects depending on which skills are targeted and which student subgroups are served. This underscores the need for a multi-faceted approach to improving outcomes and reducing inequality. Investing in more rigorous math requirements and instruction may be one important policy lever (Goodman, 2019). Complementing this with supports such as tutoring and mentoring for disadvantaged students can further enhance long-term outcomes (Cortes et al., 2015). Leveraging multiple policy tools will be critical to helping disadvantaged students translate skill gains into postsecondary and labor market success, thereby reducing inequality and promoting upward mobility.

References

Bettinger, Eric P, Brent J Evans, and Devin G Pope, "Improving College Performance and Retention the Easy Way: Unpacking the ACT Exam," *American Economic Journal: Economic Policy*, 2013, *5* (2), 26–52.

BLS, "All Employees, Federal [CES9091000001]," https://fred.stlouisfed.org/series/CES9091000001/, 2025. Last accessed on July 7, 2025.

— , "Civilian Labor Force Level [CLF16OV]," https://fred.stlouisfed.org/series/CLF16OV/, 2025. Last accessed on July 7, 2025.

— , "Employment-Population Ratio - 25-54 Yrs. [LNS12300060]," https://fred.stlouisfed.org/series/LNS12300060, 2025. Last accessed on July 7, 2025.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

— , John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, September 2014, *104* (9), 2633–79.

Cortes, Kalena E, Joshua S Goodman, and Takako Nomi, "Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra," *Journal of Human Resources*, 2015, *50* (1), 108–158.

Deming, David J, "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

DHS, "DHS STEM Designated Degree Program List," https://www.ice.gov/doclib/sevis/pdf/stemList2022.pdf, 2023. Last accessed on July 7, 2025.

Dougherty, Christopher, "Numeracy, Literacy and Earnings: Evidence from the National Longitudinal Survey of Youth," *Economics of Education Review*, 2003, *22* (5), 511–521.

Foote, Andrew and Kevin M Stange, "Attrition from Administrative Data: Problems and Solutions with an Application to Postsecondary Education," 2022. NBER Working Paper No. 30232.

Goodman, Joshua, "The Labor of Division: Returns to Compulsory High School Math Coursework," *Journal of Labor Economics*, 2019, *37* (4), 1141–1182.

Grogger, Jeff and Eric Eide, "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, 1995, pp. 280–310.

Hanushek, Eric A, Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann, "Returns to Skills Around the World: Evidence from PIAAC," *European Economic Review*, 2015, *73*, 103–130.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411–482.

Jackson, C Kirabo, "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

Lin, Dajun, Randall Lutter, and Christopher J. Ruhm, "Cognitive performance and Labour Market Outcomes," *Labour Economics*, 2018, *51*, 121–135.

Maryland State Archives, "Maryland Employment - Workforce," https://www.msa.maryland.gov/msa/mdmanual/01glance/economy/html/labor.html, 2025. Last accessed on July 7, 2025.

Minneapolis Fed, "Consumer Price Index, 1913-," https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-, 2025. Last accessed on July 7, 2025.

MSDE, "Technical Report 2003 for Maryland School Assessment," https://msa.maryland.gov/megafile/msa/speccol/sc5300/sc5339/000113/001000/001343/unrestricted/20054117e.pdf, 2003. Last accessed on July 7, 2025.

_ , "Maryland High School Assessment and Modified High School Assessment 2011 Technical Report," https://www.marylandpublicschools.org/about/Documents/DCAA/PRT/MD2011HSATechReport.pdf, 2012. Last accessed on July 7, 2025.

_ , "Report on Local, State, and Federally Mandated Assessments in Maryland," https://marylandpublicschools.org/Documents/commissiononassessments/LocalStateFederallyMandatedAssessmentsMD08312015Report.pdf, 2015. Last accessed on July 7, 2025.

_ , "Maryland School Assessment (MSA) Science Technical Report 2016," https://www.marylandpublicschools.org/about/Documents/DAAIT/Assessment/MISA/msa_sci_5_8_tech_report_2016.pdf, 2016. Last accessed on July 7, 2025.

_ , "Maryland Student Records System Manual," https://marylandpublicschools.org/about/Documents/OCP/Publications/2020-Maryland-Student-Records-System-Manual.pdf, 2020. Last accessed on July 7, 2025.

_ , "Maryland Integrated Science Assessments (MISA)," https://marylandpublicschools.org/about/Pages/DAAIT/Assessment/MISA/index.aspx, 2025. Last accessed on July 7, 2025.

Murnane, Richard J., John B. Willett, and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 1995, *77* (2), 251–266.

_ , _ , **Yves Duhaldeborde, and John H. Tyler**, "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?," *Journal of Policy Analysis and Management*, 2000, *19* (4), 547–568.

NCES, "The Classification of Instructional Programs (CIP)," https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=56, 2020. Last accessed on July 7, 2025.

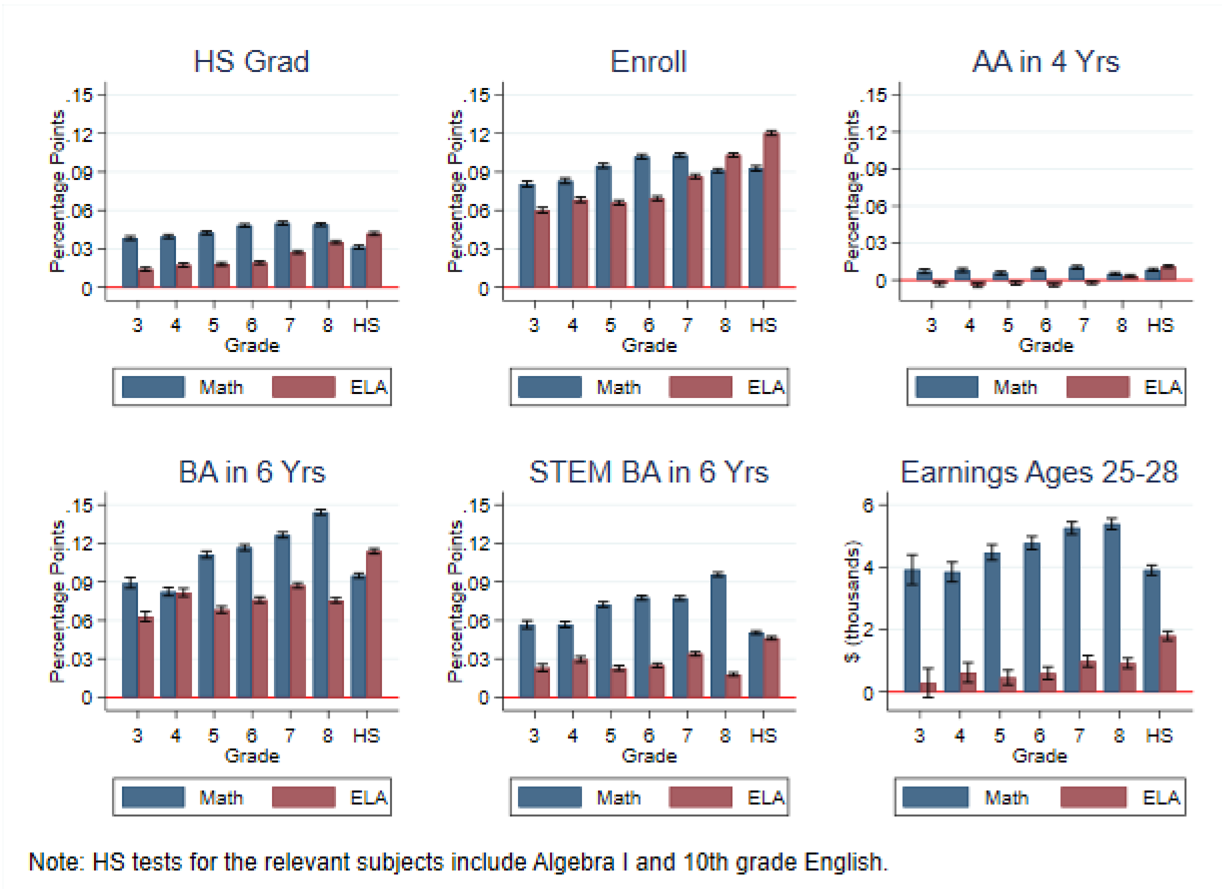
Neal, Derek A. and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 1996, *104* (5), 869–895.

Petek, Nathan and Nolan G Pope, "The Multidimensional Impact of Teachers on Students," *Journal of Political Economy*, 2023, *131* (4), 1057–1107.

U.S. Congress, "H.R.1 - No Child Left Behind Act of 2001," https://www.congress.gov/bill/107th-congress/house-bill/1, 2001. Last accessed on July 7, 2025.

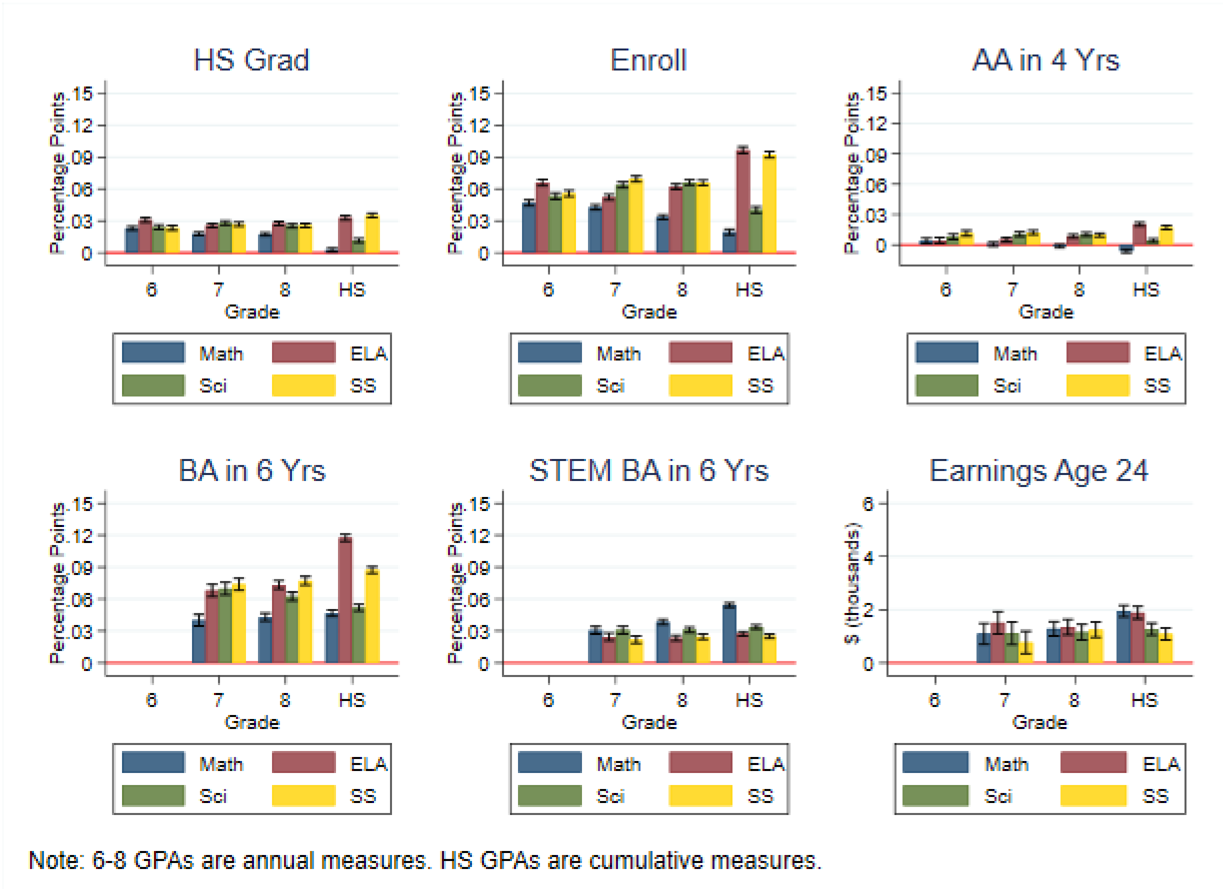
Figures and Tables

Figure 1: Relationship Between Test Scores and Outcomes by Grade



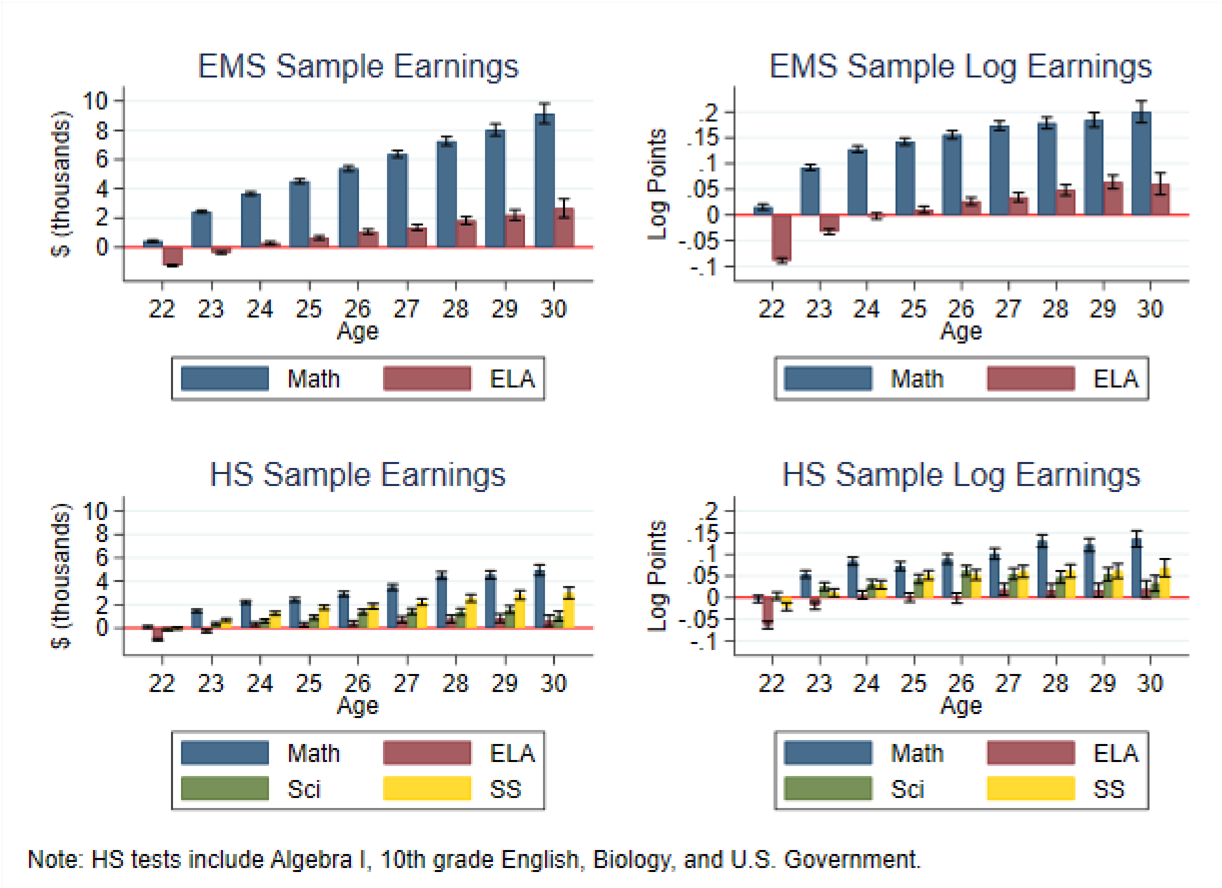
Notes: This figure shows estimates for the multivariate relationship between test scores and long-term outcomes by grade level. We use the EMS sample to construct estimates for grades 3-8 and the HS sample to construct estimates for HS. The samples are described in Section 3. Outcomes include educational attainment measures and earnings from ages 25-28. The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1, with only standardized math and ELA test scores as the main explanatory variables. The demographic controls include indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and imputed values for demographic variables. High school graduation is within four years and college enrollment is on-time relative to 9th grade enrollment. Average annual earnings from ages 25-28 are conditional on having positive observed earnings and measured in 2024 dollars. Robust 95 percent confidence intervals clustered at the student level are reported.

Figure 2: Relationship Between GPAs and Outcomes by Grade



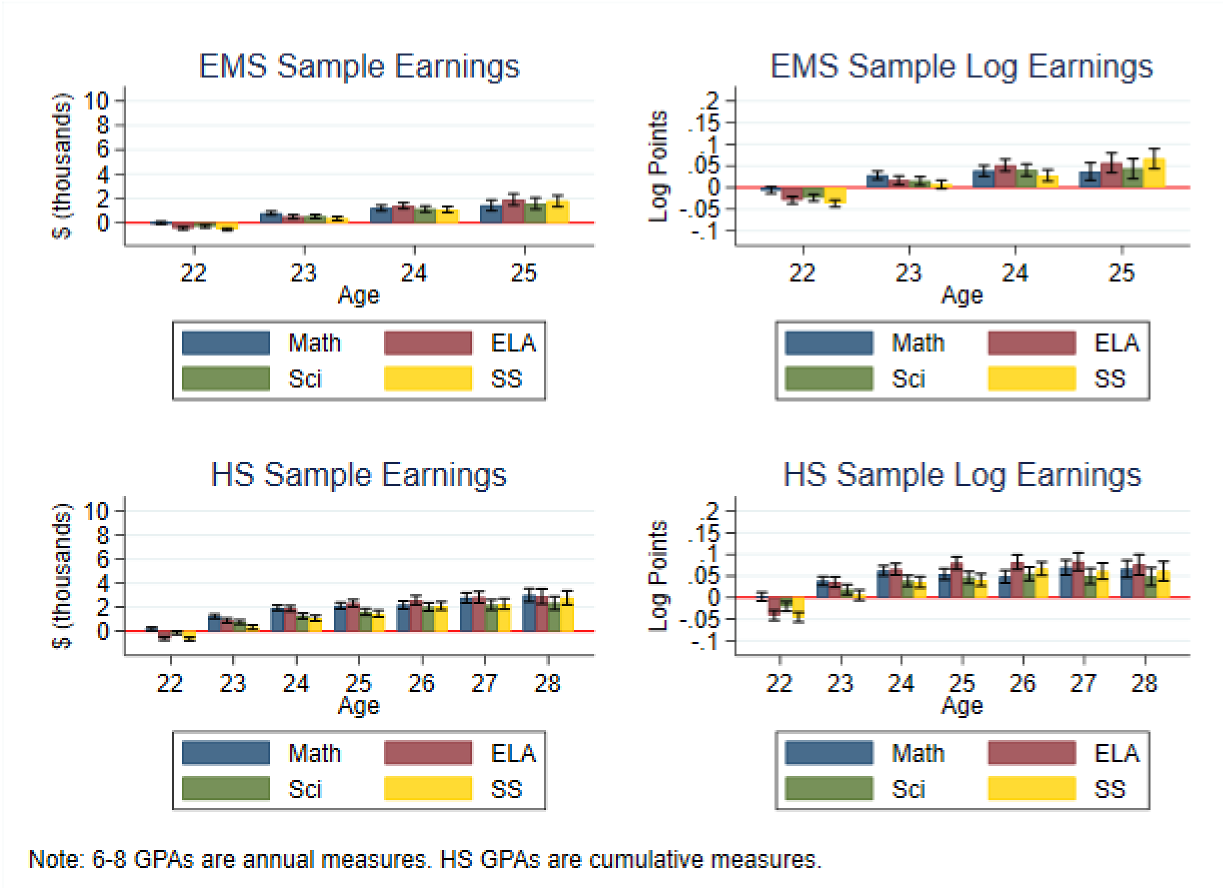
Notes: This figure shows estimates for the multivariate relationship between GPAs and long-term outcomes by grade level. We use the EMS sample to construct estimates for grades 6-8 and the HS sample to construct estimates for HS. The samples are described in Section 3. Outcomes include educational attainment measures and earnings at age 24. The specification is a multivariate GPA model with demographic controls and standardized GPAs in the four core subjects as the main explanatory variables. GPAs for grades 6-8 are annual measures and GPAs for HS are cumulative measures. High school graduation is within four years and college enrollment is on-time relative to 9th grade enrollment. Annual earnings at age 24 are conditional on having positive observed earnings and measured in 2024 dollars. Robust 95 percent confidence intervals clustered at the student level are reported.

Figure 3: Relationship Between Test Scores and Earnings by Age



Notes: This figure shows estimates for the multivariate relationship between test scores and earnings by the age when earnings are measured. Estimates for the EMS sample are shown in the top two sub-figures and estimates for the HS sample are shown in the bottom two sub-figures. The samples are described in Section 3. Outcomes are annual earnings at ages 22-30, respectively, conditional on having positive observed earnings. Earnings are measured in 2024 dollars, with the two left sub-figures showing estimates in thousands of 2024 dollars and the two right sub-figures showing estimates for log earnings. The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1. Robust 95 percent confidence intervals clustered at the student level are reported.

Figure 4: Relationship Between GPAs and Earnings by Age



Notes: This figure shows estimates for the multivariate relationship between GPAs and earnings by the age when earnings are measured. Estimates for the EMS sample are shown in the top two sub-figures and estimates for the HS sample are shown in the bottom two sub-figures. The samples are described in Section 3. Outcomes are annual earnings at ages 22-28, respectively, conditional on having positive observed earnings. Earnings are measured in 2024 dollars, with the two left sub-figures showing estimates for thousands of dollars and the two right sub-figures showing estimates for log earnings. The specification is a multivariate GPA model with demographic controls and standardized GPAs in the four core subjects as the main explanatory variables. GPAs for grades 6-8 are annual measures and GPAs for HS are cumulative measures. Robust 95 percent confidence intervals clustered at the student level are reported.

Table 1: Summary Statistics for EMS and HS Samples

	EMS Sample			HS Sample		
	(1) Mean	(2) SD	(3) N	(4) Mean	(5) SD	(6) N
Panel A: Demographics						
Female	0.497	0.500	3,281,607	0.503	0.500	457,157
FARMS	0.419	0.493	3,281,607	0.364	0.481	457,157
Black	0.342	0.474	3,281,607	0.352	0.477	457,157
Hispanic	0.131	0.337	3,281,607	0.118	0.322	457,157
White	0.416	0.493	3,281,607	0.428	0.495	457,157
Asian	0.062	0.241	3,281,607	0.056	0.230	457,157
Other Race	0.049	0.217	3,281,607	0.047	0.211	457,157
ELL	0.043	0.202	3,281,607	0.036	0.186	457,157
SPED	0.133	0.340	3,281,607	0.127	0.333	457,157
Panel B: Explanatory Variables						
Comp Score Math	0.045	0.987	3,281,607	0.076	0.921	457,157
Comp Score ELA	0.009	0.996	3,281,607	0.024	0.955	457,157
Comp Score Sci	0.038	0.972	966,205	0.009	0.953	457,157
Comp Score SS				0.012	0.944	457,157
GPA Math	-0.013	0.969	1,025,365	-0.042	0.929	344,181
GPA ELA	-0.001	0.964	1,035,480	-0.003	0.916	355,160
GPA Sci	0.001	0.968	1,027,231	-0.026	0.919	333,653
GPA SS	0.005	0.965	1,020,668	0.001	0.910	339,639
Panel C: Outcomes						
HS Grad in 4 Yrs	0.874	0.332	3,281,607	0.917	0.275	457,157
Enroll in College	0.642	0.480	2,999,633	0.658	0.474	457,157
Persist in College	0.558	0.497	2,674,895	0.565	0.496	457,157
AA in 2 Yrs	0.025	0.157	2,674,895	0.023	0.151	457,157
AA in 4 Yrs	0.073	0.260	2,025,934	0.076	0.264	427,798
BA in 4 Yrs	0.208	0.406	2,025,934	0.188	0.390	427,798
BA in 6 Yrs	0.318	0.466	1,414,238	0.290	0.454	339,663
STEM BA in 4 Yrs	0.063	0.244	2,025,533	0.055	0.228	427,314
STEM BA in 6 Yrs	0.104	0.305	1,413,813	0.087	0.282	339,101
Employed Ages 22-24	0.829	0.377	1,800,494	0.840	0.367	382,751
Employed Ages 25-28	0.723	0.447	995,098	0.757	0.429	260,277
Employed Ages 29-32	0.602	0.489	152,752	0.646	0.478	124,378
Earnings Ages 22-24	23,138	17,831	1,492,104	23,263	17,617	321,399
Earnings Ages 25-28	36,697	26,399	719,675	35,528	25,478	197,068
Earnings Ages 29-32	48,249	34,439	91,999	45,442	32,611	80,362

Notes: Summary statistics are shown for the elementary and middle school (EMS) sample in columns (1)-(3) and the high school (HS) sample in columns (4)-(6). The samples are described in Section 3. Composite test score and GPA variables are standardized in the population to have mean zero and standard deviation one. College enrollment and persistence are "on-time" outcomes relative to the first observed year of enrollment in 9th grade in Maryland public schools. Employed is an indicator for having positive observed earnings. Average annual earnings are conditional on having positive observed earnings and measured in 2024 dollars.

Table 2: Univariate Relationship Between Test Scores and Educational Attainment and Earnings

	(1) HS Grad	(2) Enroll	(3) AA	(4) BA	(5) STEM BA	(6) Earn 22-24	(7) Earn 25-28	(8) Earn 29-32
Panel A: EMS Sample								
Comp Score Math	0.0601***	0.1436***	0.0064***	0.1690***	0.0937***	917***	5,376***	9,500***
	(0.0004)	(0.0005)	(0.0004)	(0.0007)	(0.0006)	(35)	(68)	(175)
Adj R-squared	0.0836	0.2159	0.0158	0.2509	0.1436	0.0389	0.1139	0.1514
Comp Score ELA	0.0523***	0.1377***	0.0032***	0.1502***	0.0729***	99***	3,726***	6,974***
	(0.0004)	(0.0005)	(0.0004)	(0.0007)	(0.0006)	(32)	(62)	(158)
Adj R-squared	0.0784	0.2122	0.0155	0.2363	0.1201	0.0372	0.1008	0.1325
Outcome Mean	[.8743]	[.6416]	[.0728]	[.3178]	[.1037]	[23,138]	[36,697]	[48,249]
N	3,281,607	2,999,633	2,025,934	1,414,238	1,413,813	1,492,104	719,675	91,999
Panel B: HS Sample								
Comp Score Math	0.0561***	0.1642***	0.0150***	0.1622***	0.0778***	882***	4,959***	7,806***
	(0.0006)	(0.0008)	(0.0004)	(0.0009)	(0.0006)	(40)	(70)	(141)
Adj R-squared	0.0611	0.1935	0.0181	0.2077	0.1059	0.0414	0.1064	0.1244
Comp Score ELA	0.0589***	0.1709***	0.0157***	0.1638***	0.0727***	187***	3,847***	6,236***
	(0.0006)	(0.0008)	(0.0004)	(0.0009)	(0.0006)	(38)	(67)	(135)
Adj R-squared	0.0661	0.2064	0.0184	0.2189	0.1051	0.0400	0.0992	0.1140
Comp Score Sci	0.0585***	0.1600***	0.0150***	0.1568***	0.0804***	568***	4,214***	6,239***
	(0.0006)	(0.0008)	(0.0004)	(0.0009)	(0.0006)	(37)	(65)	(129)
Adj R-squared	0.0662	0.1975	0.0183	0.2122	0.1155	0.0406	0.1027	0.1162
Comp Score SS	0.0624***	0.1762***	0.0155***	0.1681***	0.0756***	584***	4,401***	6,957***
	(0.0006)	(0.0008)	(0.0004)	(0.0008)	(0.0006)	(37)	(66)	(136)
Adj R-squared	0.0695	0.2113	0.0184	0.2246	0.1093	0.0407	0.1045	0.1203
Outcome Mean	[.9174]	[.658]	[.0756]	[.2901]	[.087]	[23,263]	[35,528]	[45,442]
N	457,157	457,157	427,798	339,663	339,101	321,399	197,068	80,362

Notes: This table shows estimates for the univariate relationship between test scores and long-term outcomes. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 3. Outcomes include educational attainment measures in columns (1)-(5) and earnings measures in columns (6)-(8). The specification is based on our main test score model with demographic controls in Section 4.1, Equation 1, but with only one test score variable included in the regression. Thus each cell in this table represents a separate regression. The demographic controls include indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and imputed values for demographic variables. High school graduation is within four years, college enrollment is on-time relative to 9th grade enrollment, AA receipt is within four years, and BA receipt outcomes are within six years. Average annual earnings are conditional on having positive observed earnings and measured in 2024 dollars. Robust standard errors are clustered at the student level.

*** p<0.01, ** p<0.05, * p<0.10

Table 3: Multivariate Relationship Between Test Scores and Educational Attainment and Earnings

	(1) HS Grad	(2) Enroll	(3) AA	(4) BA	(5) STEM BA	(6) Earn 22-24	(7) Earn 25-28	(8) Earn 29-32
Panel A: EMS Sample								
Comp Score Math	0.0446***	0.0922***	0.0073***	0.1183***	0.0771***	1,413***	4,874***	8,096***
	(0.0004)	(0.0006)	(0.0004)	(0.0008)	(0.0006)	(38)	(74)	(207)
Comp Score ELA	0.0237***	0.0791***	-0.0014***	0.0773***	0.0253***	-768***	775***	2,167***
	(0.0004)	(0.0006)	(0.0004)	(0.0008)	(0.0006)	(34)	(66)	(183)
Adj R-squared	0.0857	0.2273	0.0158	0.2630	0.1466	0.0397	0.1142	0.1530
Outcome Mean	[.8743]	[.6416]	[.0728]	[.3178]	[.1037]	[23,138]	[36,697]	[48,249]
N	3,281,607	2,999,633	2,025,934	1,414,238	1,413,813	1,492,104	719,675	91,999
Panel B: HS Sample								
Comp Score Math	0.0165***	0.0574***	0.0054***	0.0587***	0.0299***	910***	2,873***	4,676***
	(0.0008)	(0.0011)	(0.0006)	(0.0010)	(0.0007)	(51)	(88)	(174)
Comp Score ELA	0.0200***	0.0668***	0.0065***	0.0642***	0.0197***	-710***	369***	841***
	(0.0009)	(0.0011)	(0.0006)	(0.0011)	(0.0008)	(54)	(90)	(181)
Comp Score Sci	0.0184***	0.0353***	0.0043***	0.0361***	0.0370***	241***	1,159***	1,517***
	(0.0009)	(0.0012)	(0.0006)	(0.0011)	(0.0008)	(53)	(89)	(164)
Comp Score SS	0.0270***	0.0747***	0.0053***	0.0692***	0.0209***	397***	1,852***	2,787***
	(0.0010)	(0.0012)	(0.0006)	(0.0011)	(0.0007)	(55)	(93)	(188)
Adj R-squared	0.0776	0.2348	0.0192	0.2502	0.1289	0.0419	0.1120	0.1319
Outcome Mean	[.9174]	[.658]	[.0756]	[.2901]	[.087]	[23,263]	[35,528]	[45,442]
N	457,157	457,157	427,798	339,663	339,101	321,399	197,068	80,362

Notes: This table shows estimates for the multivariate relationship between test scores and long-term outcomes. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 3. Outcomes include educational attainment measures in columns (1)-(5) and earnings measures in columns (6)-(8). The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1, so each column in a panel represents a separate regression. High school graduation is within four years, college enrollment is on-time relative to 9th grade enrollment, AA receipt is within four years, and BA receipt outcomes are within six years. Average annual earnings are conditional on having positive observed earnings and measured in 2024 dollars. Robust standard errors are clustered at the student level.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 4: Multivariate Relationship Between Test Scores and GPAs and Outcomes

	(1) HS Grad	(2) Enroll	(3) AA	(4) BA	(5) STEM BA	(6) Earn 22-24	(7) Earn 25-28	(8) Earn 29-32
Panel A: EMS Sample								
Comp Score Math	0.0051***	0.0286***	-0.0097***	0.0773***	0.0765***	273***	3,471***	
	(0.0006)	(0.0009)	(0.0008)	(0.0020)	(0.0017)	(70)	(282)	
Comp Score ELA	0.0044***	0.0511***	-0.0058***	0.0393***	0.0135***	-1,187***	-902***	
	(0.0006)	(0.0009)	(0.0008)	(0.0018)	(0.0014)	(66)	(255)	
GPA Math	0.0173***	0.0268***	0.0032***	0.0225***	0.0185***	204***	823***	
	(0.0006)	(0.0008)	(0.0007)	(0.0017)	(0.0011)	(54)	(224)	
GPA ELA	0.0269***	0.0508***	0.0081***	0.0599***	0.0155***	16	1,779***	
	(0.0006)	(0.0009)	(0.0007)	(0.0017)	(0.0011)	(57)	(238)	
GPA Sci	0.0243***	0.0486***	0.0126***	0.0414***	0.0128***	155***	1,059***	
	(0.0006)	(0.0009)	(0.0007)	(0.0017)	(0.0011)	(56)	(241)	
GPA SS	0.0240***	0.0522***	0.0132***	0.0581***	0.0104***	-9	1,432***	
	(0.0006)	(0.0009)	(0.0007)	(0.0017)	(0.0011)	(56)	(237)	
Adj R-squared	0.0974	0.2707	0.0186	0.3253	0.1736	0.0317	0.1101	
Outcome Mean	[.8947]	[.6674]	[.0775]	[.3542]	[.1185]	[23,208]	[38,024]	
N	978,252	873,674	417,429	143,436	143,401	290,425	26,944	
Panel B: HS Sample								
Comp Score Math	0.0033***	0.0297***	-0.0013*	0.0325***	0.0205***	422***	2,051***	2,917***
	(0.0009)	(0.0012)	(0.0008)	(0.0013)	(0.0010)	(65)	(127)	(532)
Comp Score ELA	0.0006	0.0260***	-0.0014*	0.0304***	0.0101***	-1,182***	-548***	87
	(0.0009)	(0.0013)	(0.0008)	(0.0013)	(0.0010)	(68)	(128)	(596)
Comp Score Sci	0.0015	0.0050***	-0.0023***	-0.0064***	0.0262***	13	-108	1,125*
	(0.0010)	(0.0013)	(0.0008)	(0.0013)	(0.0011)	(69)	(139)	(581)
Comp Score SS	0.0140***	0.0479***	-0.0015**	0.0271***	0.0065***	312***	840***	2,405***
	(0.0011)	(0.0014)	(0.0008)	(0.0013)	(0.0010)	(70)	(132)	(603)
GPA Math	0.0004	0.0052***	-0.0052***	0.0363***	0.0453***	520***	1,805***	1,527***
	(0.0009)	(0.0014)	(0.0009)	(0.0015)	(0.0010)	(69)	(130)	(435)
GPA ELA	0.0312***	0.0872***	0.0209***	0.1091***	0.0222***	389***	2,339***	2,221***
	(0.0011)	(0.0016)	(0.0010)	(0.0016)	(0.0011)	(76)	(145)	(498)
GPA Sci	0.0087***	0.0248***	0.0053***	0.0409***	0.0241***	377***	1,494***	2,196***
	(0.0010)	(0.0015)	(0.0010)	(0.0016)	(0.0010)	(73)	(135)	(448)
GPA SS	0.0308***	0.0704***	0.0185***	0.0711***	0.0133***	79	1,399***	2,280***
	(0.0011)	(0.0016)	(0.0010)	(0.0016)	(0.0010)	(74)	(138)	(460)
Adj R-squared	0.0955	0.3028	0.0235	0.3718	0.1881	0.0381	0.1322	0.2013
Outcome Mean	[.9345]	[.6867]	[.0788]	[.3338]	[.1105]	[23,646]	[37,887]	[52,042]
N	319,470	319,470	291,985	206,941	206,915	215,793	101,746	7,734

Notes: Robust standard errors are clustered at the student level. *** p<0.01, ** p<0.05, * p<0.10

Table 5: Multivariate Relationship Between Test Scores and Degree Field

	(1) STEM	(2) Math/ Stats/CS	(3) Engi/ Arch	(4) Sci	(5) Soc Sci	(6) Hum	(7) Health	(8) Bus	(9) Edu
Panel A: EMS Sample									
Comp Score Math	0.0771***	0.0253***	0.0274***	0.0229***	0.0109***	0.0052***	0.0063***	0.0174***	0.0028***
	(0.0006)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0004)	(0.0003)	(0.0004)	(0.0003)
Comp Score ELA	0.0253***	0.0040***	0.0050***	0.0139***	0.0276***	0.0192***	0.0047***	0.0032***	0.0007***
	(0.0006)	(0.0003)	(0.0003)	(0.0004)	(0.0005)	(0.0004)	(0.0003)	(0.0004)	(0.0002)
Adj R-squared	0.1466	0.0602	0.0497	0.0472	0.0425	0.0275	0.0257	0.0219	0.0175
Outcome Mean	[.1037]	[.0347]	[.0254]	[.0391]	[.074]	[.0362]	[.0227]	[.0413]	[.0176]
N	1,413,813	1,413,813	1,413,813	1,413,813	1,413,813	1,413,813	1,413,813	1,413,813	1,413,813
Panel B: HS Sample									
Comp Score Math	0.0299***	0.0105***	0.0123***	0.0068***	0.0061***	0.0004	0.0033***	0.0145***	0.0040***
	(0.0007)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0004)	(0.0003)	(0.0005)	(0.0003)
Comp Score ELA	0.0197***	0.0063***	0.0039***	0.0078***	0.0187***	0.0152***	0.0036***	0.0051***	0.0022***
	(0.0008)	(0.0005)	(0.0004)	(0.0005)	(0.0007)	(0.0005)	(0.0004)	(0.0005)	(0.0003)
Comp Score Sci	0.0370***	0.0072***	0.0109***	0.0183***	-0.0022***	0.0050***	0.0041***	-0.0040***	-0.0008***
	(0.0008)	(0.0005)	(0.0004)	(0.0005)	(0.0006)	(0.0005)	(0.0003)	(0.0004)	(0.0003)
Comp Score SS	0.0209***	0.0073***	0.0045***	0.0082***	0.0247***	0.0107***	0.0027***	0.0080***	0.0011***
	(0.0007)	(0.0004)	(0.0004)	(0.0005)	(0.0006)	(0.0005)	(0.0003)	(0.0005)	(0.0003)
Adj R-squared	0.1289	0.0490	0.0411	0.0474	0.0439	0.0308	0.0247	0.0207	0.0187
Outcome Mean	[.087]	[.0287]	[.0207]	[.0341]	[.0692]	[.035]	[.0205]	[.0388]	[.0179]
N	339,101	339,101	339,101	339,101	339,101	339,101	339,101	339,101	339,101

Notes: This table shows estimates for the multivariate relationship between test scores and BA degree field outcomes within six years. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples and outcomes are described in Section 3. The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1, so each column in a panel represents a separate regression. Robust standard errors are clustered at the student level.
*** p<0.01, ** p<0.05, * p<0.10

Table 6: Test Scores and Earnings: Mediation Analysis Across Specifications

	(1) Main	(2) Ctrl Enroll	(3) Ctrl Enroll & BA	(4) Ctrl Enroll, BA, & Field
Panel A: EMS Sample				
Comp Score Math	8,114***	6,812***	4,763***	3,732***
	(208)	(210)	(206)	(202)
Comp Score ELA	2,152***	1,433***	339*	700***
	(183)	(183)	(180)	(177)
Adj R-squared	0.1531	0.1656	0.2060	0.2362
Outcome Mean	[48238]	[48238]	[48238]	[48238]
N	91,881	91,881	91,881	91,881
Panel B: HS Sample				
Comp Score Math	4,685***	4,129***	3,341***	2,988***
	(174)	(173)	(168)	(165)
Comp Score ELA	833***	371**	-159	-58
	(181)	(180)	(175)	(173)
Comp Score Sci	1,524***	1,143***	678***	553***
	(164)	(164)	(160)	(159)
Comp Score SS	2,778***	2,218***	1,246***	1,228***
	(188)	(187)	(182)	(180)
Adj R-squared	0.1321	0.1434	0.1851	0.2089
Outcome Mean	[45425]	[45425]	[45425]	[45425]
N	80,119	80,119	80,119	80,119

Notes: This table shows estimates for the multivariate relationship between test scores and earnings using different specifications. The earnings outcome is average annual earnings from ages 29-32 conditional on having positive observed earnings and measured in 2024 dollars. Column (1) shows our main multivariate test score model specification from Section 4.1, Equation 1. Column (2) adds a control for on-time college enrollment, column (3) adds a control for BA receipt within six years, and column (4) adds controls for the major fields in Table 5. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 3. Robust standard errors are clustered at the student level.
*** p<0.01, ** p<0.05, * p<0.10

Table 7: Heterogeneity Estimates of the Relationship Between Test Scores and Outcomes for EMS Sample

	(1) FARMS	(2) Non-FARMS	(3) Black	(4) Hispanic	(5) White	(6) Asian	(7) Q1	(8) Q2	(9) Q3
Panel A: Enroll in College									
Comp Score Math	0.1018***	0.0854***	0.0994***	0.0875***	0.0958***	0.0393***	0.1043***	0.1141***	0.0347***
	(0.0010)	(0.0007)	(0.0011)	(0.0019)	(0.0009)	(0.0016)	(0.0012)	(0.0015)	(0.0008)
Comp Score ELA	0.0993***	0.0657***	0.0929***	0.0919***	0.0720***	0.0210***	0.0938***	0.0995***	0.0248***
	(0.0010)	(0.0007)	(0.0011)	(0.0018)	(0.0008)	(0.0015)	(0.0012)	(0.0015)	(0.0007)
Adj R-squared	0.1573	0.1368	0.1879	0.1799	0.2384	0.0647	0.1048	0.0930	0.0561
Outcome Mean	[.4599]	[.7701]	[.5549]	[.5429]	[.6993]	[.9019]	[.3915]	[.6682]	[.8641]
N	1,243,120	1,756,513	1,026,135	380,014	1,262,634	183,495	995,705	1,005,580	998,348
Panel B: BA in 6 Yrs									
Comp Score Math	0.0743***	0.1393***	0.0938***	0.1018***	0.1341***	0.1234***	0.0491***	0.1494***	0.1024***
	(0.0011)	(0.0011)	(0.0013)	(0.0024)	(0.0013)	(0.0033)	(0.0010)	(0.0020)	(0.0016)
Comp Score ELA	0.0530***	0.0869***	0.0625***	0.0708***	0.0885***	0.0635***	0.0269***	0.1001***	0.0603***
	(0.0010)	(0.0010)	(0.0012)	(0.0022)	(0.0012)	(0.0030)	(0.0009)	(0.0020)	(0.0014)
Adj R-squared	0.1319	0.2093	0.1677	0.2015	0.2431	0.1740	0.0543	0.0917	0.1094
Outcome Mean	[.1294]	[.4352]	[.1908]	[.2114]	[.4]	[.6388]	[.0828]	[.2822]	[.5988]
N	542,826	871,412	487,347	155,015	621,338	78,417	472,646	484,618	456,974
Panel C: Earnings Ages 29-32									
Comp Score Math	6,649***	8,680***	7,597***	6,743***	8,397***	8,499***	5,956***	9,185***	8,503***
	(301)	(270)	(310)	(641)	(324)	(1,167)	(332)	(472)	(530)
Comp Score ELA	2,557***	1,902***	2,912***	1,034*	1,559***	3,183***	2,367***	2,679***	1,018**
	(262)	(243)	(266)	(577)	(289)	(1,108)	(273)	(468)	(463)
Adj R-squared	0.1095	0.1158	0.0942	0.0613	0.1077	0.0678	0.0997	0.0717	0.0481
Outcome Mean	[38,384]	[53,857]	[37,449]	[48,731]	[55,696]	[63,157]	[36,008]	[49,028]	[62,490]
N	33,340	58,659	34,685	8,704	40,812	3,671	32,412	33,558	26,029

Notes: This table shows estimates for the multivariate relationship between test scores and outcomes across subgroups for the EMS sample. The subgroups in columns (7)-(9) are achievement terciles from the lowest achievement tercile (Q1) to the highest achievement tercile (Q3). Terciles are defined based on the average of the math and ELA test scores. Educational attainment and earnings outcomes are shown in Panels A-C. The earnings outcome is average annual earnings from ages 29-32 conditional on having positive observed earnings and measured in 2024 dollars. The sample and attainment outcomes are described in Section 3. The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1, so each column in a panel represents a separate regression. Robust standard errors are clustered at the student level.
*** p<0.01, ** p<0.05, * p<0.10

Table 8: Heterogeneity Estimates of the Relationship Between Test Scores and Outcomes for HS Sample

	(1) FARMS	(2) Non-FARMS	(3) Black	(4) Hispanic	(5) White	(6) Asian	(7) Q1	(8) Q2	(9) Q3
Panel A: Enroll in College									
Comp Score Math	0.0530***	0.0594***	0.0492***	0.0569***	0.0706***	0.0238***	0.0440***	0.0861***	0.0294***
	(0.0017)	(0.0014)	(0.0018)	(0.0031)	(0.0017)	(0.0034)	(0.0016)	(0.0027)	(0.0016)
Comp Score ELA	0.0719***	0.0629***	0.0669***	0.0699***	0.0737***	0.0189***	0.0622***	0.0809***	0.0291***
	(0.0018)	(0.0014)	(0.0019)	(0.0033)	(0.0017)	(0.0039)	(0.0017)	(0.0028)	(0.0017)
Comp Score Sci	0.0383***	0.0337***	0.0406***	0.0416***	0.0316***	0.0203***	0.0331***	0.0510***	0.0035*
	(0.0017)	(0.0015)	(0.0018)	(0.0032)	(0.0019)	(0.0043)	(0.0016)	(0.0032)	(0.0018)
Comp Score SS	0.0772***	0.0713***	0.0778***	0.0727***	0.0672***	0.0497***	0.0626***	0.0976***	0.0273***
	(0.0017)	(0.0016)	(0.0019)	(0.0030)	(0.0019)	(0.0050)	(0.0016)	(0.0032)	(0.0021)
Adj R-squared	0.1882	0.1859	0.1955	0.2506	0.2450	0.1122	0.1084	0.0764	0.0421
Outcome Mean	[.4993]	[.7487]	[.5954]	[.5456]	[.7041]	[.8752]	[.3948]	[.6988]	[.8803]
N	166,298	290,859	160,773	53,897	195,445	25,723	152,386	152,386	152,385
Panel B: BA in 6 Yrs									
Comp Score Math	0.0280***	0.0789***	0.0360***	0.0476***	0.0857***	0.0667***	0.0107***	0.1053***	0.0834***
	(0.0011)	(0.0016)	(0.0013)	(0.0025)	(0.0020)	(0.0053)	(0.0007)	(0.0030)	(0.0030)
Comp Score ELA	0.0412***	0.0732***	0.0488***	0.0564***	0.0709***	0.0612***	0.0126***	0.0754***	0.0752***
	(0.0012)	(0.0016)	(0.0014)	(0.0030)	(0.0020)	(0.0056)	(0.0008)	(0.0031)	(0.0029)
Comp Score Sci	0.0265***	0.0425***	0.0325***	0.0386***	0.0420***	0.0646***	0.0088***	0.0597***	0.0366***
	(0.0011)	(0.0017)	(0.0012)	(0.0026)	(0.0022)	(0.0063)	(0.0007)	(0.0036)	(0.0032)
Comp Score SS	0.0439***	0.0880***	0.0560***	0.0468***	0.0834***	0.0936***	0.0173***	0.1216***	0.0876***
	(0.0011)	(0.0018)	(0.0014)	(0.0027)	(0.0022)	(0.0068)	(0.0007)	(0.0034)	(0.0037)
Adj R-squared	0.1382	0.2328	0.1670	0.1997	0.2466	0.2527	0.0342	0.0779	0.1103
Outcome Mean	[.1291]	[.3817]	[.1871]	[.1839]	[.3652]	[.5630]	[.0552]	[.2577]	[.5857]
N	123,087	216,576	122,170	37,298	146,380	17,816	117,225	116,450	105,988
Panel C: Earnings Ages 29-32									
Comp Score Math	3,365***	5,771***	3,328***	3,870***	6,960***	4,872***	2,777***	7,349***	7,236***
	(222)	(257)	(213)	(516)	(355)	(1,216)	(182)	(466)	(662)
Comp Score ELA	850***	722***	1,124***	999*	546	758	890***	1,138**	-1,049
	(232)	(268)	(224)	(584)	(342)	(1,457)	(189)	(522)	(677)
Comp Score Sci	1,357***	1,633***	1,860***	914*	852**	1,416	798***	1,666***	2,058***
	(211)	(248)	(192)	(536)	(358)	(1,284)	(168)	(523)	(782)
Comp Score SS	2,602***	2,725***	2,835***	2,235***	2,128***	6,528***	1,165***	4,420***	3,362***
	(244)	(278)	(234)	(600)	(369)	(1,397)	(204)	(510)	(730)
Adj R-squared	0.0973	0.1154	0.0739	0.0628	0.1005	0.0825	0.0801	0.0579	0.0372
Outcome Mean	[38,155]	[49,613]	[36,824]	[46,960]	[52,543]	[57,792]	[35,159]	[47,784]	[60,107]
N	29,251	51,111	33,670	7,575	33,029	2,881	32,715	29,402	18,245

Notes: Heterogeneity estimates for the HS sample; robust standard errors are reported. *** p<0.01, ** p<0.05, * p<0.10

Table 9: Effect of Teacher Test Score Value-Added on Educational Attainment and Earnings

	(1) HS Grad	(2) Enroll	(3) AA 2 Yrs	(4) AA 4 Yrs	(5) BA 4 Yrs	(6) BA 6 Yrs	(7) STEM	(8) Earn 22-24
Panel A: Univariate Models								
Math TVA	-0.0010*	0.0012	-0.0007	-0.0005	0.0063***	0.0019	0.0134***	164*
	(0.0006)	(0.0013)	(0.0005)	(0.0012)	(0.0017)	(0.0039)	(0.0032)	(95)
Adj R-squared	0.0942	0.2516	0.0223	0.0321	0.2567	0.3080	0.1820	0.0601
ELA TVA	0.0001	0.0025	-0.0000	0.0001	0.0055**	0.0054	0.0002	-110
	(0.0010)	(0.0017)	(0.0007)	(0.0017)	(0.0025)	(0.0066)	(0.0043)	(172)
Adj R-squared	0.0937	0.2500	0.0219	0.0316	0.2522	0.3022	0.1733	0.0599
Outcome Mean	[.8988]	[.6853]	[.0316]	[.0826]	[.2589]	[.395]	[.1345]	[23,289]
N	1,080,311	919,422	723,722	341,840	341,840	103,850	103,817	236,408
Panel B: Multivariate Models								
Math TVA	-0.0011*	0.0007	-0.0007	-0.0005	0.0056***	0.0011	0.0133***	184*
	(0.0006)	(0.0013)	(0.0005)	(0.0012)	(0.0017)	(0.0039)	(0.0032)	(95)
ELA TVA	0.0004	0.0022	0.0001	0.0002	0.0043*	0.0052	-0.0028	-155
	(0.0010)	(0.0017)	(0.0007)	(0.0017)	(0.0026)	(0.0064)	(0.0044)	(175)
Adj R-squared	0.0943	0.2526	0.0225	0.0326	0.2580	0.3100	0.1821	0.0604
Outcome Mean	[.8988]	[.6853]	[.0316]	[.0826]	[.2589]	[.395]	[.1345]	[23,289]
N	1,080,311	919,422	723,722	341,840	341,840	103,850	103,817	236,408

Notes: This table shows estimates for the effect of teacher valued-added (TVA) for students' test scores on educational attainment and earnings. The sample is our student-year level TVA sample described in Section 3, restricted to observations with a non-missing leave-out TVA measure. This table reports the effect of a one standard deviation increase in test score TVA on outcomes. The specification is our main TVA model with TVA controls in Section 4.2, Equation 6. The TVA controls include: i) lags of a cubic polynomial in students' math and ELA test scores; ii) lags of a cubic polynomial in class- and grade-level means of those test scores; iii) the same current demographic controls used in our descriptive test score model (Equation 1): indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and imputed values for missing demographics; and iv) class- and grade-level means of those demographic variables. All these covariates are interacted with grade fixed effects, and we include a control for class size. When lagged scores in the other subject are missing, we set the other subject lagged score to zero and include an indicator for missing data in the other subject interacted with controls for lagged own-subject test scores. Panel A shows results using a univariate model, so each cell in Panel A represents a separate regression. Panel B shows results using a multivariate model, so each column in Panel B represents a separate regression. Average annual earnings from ages 22-24 are conditional on having positive observed earnings and measured in 2024 dollars. Robust standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

A Appendix

Table A1: HSA Subscore Descriptions

HSA Test	HSA Subscore
Algebra I	Analyzing Patterns and Functions
Algebra I	Modeling Real World Situations
Algebra I	Collecting, Organizing, and Analyzing Data
Algebra I	Using Data to Make Predictions
English 10	Reading/Literature: Comprehension and Interpretation
English 10	Reading/Literature: Making Connections and Evaluation
English 10	Writing: Composing
English 10	Writing: Language Usage and Conventions
Biology	Skills and Processes of Biology
Biology	Structure and Function of Biological Molecules
Biology	Structure and Function of Cells and Organisms
Biology	Inheritance of Traits
Biology	Mechanism of Evolutionary Change
Biology	Interdependence of Organisms in the Biosphere
U.S. Government	U.S. Government Structure, Functions, and Principles
U.S. Government	Protecting Rights and Maintaining Order
U.S. Government	Systems of Government and U.S. Foreign Policy
U.S. Government	Impact of Geography on Governmental Policy
U.S. Government	Economic Principles, Institutions, and Processes

Notes: This table shows the subscores for the Maryland High School Assessments (HSA) (MSDE, 2012).

Table A2: MSA Subscore Descriptions

MSA Test	MSA Subscore
Math	Algebra, Patterns, or Functions
Math	Geometry and Measurement
Math	Statistics and Probability
Math	Number and Relationships Computation
Math	Processes of Mathematics
ELA	General Reading Processes
ELA	Informational Reading Processes
ELA	Literary Reading Processes
Science	Skills and Processes
Science	Earth/Space Science
Science	Life Science
Science	Chemistry
Science	Physics
Science	Environmental Science

Notes: This table shows the subscores for the Maryland School Assessments (MSA) (MSDE, 2003, 2016).

Table A3: Teacher Value-Added Estimates of Forecast Bias

	(1) Math	(2) ELA
TVA	1.1153***	1.0992***
	(0.0281)	(0.0448)
Adj R-squared	0.7517	0.7204
N	1,712,724	1,671,687

Notes: This table reports the effect of teacher value-added (TVA) on current-year student test scores, serving as a validity check of the TVA identification strategy. The sample includes student-year observations from the TVA sample (Section 3) with non-missing leave-out TVA measures, where TVA is scaled in student-level test score standard deviations and estimated using data from other years taught by the same teacher (Section 4.2). Each column shows the effect of subject-specific TVA on the corresponding standardized test score. The specification follows our main TVA model (Equation 6), including: i) lags of a cubic polynomial in students' math and ELA scores; ii) lags of a cubic polynomial in class- and grade-level means of those scores; iii) demographic controls (indicators for gender, race, gender-race interactions, FARMS, ELL, SPED, and missing demographics); and iv) class- and grade-level means of these demographic variables. All controls are interacted with grade fixed effects, and class size is included. For missing lagged scores in the non-matching subject, we set the value to zero and include a missing indicator interacted with the own-subject lag. Robust standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A4: Multivariate Relationship Between Test Scores and Labor Market Outcomes

	(1) Emp 22-24	(2) Emp 25-28	(3) Emp 29-32	(4) Earn 22-24 Uncond	(5) Earn 25-28 Uncond	(6) Earn 29-32 Uncond	(7) Earn 22-24 Cond	(8) Earn 25-28 Cond	(9) Earn 29-32 Cond
Panel A: EMS Sample									
Comp Score Math	-0.0198***	-0.0303***	-0.0314***	513***	1,863***	2,772***	1,413***	4,874***	8,096***
	(0.0007)	(0.0011)	(0.0023)	(34)	(68)	(172)	(38)	(74)	(207)
Comp Score ELA	-0.0201***	-0.0350***	-0.0358***	-1,224***	-1,033***	-675***	-768***	775***	2,167***
	(0.0007)	(0.0010)	(0.0022)	(31)	(60)	(155)	(34)	(66)	(183)
Adj R-squared	0.0139	0.0232	0.0190	0.0201	0.0250	0.0252	0.0397	0.1142	0.1530
Outcome Mean	[.8287]	[.7232]	[.6023]	[17,806]	[24,930]	[28,697]	[23,138]	[36,697]	[48,249]
N	1,800,494	995,098	152,752	1,800,494	995,098	152,752	1,492,104	719,675	91,999
Panel B: HS Sample									
Comp Score Math	-0.0098***	-0.0102***	-0.0046**	395***	1,519***	2,607***	910***	2,873***	4,676***
	(0.0010)	(0.0014)	(0.0023)	(48)	(82)	(150)	(51)	(88)	(174)
Comp Score ELA	-0.0124***	-0.0165***	-0.0316***	-891***	-460***	-921***	-710***	369***	841***
	(0.0011)	(0.0015)	(0.0025)	(51)	(85)	(158)	(54)	(90)	(181)
Comp Score Sci	-0.0078***	-0.0143***	-0.0114***	-65	306***	498***	241***	1,159***	1,517***
	(0.0011)	(0.0015)	(0.0023)	(51)	(84)	(145)	(53)	(89)	(164)
Comp Score SS	-0.0146***	-0.0246***	-0.0212***	-134**	298***	597***	397***	1,852***	2,787***
	(0.0011)	(0.0015)	(0.0026)	(53)	(87)	(163)	(55)	(93)	(188)
Adj R-squared	0.0129	0.0206	0.0184	0.0215	0.0289	0.0269	0.0419	0.1120	0.1319
Outcome Mean	[.8397]	[.7571]	[.6461]	[18,109]	[24,917]	[28,455]	[23,263]	[35,528]	[45,442]
N	382,751	260,277	124,378	382,751	260,277	124,378	321,399	197,068	80,362

Notes: This table shows estimates for the multivariate relationship between test scores and employment and earnings. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 3. Columns (1)-(3) show estimates for employment at ages 22-24, 25-28, and 29-32, respectively. Employment is defined using a binary indicator for having positive earnings observed in Maryland UI data in each respective age range. Columns (4)-(6) show estimates for unconditional average annual earnings for the same respective age ranges with zeros imputed for those who are missing earnings. Columns (7)-(9) show estimates for conditional average annual earnings for the same respective age ranges, which are identical to our estimates in Table 3, columns (6)-(8). The specification is our main multivariate test score model with demographic controls in Section 4.1, Equation 1, so each column in a panel represents a separate regression. Average annual earnings are measured in 2024 dollars. Robust standard errors are clustered at the student level.

*** $p<0.01$, ** $p<0.05$, * $p<0.10$