Skills That Pay: Subject-Specific Test Scores and Long-Run Outcomes*

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

Abstract

We estimate earnings gradients and postsecondary outcomes associated with K-12 test scores and subscores across subjects and grades using linked administrative data from Maryland. While math and English both strongly predict college enrollment and completion, math scores—especially in middle school—are linked to significantly higher earnings in adulthood. Educational attainment explains nearly all of the English earnings gradient and about half of the math gradient. Subscores reveal substantial within-subject heterogeneity in skills and earnings gradients that composite scores mask. Finally, math and ELA act as partially substitutable inputs for high school graduation and early postsecondary attainment, but become increasingly complementary for BA attainment and earnings.

*We thank Melissa Kearney, Sergio Urzúa, Jing Liu, Matthew Chingos, Maggie Reeves, and Victoria Nelson, as well as conference participants at Student Upward Mobility Initiative (SUMI), MidSouth Education Policy Workshop, and APPAM for their 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 supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2236417. This research was conducted using data from the Maryland Longitudinal Data System (MLDS) and was supported by the MLDS Center, whose staff provided valuable technical assistance. 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 State of Maryland, the MLDS Center, the MLDS Governing Board, or its partner agencies. All remaining errors are our own. This study was determined to be exempt by the University of Maryland IRB (Project #2205386).

†Postdoctoral Researcher, RAND Corporation (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).

Schools cultivate a wide range of academic skills that shape students' postsecondary and labor market trajectories. Standardized test scores are the primary metrics used by educators, researchers, and policymakers to evaluate these skills and to gauge the effectiveness of educational investments. Yet despite their central role, data remain surprisingly sparse on the core empirical statistic that underlies nearly all policy discussions about raising achievement: how a one-standard deviation increase in performance—whether in math, reading, or other subjects—relates to long-run educational attainment and earnings. Important questions also remain about the deeper structure of these relationships: how they differ across grades and assessments, how much composite test scores mask variation in underlying subskills, and how achievement across multiple subjects jointly shapes later outcomes.

This paper links longitudinal administrative records with detailed test score data—including subject-specific subscores—to estimate how achievement across grades, subjects, and underlying skill components predicts postsecondary outcomes and earnings. Using these linked data, we generate a comprehensive set of estimates of these relationships across a broad span of grade levels (from third grade onward) and core subjects (math, English, science, and social studies), allowing researchers to translate test-score-based treatment effects and cost-benefit analyses into more concrete and comparable terms. We then build upon these estimates to examine the mechanisms and structure underlying the relationship between achievement and later outcomes: how much postsecondary attainment and degree field account for each subject's earnings gradient, which constituent skills drive composite score relationships, and how nonlinearities across the achievement distribution and complementarities across subjects shape these patterns.

We draw on the Maryland Longitudinal Data System (MLDS), which links statewide K-12 data for Maryland public school students to postsecondary enrollment and completion records and to quarterly earnings records, allowing us to follow individuals through age 32. Focusing first on the state's annual math and ELA assessments for grades 3-8, we find clear differences in how achievement relates to long-run earnings. Conditioning on other-subject performance and demographic characteristics, a one standard deviation (SD) increase in math scores predicts roughly \$8,100 (17 percent) higher average annual earnings at ages 29-32, compared with about \$2,200 (4.5 percent) for the same increase in English Language Arts (ELA) scores, even though math and ELA exhibit much smaller differences in their educational attainment gradients.

We find that the relative strength of the math-earnings gradient persists across assessments, grades, and student groups. Analyses of Maryland's high school end-of-course assessments show that a one SD increase in Algebra I performance predicts approximately $4,700 higher average annual earnings at ages 29-32, compared with $2,800 for U.S. Government, $1,500 for Biology, and $800 for English 10. Across grades 3-8, the math earnings gradient also exceeds the ELA gradient at every grade and generally grows as students progress through school. The math-ELA gap also appears across all demographic groups, although it narrows among lower-income and lower-achieving students. Finally, earnings gradients for both subjects become larger as earnings are measured at older ages, though the math gradient grows more quickly.

We further examine how subject-specific achievement predicts bachelor's degree field. Across nine degree fields, achievement patterns closely align with the skills emphasized in different areas of study, consistent with comparative advantage shaping educational specialization. Math achievement is most strongly associated with completion of bachelor's degrees in math-intensive fields such as mathematics, statistics, and computer science, while science achievement is most strongly associated with life science degrees. In contrast, ELA achievement is the strongest predictor of humanities degree completion, while social studies achievement most strongly predicts social science degree completion. Together, these findings suggest that academic achievement reflects not only overall skill levels but also distinct bundles of skills that shape students' pathways through higher education and into the labor market. Consistent with this interpretation, differences in degree field explain part of the earnings gap across subjects: attainment and degree field together account for roughly half of the math earnings gradient, while a substantial portion of the relationship between math achievement and earnings remains unexplained by observed higher education pathways.

We next turn from comparisons across subjects to variation among the component skills measured *within* each subject. Maryland's assessments include composite scores as well as subscores that separate achievement into narrower skill domains. Using subscores in place of the composite measures reveals substantial within-subject heterogeneity. Bachelor's degree completion generally relates to a broad set of component skills, whereas earnings tend to be associated with a smaller subset. Within math and English, subscores capturing analytical and applied reasoning—such as conceptual reasoning, mathematical modeling, text analysis, and written composition—are

more strongly associated with later earnings than subscores focused primarily on procedural skills or content knowledge. This distinction is less pronounced in science and social studies, where subscores more often combine elements of analytical reasoning and content knowledge. These findings suggest that while composite test scores remain valuable summaries of overall achievement, important variation lies beneath them. In particular, assessments that distinguish transferable cognitive skills from procedural and content mastery may provide additional insight into the dimensions of achievement most predictive of long-run labor market success. Subscores also predict degree field specialization, with specific subscore profiles aligning with fields that emphasize different combinations of academic skills.

We next relax the assumptions that outcome gradients are linear in achievement and additively separable across math and ELA. Allowing for flexible functional forms reveals meaningful nonlinearities for educational attainment, though the earnings gradients remain close to linear across most of the achievement distribution. In terms of cross-subject interactions, we find that math and ELA behave as substitutes for high school graduation, college enrollment, and associate degree completion, but as complements for bachelor's degree completion, STEM degree completion, and earnings. Strength in one subject therefore appears to offset weakness in the other at lower educational margins, whereas math and ELA achievement reinforce one another in predicting bachelor's degree completion, STEM degree completion, and earnings.

We also provide two checks on the interpretation of the descriptive gradients. First, conditioning on subject GPAs—which summarize sustained classroom performance—we assess whether test scores capture skills that grades do not. Adding GPAs attenuates many coefficients, but math scores continue to show strong relationships with STEM degree completion and earnings, whereas the ELA earnings gradient largely disappears. Second, using teacher value-added (TVA), we examine whether plausibly exogenous teacher-induced variation in subject-specific achievement yields similar patterns. Math and ELA teacher value-added generate comparable gains in on-time bachelor's degree attainment, but math value added has a larger effect on STEM degree completion and produces a positive, albeit imprecise, earnings estimate. Although the TVA coefficients are smaller than the descriptive gradients, their ordering across subjects and outcomes aligns with our central findings.

Our findings contribute to a longstanding literature relating cognitive achievement to educa-

tional attainment and earnings. Earlier studies primarily use longitudinal surveys or adult skill assessments (Murnane et al., 1995, 2000; Dougherty, 2003; Hanushek et al., 2015). Chetty et al. (2014b) use administrative data to show that students assigned to teachers with higher test score value-added subsequently attend college at higher rates and earn more. Their analysis focuses on teacher value added in grades 4-8 and pools math and English in the main estimates, leaving open how achievement across a broader range of grades and subjects relates to long-run outcomes. Aucejo and James (2021) use administrative data from England to show that verbal skills play a larger role than math skills in college attendance and graduation. Delaney and Devereux (2026) similarly follow students in England into the labor market and show that mathematical and verbal achievement follow different pathways through higher education and earnings. We complement this work by providing a systematic comparison in a U.S. setting across grades, four academic subjects, and the component skills within subjects. These findings echo broader evidence from labor economics that the labor market increasingly rewards nonroutine cognitive skills (Autor et al., 2003). By linking the task-based distinction between nonroutine and routine cognitive skills to specific dimensions of K-12 achievement, our analysis provides new evidence on how these skills emerge in the education system. We also characterize how these relationships vary across student groups, the achievement distribution, and combinations of math and ELA performance.

We also contribute to an emerging literature linking granular assessment data to long-run economic outcomes. Bruhn et al. (2025) show that item-level response patterns contain substantial predictive information beyond conventional summary scores, while Moreno-Medina et al. (2026) use test-item text to identify the curricular skills most strongly associated with adult earnings. Whereas these studies extract new information from individual test items, we study the subscore domains already embedded in many state assessments and trace their relationships with postsecondary outcomes and earnings. Our analysis focuses on the middle ground between composite scores and outcome-specific item measures, showing how existing scoring scales can reveal economically consequential skill variation that composite scores mask.

1 Data

We use data from the Maryland Longitudinal Data System (MLDS) Center covering the years 2008 to 2024. The MLDS Center maintains a centralized repository of linked student and workforce

records created through partnerships among multiple state agencies and non-profit entities. Our analysis draws on three core datasets: (i) K–12 records from the Maryland State Department of Education (MSDE), (ii) postsecondary enrollment and completion records from the Maryland Higher Education Commission (MHEC) and the National Student Clearinghouse (NSC), and (iii) quarterly earnings records from the Maryland Department of Labor's Division of Unemployment Insurance (UI).

The K-12 data from MSDE include information on all students enrolled in Maryland public schools, covering attendance and demographics, assessments, course enrollments, school information, and high school graduation.[1] For postsecondary outcomes, MHEC provides records on all students enrolled in Maryland colleges, and NSC captures out-of-state enrollment for Maryland public high school graduates. The two sources overlap extensively, yielding near-complete coverage of postsecondary outcomes for the cohorts for which these data are available.[2] The earnings data from the Maryland Division of UI capture wages reported by Maryland employers covered by the state unemployment insurance system. These data do not include earnings for federal employees, military personnel, self-employed workers, independent contractors, or individuals working out of state.[3] As a result, we cannot distinguish between nonemployment versus non-coverage for individuals with missing earnings records.

Analytic Samples. We construct three samples, each tailored to different assessments and empirical tasks:

1. **Elementary and Middle School ("EMS") sample:** This sample includes all students in grades 3 through 8 from 2008-2019 with non-missing math and English Language Arts (ELA) test scores. Math and ELA were assessed annually for these grades and years.[4]

2. **High School ("HS") sample:** This sample includes all students from 2008-2019 with non-missing end-of-course test scores in Algebra I, English 10, Biology, and U.S. Government. Students generally take each exam once, although some retake an exam after an unsuccessful

[1] Course enrollment data begin in 2013.

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

[3] 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 more broadly (BLS, 2025a,b).

[4] The math and ELA assessments from 2008-2014 were the Maryland School Assessments (MSA), and the assessments from 2015-2019 were assessments designed by the Partnership for Assessment of Readiness for College and Careers (PARCC). 5th and 8th graders were also assessed in science in 2008-2015 (MSA) and 2018-2019 (Maryland Integrated Science Assessments, MISA).

first attempt; we use the first-attempt score. Although these are high school-level assessments, some students first take them in middle school, particularly for Algebra I.[5]

3. **Teacher Value-Added ("TVA") sample:** This sample consists of student-year-subject observations for all Maryland 4th-8th grade students from 2013-2019 with non-missing math or ELA test scores. Using MLDS course data, we link students in the EMS sample to their math and ELA classroom teachers each year to construct student-year-subject data for the relevant grades and years.[6] We then apply standard sample restrictions taken from the TVA literature (Chetty et al., 2014a; Jackson, 2018; Petek and Pope, 2023).[7]

All samples include students enrolled in Maryland regular public schools.[8] The EMS and HS samples also exclude outgoing transfer students who move from a Maryland public high school to a private or out-of-state high school.[9]

Key Explanatory Variables. The main explanatory variables are standardized test scores and GPAs. In the EMS sample, the two primary measures we use are math and ELA test scores, each standardized to have a mean of zero and a standard deviation of one. We also use annual math, ELA, science, and social studies (SS) GPAs, also standardized to have a mean of zero and a standard deviation of one. In the HS sample, our four main explanatory variables are standardized test scores for Algebra I, English 10, Biology, and U.S. Government. This sample also uses standardized GPAs in the four core subjects, though unlike the EMS sample, these GPAs are cumulative and reflect grades earned throughout high school.

Student Outcomes. Our main outcome variables capture educational attainment, degree field, and earnings. For educational attainment, we construct binary indicators for high school

[5]The Algebra I and English 10 assessments from 2008-2014 were from Maryland's High School Assessments (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.

[6]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.

[7]These restrictions include: 1) each observation must have non-missing observable covariates to compute TVA measures, especially lagged own-subject test score; 2) we omit observations when there is more than one subject area course observed for a student in an academic year; 3) we drop classrooms with over 40 percent of students receiving special education services (SPED); 4) we drop classrooms in which the assigned teacher is linked to 200 or more students in an academic year; 5) we drop classrooms with fewer than 7 students or 40 or more students.

[8]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.

[9]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.

graduation within four years, on-time college enrollment, associate degree (AA) completion within four years, and bachelor's degree (BA) completion within 6 years.[10] We classify degree fields using the STEM-designated degree list maintained by the Department of Homeland Security (DHS, 2023) and 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.[11]

Workforce Outcomes: Our main workforce outcomes are conditional average annual earnings at ages 22-24, 25-28, and 29-32. We construct these measures in four steps: (i) sum quarterly earnings from all Maryland UI-covered employers within each 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. We also analyze employment and unconditional earnings. Employment is defined as having positive earnings in the Maryland UI data within each age range. Unconditional earnings impute zeros for individuals with missing earnings.[12]

Descriptive Statistics. Table 1 provides summary statistics for the EMS and HS samples. The two samples are broadly similar, with modest positive selection arising from the exclusion of specialized schools and the requirement of non-missing test scores. One notable difference is that almost 42 percent of EMS observations are eligible for Free and Reduced-price Meals (FARMS), compared with 36 percent in the HS sample. Standardized test score means are above zero in both samples, suggesting that conditioning on all test scores being non-missing induces mild positive selection and reduces the variance of test score and GPA measures. In contrast, GPA measures

[10]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.

[11]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 most common majors.

[12]The largest sources of missingness in earnings data are non-employment and out-of-state migration. During the period when earnings are measured in our study, the national prime-age employment-to-population ratio ranged from 75 to 81 percent (BLS, 2025c). While this statistic does not directly identify the sources of missing earnings in our data, it suggests that non-employment likely accounts for a substantial share of observations without recorded earnings.

remain closer to the population mean, potentially reflecting differences in grading environments where grades capture performance relative to school or classroom peers. High school graduation rates are higher in the HS sample, likely because students with non-missing high school test scores drop out at lower rates than those selected on earlier-grade test scores in the EMS sample. However, BA receipt, STEM BA receipt, and earnings from ages 25-28 and 29-32 are higher in the EMS sample.

2 Subject-Specific Achievement and Long-Run Outcomes

2.1 Estimating Subject-Specific Gradients

We estimate how test scores in each subject relate to long-run educational attainment and earnings, following the general approach used in Bettinger et al. (2013). We estimate the following descriptive model:

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

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, C_{it} is a vector of demographic covariates including indicators for gender, race, gender-race interactions, FARMS, English-language learner (ELL), and special education (SPED) status, and ϵ_i is an error term clustered at the student level. In this specification, β_k gives the difference in the outcome associated with a one SD higher score in subject k, holding scores in the other included subjects fixed. For binary outcomes, β_k represents a change in the probability of the outcome, which we express in the text as percentage-point differences.; for earnings, it represents a difference in average annual earnings.

The EMS specification includes math and ELA scores, while the HS specification includes Algebra I, English 10, Biology, and U.S. Government scores. In the EMS sample, each observation corresponds to a student-year, and students contribute multiple observations when we observe them in more than one tested grade. In the HS sample, each student contributes one observation containing the student's first-attempt scores on the four end-of-course assessments. We cluster standard errors at the student level to account for repeated observations in the EMS sample.

For comparison with prior work that examines one achievement measure at a time (Chetty et al., 2014c), we also report univariate models that include each subject score separately. These

models show how predictive each score is when it serves as the only achievement measure in the regression. Because students who score higher in one subject also tend to score higher in other subjects, the univariate coefficient incorporates the outcome relationships associated with those omitted scores. Comparing the univariate and multivariate results shows how the estimated relationship for each subject changes once the model accounts for performance in the other subjects.

We also estimate specifications that add standardized GPAs in the four core subjects. Grades summarize sustained classroom performance across assignments, examinations, and teacher evaluations, while standardized tests measure performance on a common assessment. Including both measures examines whether test scores contain predictive information beyond the information reflected in students' course grades.

We interpret the resulting β_k estimates as *descriptive* gradients rather than the causal effect of raising achievement in subject k. Family resources, school environments, motivation, and other unobserved inputs may influence both test performance and later outcomes. Section 2.5 supplements our descriptive analysis with a teacher value-added design that estimates the long-run effects of assignment to teachers who generate larger achievement gains in math or ELA.

2.2 How Test Scores Relate to Long-Term Outcomes

Table 2 reports estimates from our preferred multivariate specification, which includes all available subject scores simultaneously. We highlight two broad patterns. First, achievement across subjects predicts general educational attainment at broadly similar rates. Second, math achievement predicts STEM BA completion and earnings substantially more strongly than achievement in other subjects.

In the EMS sample, math and ELA scores both predict high school completion within four years, with gradients of 4.5 and 2.4 percentage points, respectively. For postsecondary outcomes, math and ELA scores are comparably predictive of college enrollment and BA completion. A one-SD higher math score corresponds to a 9.2-percentage-point increase in on-time college enrollment, compared with 7.9 percentage points for ELA. For BA completion within six years, the corresponding gradients are 11.8 percentage points for math and 7.7 percentage points for ELA. The HS sample shows even less evidence that any single subject dominates general attainment: a one-SD higher score predicts BA completion increases ranging from 3.6 percentage points for Biology to 6.9 percentage points

for U.S. Government, with Algebra I and English 10 falling between these values. Performance in all tested subjects therefore similarly predicts college enrollment and BA completion, with no single subject consistently producing the largest gradient across these outcomes.

Subject differences begin to diverge when we examine STEM degree completion. In the EMS sample, a one-SD increase in math scores corresponds to a 7.7-percentage-point increase in STEM BA completion—more than three times the 2.5-percentage-point gradient for ELA. Our HS results reflect this same pattern, although they are less pronounced: the STEM BA gradients are 3.7 percentage points for Biology and 3.0 for Algebra I, compared with 2.0 for English 10 and 2.1 for U.S. Government.

Subject differences grow even larger for earnings. In the EMS sample, a one-SD increase in math achievement corresponds to roughly $8,100 in higher average annual earnings at ages 29-32 (17 percent of mean earnings), compared with about $2,200 (4.5 percent) for ELA. We also note that this gap appears to further widen with age. At ages 22-24, the math and ELA gradients equal about $1,400 and -$800, respectively; by ages 29-32, both have increased, but the math gradient rises much more sharply ($8,100 vs. $2,200). The HS sample reinforces this ordering across a broader set of subjects. At ages 29-32, a one-SD increase in test scores corresponds to approximately $4,700 in higher annual earnings for Algebra I, $2,800 for U.S. Government, $1,500 for Biology, and $800 for English 10.

Univariate results: Table A3 provides a univariate comparison that includes one subject score at a time. In these models, all subject scores strongly predict attainment and earnings, and cross-subject differences appear considerably smaller since each score also reflects the tendency of high-scoring students to perform well on the omitted assessments. For example, in the EMS sample, the univariate earnings gradients at ages 29-32 equal roughly $9,500 for math and $7,000 for ELA, compared to the multivariate model where the gradients are $8,100 and $2,200, respectively. The HS sample shows the same pattern: conditioning on other subjects reduces all four earnings coefficients, though math retains the largest gradient.

Controlling for course grades: Table 3 extends the baseline multivariate specification by adding standardized GPAs in the four core subjects alongside the composite test scores. Grades reflect sustained classroom performance and directly influence several educational transitions, while standardized tests measure performance on a common exam. The joint specification therefore

asks whether test scores retain predictive information after accounting for students' classroom performance.

For BA completion, adding GPAs reduces the math and ELA test score gradients by roughly 35 and 50 percent, respectively. The test score gradient exceeds the GPA gradient for math (7.7 versus 2.3 percentage points) but not for ELA (3.9 vs. 6.0). For the HS sample, adding GPAs reduces the Algebra I, English 10, and U.S. Government test-score gradients by roughly 45-60 percent and turns the Biology gradient slightly negative. GPA gradients exceed the corresponding test score gradients in all four subjects: the difference is modest for math (3.6 versus 3.3 percentage points) but substantially larger for ELA (10.9 versus 3.0), science (4.1 versus -0.6), and social studies (7.1 versus 2.7).

We observe clearer differences for STEM BA completion. In the EMS sample, adding GPAs leaves the math test score gradient for STEM completion essentially unchanged at 7.7 percentage points but reduces the ELA gradient from 2.5 to 1.4 percentage points. Comparing test scores versus grades, the math test score gradient substantially exceeds the math GPA gradient (7.7 versus 1.9 percentage points), while the ELA test score and GPA gradients are similar (1.4 versus 1.6). In the HS sample, adding GPAs reduces the Algebra I, English 10, Biology, and U.S. Government test score gradients by roughly 31, 49, 29, and 69 percent, respectively. GPA gradients exceed test score gradients for math (4.5 versus 2.1 percentage points), ELA (2.2 versus 1.0), and social studies (1.3 versus 0.7), while the Biology test score and science GPA gradients are similar (2.6 versus 2.4).

These differences persist for earnings. For EMS earnings at ages 25-28, adding GPAs reduces the math test score gradient by roughly 30 percent, from \$4,900 to \$3,500, and turns the ELA gradient negative.[13] The test score gradient again exceeds the GPA gradient for math (\$3,500 versus \$800) but not for ELA (-\$900 vs. \$1,800). For HS earnings at ages 29-32, adding GPAs reduces the Algebra I, English 10, Biology, and U.S. Government test score gradients by roughly 38, 90, 26, and 14 percent, respectively. Comparing test scores and grades again, the Algebra I test score gradient exceeds its GPA gradient (\$2,900 versus \$1,500); the GPA gradients exceed the test score gradients for English 10 (\$2,200 versus \$100) and Biology (\$2,200 versus \$1,100); and the two gradients are similar for U.S. Government (\$2,400 versus \$2,300).

[13] We are unable to observe earnings from ages 29-32 in the EMS sample when GPAs enter the model because course data for computing GPAs begin in 2013.

In summary, test scores and grades contain distinct predictive information. GPAs absorb much of the test-score gradient for general BA completion and often produce larger gradients, particularly in ELA, science, and social studies. Test scores retain more predictive content for subject-aligned STEM completion, especially math in the EMS sample and Biology in the HS sample. For earnings, math test scores retain substantial predictive information about earnings beyond course grades, whereas grades carry most of ELA's smaller baseline earnings signal.

2.3 Postsecondary Pathways and Earnings Gradients

Subject-specific achievement may predict earnings partly because it predicts whether students attend college, complete a degree, and earn degrees in fields with different labor market returns. We first examine how test scores relate to completed degree field and then assess how conditioning on these observed postsecondary pathways changes the subject-specific earnings gradients. We interpret this exercise as a descriptive accounting rather than a causal decomposition.

Degree field: Table 4 relates subject-specific test scores to BA completion within six years across nine degree fields. In the EMS sample, math produces substantially larger gradients than ELA for degrees in mathematics, statistics, and computer science; engineering and architecture; science; and business. ELA produces larger gradients for social science and humanities degrees. The HS sample shows even closer alignment between assessment subject and degree field. Among the four test scores, Algebra I produces the largest gradients for degrees in mathematics, engineering, and business; English 10 produces the largest gradient for humanities; Biology produces the largest gradient for science; and U.S. Government produces the largest gradient for social science. Subject-specific performance therefore aligns closely with the fields in which students complete their degrees.

Accounting for postsecondary pathways: Table 5 examines how observable postsecondary outcomes account for the earnings gradients at ages 29-32. Column 1 reproduces the baseline multivariate specification on a common sample.[14] Column 2 adds on-time college enrollment, column 3 adds BA completion within six years, and column 4 adds completed degree field. Changes across columns show how the test score gradients shift as the model holds these pathways fixed.

In the EMS sample, adding college enrollment and BA completion reduces the math earnings

[14]The estimates in Table 5, column (1) are nearly identical to those in Table 2, column (8). Minor differences reflect sample restrictions due to missing degree field data in Table 5.

gradient from approximately $8,100 to $4,800 (an attenuation of 41 percent), and the ELA gradient from approximately $2,200 to $300 (an attenuation of 84 percent). Adding degree field further reduces the math coefficient to $3,700 and raises the ELA coefficient to $700. Relative to the baseline specification, the full set of postsecondary controls reduces the math and ELA earnings gradients by 54 and 67 percent.

The HS sample shows a similar ordering across subjects. The full set of controls reduces the Algebra I gradient from approximately $4,700 to $3,000, a decline of 36 percent. The English 10 coefficient falls from approximately $800 to zero. The Biology and U.S. Government gradients decline from approximately $1,500 to $600 and from $2,800 to $1,200, reductions of 64 and 56 percent, respectively. BA completion produces most of the attenuation in the English 10, Biology, and U.S. Government gradients, while degree field contributes relatively little additional attenuation for English 10 and U.S. Government.

Across both samples, observed postsecondary pathways account for nearly all of the ELA earnings gradient and only part of the math gradient. The ELA attenuation occurs primarily after the model adds college enrollment and BA completion rather than degree field, suggesting that general attainment (and not sorting across majors) accounts for most of ELA's earnings relationship. Math continues to predict earnings among students with similar observed postsecondary pathways, with fully controlled gradients of approximately $3,700 in the EMS sample and $3,000 in the HS sample. These remaining gradients indicate that math scores contain predictive information about earnings that college attendance, degree completion, and field of study do not capture.

2.4 Heterogeneity Across Students, Grades, and Ages

Heterogeneity by Gender, Socioeconomic Status, and Achievement. Table 6 reports conditional test score gradients in the EMS sample for on-time high school graduation, BA attainment within six years, and earnings at ages 29-32, separately by gender, FARMS status, and achievement tercile.

Gender differences change direction across outcomes. Math and ELA gradients for high school graduation are larger for male students, whereas the corresponding BA and earnings gradients are larger for female students. For example, the math coefficient for high school graduation is 5.2 p.p. for males and 3.6 p.p. for females, while the math earnings coefficient is $7,500 for males and

$8,800 for females.

Socioeconomic differences are larger and also vary by outcome. Relative to non-FARMS students, FARMS students have more than twice the gradient for high school graduation but substantially smaller gradients for BA attainment. For earnings, the math coefficient is smaller among FARMS students ($6,600 versus $8,700), while the ELA coefficient is slightly larger ($2,600 versus $1,900). The math-ELA earnings gap is therefore considerably smaller among FARMS students.

We also show how outcomes vary across achievement terciles. Mechanically, high school graduation gradients decline sharply across the achievement distribution. The math coefficient falls from 9.9 p.p. in the bottom tercile to 1.1 p.p. in the top tercile, while the ELA coefficient falls from 6.0 to 0.3 p.p. BA gradients peak in the middle tercile, reaching 14.9 p.p. for math and 10.0 p.p. for ELA. The math earnings gradient follows a similar pattern, rising from $6,000 in the bottom tercile to $9,200 in the middle tercile and remaining high at $8,500 in the top tercile. The ELA gradient remains much smaller and falls to just $1,000 in the top tercile.

These comparisons partly reflect differences in baseline outcome rates. High school graduation already reaches 96 percent in the top achievement tercile, while BA attainment reaches only 8 percent in the bottom tercile. Beyond this compression at the extremes, the weaker BA gradients among FARMS and lower-achieving students may reflect differences in college destinations, including greater attendance at less selective institutions where time to degree has increased (Bound et al., 2012), as well as larger barriers to completion among disadvantaged students (Bailey and Dynarski, 2011). The HS sample in Table A4 yields similar patterns. Earnings gradients for math, science, and social studies are largest in the upper two achievement terciles, while the ELA gradient does not increase with achievement.

Heterogeneity by Grade and Age of Earnings. Figure A15 plots conditional test score gradients by grade. From grades 3 through 8, the gradients generally rise, especially for BA attainment, STEM BA attainment, and earnings. Math and ELA differ relatively little for high school graduation, college enrollment, and AA attainment, but diverge sharply for BA attainment, STEM BA attainment, and earnings. Among eighth graders, a one SD increase in math scores corresponds to a 14 p.p. higher BA attainment rate, a 10 p.p. higher STEM BA attainment rate, and more than $5,000 in higher annual earnings at ages 25-28. The corresponding ELA gradients are 7.5

p.p., 1.5 p.p., and about $1,000. Relative to eighth grade, the high school math gradients decline for most outcomes, while the ELA gradients rise. Comparisons with high school require some caution because the high school estimates use different assessments and a different sample.

Figure A17 shows that earnings gradients steepen with age, especially for math. In the EMS sample, the math coefficient rises from just over $2,000 at age 23, or approximately 10 percent, to about $9,000 at age 30, or approximately 20 percent. The ELA coefficient rises from approximately -$500 to $2,500 over the same ages. The HS sample follows a similar age profile. Math has the largest earnings gradient at nearly every age, while social studies and science generally exceed ELA.

Figures A16 and A18 show a different pattern for subject GPAs. GPA gradients change less across grades than test score gradients. For broad attainment outcomes, math GPA gradients generally fall below those for ELA and social studies; STEM BA attainment provides the main exception. ELA GPA gradients often approach the corresponding ELA test score gradients. GPA earnings gradients rise modestly with age, but they differ relatively little across subjects and remain substantially smaller than the math test score gradients.

2.5 Teacher value-added effects on long-term outcomes

Approach: To supplement and validate our descriptive regressions, we use teacher value-added (TVA) to examine whether assignment to teachers who raise math or ELA achievement relates to later outcomes in similar ways across subjects. Following Chetty et al. (2014a) and subsequent work (Jackson, 2018; Petek and Pope, 2023), we construct subject-specific, leave-year-out TVA measures. We first residualize current-year test scores using prior scores, student demographics, and classroom characteristics. We then predict each teacher's mean residual score in a given year using students whom that teacher taught in other years. Identification requires that, conditional on these controls, teacher assignment is uncorrelated with unobserved determinants of students' later outcomes. Under this assumption, the resulting coefficients identify the effects of assignment to a teacher with higher test score TVA. Appendix C details our estimation procedure.

Appendix Table A6 assesses forecast bias by testing whether leave-year-out TVA predicts current-year residual test scores with a coefficient of one. The estimated coefficients are 1.12 for math, with a 95 percent confidence interval of 1.06 to 1.17, and 1.10 for ELA, with a confidence

interval of 1.01 to 1.19. These estimates imply forecast-bias point estimates of approximately -12 percent for math and -10 percent for ELA. Although modest, both estimates differ statistically from the no-bias benchmark of zero. We therefore interpret the causal magnitudes cautiously.

Results: Table 7 reports univariate models in Panel A and our preferred multivariate models, which include math and ELA TVA simultaneously, in Panel B. The multivariate coefficients for high school graduation, college enrollment, and AA attainment lie close to zero. For BA attainment within four years, a one SD increase in TVA corresponds to increases of 0.56 p.p. for math and 0.43 p.p. for ELA, although the ELA estimate lacks precision. Neither subject has a precise coefficient for BA attainment within six years. Because the six-year estimates draw on a much smaller set of earlier cohorts, differences from the four-year estimates may reflect sample composition and lower precision rather than changes in TVA effects over time.

Our clearest subject-specific result comes from STEM attainment. A one SD increase in math TVA significantly raises STEM BA attainment by 1.33 p.p. (relative to a mean of 13.5 percent), while the ELA coefficient is small, negative, and statistically indistinguishable from zero. For early-career earnings at ages 22-24, the math coefficient is $184 but only significant at the 10% threshold, while the ELA coefficient -$155 and remains imprecisely estimated. The univariate estimates in Panel A closely resemble these multivariate results.

Our TVA results therefore echo our descriptive patterns across several dimensions: math and ELA TVA have similar point estimates for general BA attainment, while math TVA has a distinct positive relationship with STEM BA attainment. The earnings estimates point in the same direction as the descriptive gradients but remain too imprecise to support a strong conclusion. Moreover, the TVA coefficients capture the effects of assignment to higher-TVA teachers, not the causal effect of a test score increase itself. Because a one-SD increase in teacher TVA differs from a one-SD increase in student test scores, the difference in magnitudes prevents direct comparison with our descriptive gradients.

2.6 Alternative Earnings Definitions and UI Coverage

Our baseline labor market outcome measures earnings among individuals with positive Maryland UI records. State UI data do not distinguish nonemployment from employment outside Maryland or in jobs that the UI system does not cover. We therefore examine two alternative

outcomes: an indicator for positive Maryland UI earnings ("observed UI employment"), and earnings that assign zero to every individual without a UI record ("zero-imputed earnings"). Again, we emphasize that both measures imperfectly capture total employment and earnings: observed UI employment misses jobs outside Maryland's UI system, while zero-imputed earnings assigns $0 to everyone without a Maryland UI record—including people working out of state or in uncovered jobs.

Table A7 compares these outcomes with our baseline earnings measure, with the goal of illustrating the limitations of these alternative outcomes. Because the patterns remain similar across age ranges, we focus on ages 29-32. In the EMS sample, a one-SD increase in math and ELA scores corresponds to a 3.1 and 3.6 p.p. lower probability of positive Maryland UI earnings, respectively. These negative gradients do not imply that achievement reduces employment. Rather, higher-scoring students appear less often in Maryland UI records, potentially because they are more likely to work outside the state or in jobs that the UI system does not cover. The data cannot distinguish these workers from individuals without earnings.

Assigning zeros to missing records sharply reduces the estimated earnings gradients. In the EMS sample, the math coefficient falls from $8,100 among individuals with positive UI earnings to $2,800 under zero imputation. The ELA coefficient changes from $2,200 to -$700. The HS sample shows the same general pattern: zero imputation reduces the coefficient for every subject and changes the ELA coefficient from positive to negative.

These comparisons motivate our focus on earnings conditional on a positive Maryland UI record, which avoids coding all missing observations as zero. This approach instead excludes individuals without in-state earnings and therefore changes the estimand to the achievement-earnings gradient among workers with positive Maryland UI earnings. If achievement and potential earnings both influence out-of-state migration, workers who remain in the sample may differ systematically from those who leave. Foote and Stange (2022) show that conditioning on positive in-state earnings reduces the resulting migration bias but does not necessarily eliminate it.

3 The Structure of Skill-Outcome Relationships

3.1 Heterogeneity Across Subskills

Maryland's assessments report composite subject scores alongside subscores that capture narrower skill domains. Appendix Tables A1 and A2 list the subscores available in the EMS and HS samples, respectively. To examine whether composite scores conceal meaningful within-subject heterogeneity, we modify the baseline multivariate specification in Equation 1. For one subject at a time, we replace its composite score with the full set of standardized subscores while retaining composite scores for all other subjects:

$$Y_i = \alpha + \sum_{m=1}^{M_k} \beta_{k,m} s_{it}^{k,m} + \sum_{\ell \neq k} \beta_\ell S_{it}^\ell + \gamma' C_{it} + \epsilon_{it}, \tag{2}$$

where $s_{it}^{k,m}$ denotes subscore m within subject k. We standardize all test measures to have mean zero and standard deviation one. The coefficient $\beta_{k,m}$ measures the conditional gradient associated with a one-SD higher subscore, holding fixed the other subscores within the same subject, composite scores in other subjects, and demographic covariates.

Strong correlations among subscores mean that these coefficients do not form an additive decomposition of the composite-score coefficient. We therefore use them to compare conditional patterns across subskills rather than to quantify each subskill's independent contribution. Many confidence intervals also overlap, so we emphasize broad patterns rather than fine rankings among closely estimated coefficients.

Our results reveal three broad patterns. First, general BA completion relates to a relatively broad set of subskills. Second, earnings gradients concentrate more heavily in skills that require students to apply, interpret, model, or produce information. We view these application- and production-oriented skills as more transferable across academic and workplace settings than proficiency in specific content domains. Third, content-specific subskills show especially strong alignment with related BA fields. This distinction provides an organizing interpretation rather than an exhaustive classification, particularly because science and social studies do not always separate cleanly into process- and content-oriented domains.

Figure 2 presents the EMS results. General BA completion relates to a relatively broad set of

subskills: all five math subscores have positive coefficients, while informational and literary reading have larger ELA coefficients than general reading. Other outcomes show more concentration. "Processes of Mathematics" has the largest math point estimate for high school completion, college enrollment, and earnings. For STEM BA completion, algebra, geometry, and number and computation have larger coefficients than mathematical processes, consistent with a stronger role for content-specific quantitative proficiency. Within ELA, informational reading has the largest earnings coefficient, while general reading has little relationship with earnings. Thus, the EMS results distinguish subskills that predict broad educational progression and earnings from those that align more closely with STEM specialization.

Appendix Figure A7 extends the EMS analysis to BA degree fields and refines the subject-level patterns in Table 4. Algebra, geometry, and number and computation have their largest gradients for quantitative and technical degrees. Informational and literary reading feature more prominently for social science and humanities degrees. These patterns show that content-specific achievement remains closely related to educational specialization even when application-oriented skills display larger gradients for general attainment or earnings.

Figure 3 presents the HS results across four subjects. Algebra I subscores have relatively similar coefficients for general BA completion, while English 10 concentrates more heavily in the two writing domains. Most Biology subscores also have positive BA coefficients, although "Interdependence of Organisms in the Biosphere" has little relationship with completion. Within U.S. Government, "U.S. Government Principles" has the largest BA point estimate. Earnings show a clearer emphasis on application- and production-oriented skills: "Modeling Real World Situations" has the largest Algebra I point estimate, "Writing: Composing" has the largest English 10 estimate, and "Skills and Processes of Biology" has the largest Biology estimate. U.S. Government provides a less distinct process-content contrast, with positive earnings coefficients spread across several content domains.

Appendix Figure A8 shows that HS subscores also align with related BA fields. Algebra I subskills have larger gradients for quantitative fields, Biology subskills for science degrees, English 10 reading and writing subskills for humanities degrees, and U.S. Government subskills for social science degrees. These results again suggest that content-specific proficiency plays a particularly visible role in field choice.

Taken together, the results suggest that composite scores combine two distinct dimensions of achievement. General BA completion relates to proficiency across multiple domains, while earnings gradients concentrate more heavily in subskills that emphasize application, interpretation, modeling, and communication. Content-specific achievement remains consequential for educational specialization, with subscore gradients often aligning closely with related BA fields.

3.2 Nonlinearities and complementarities

We next relax two restrictions imposed by our baseline specification. First, we allow the relationship between each subject score and later outcomes to vary across the achievement distribution. The resulting shapes differ sharply across outcomes: the gradients flatten at higher achievement for high school completion and college enrollment, are hump-shaped for AA receipt, steepen for BA and STEM BA completion, and remain approximately linear for earnings. Second, we allow the gradient associated with one subject to depend on achievement in the other. Math and ELA appear to offset each other as partial substitutes for early educational milestones but reinforce each other as complements for BA completion, STEM BA completion, and earnings.

Figure 1 provides an initial view of the joint relationship between math, ELA, and each outcome in the EMS sample. High school completion and college enrollment rise with both scores but flatten as either score reaches higher levels. AA attainment peaks closer to the middle of the joint distribution. BA completion, STEM BA completion, and earnings show a different pattern, with the highest outcomes concentrated among students who score highly in both subjects. Math creates a particularly strong vertical gradient for STEM completion and earnings, but the darkest regions still occur where high math coincides with high ELA. Appendix Figure A1 shows similar joint patterns in the HS sample. Because students with highly divergent math and ELA scores occur relatively infrequently, as Appendix Figure A2 shows, we interpret the corners of these contour plots cautiously.

Nonlinearities: Figure 4 traces each subject's relationship with outcomes across its own score distribution. For high school completion and college enrollment, the gradients are steepest below the mean and flatten at higher scores. This concavity fits outcomes that depend partly on crossing an academic threshold and leave less room for differentiation among high-achieving students. AA attainment follows a pronounced hump-shaped pattern: attainment rises through the middle

of the score distribution and then declines. This decline likely reflects educational sorting rather than a negative relationship between skills and postsecondary success, as higher-scoring students increasingly pursue BA degrees instead of associate degrees.

BA and STEM BA completion show the opposite pattern. Both rise most sharply above the mean, especially with respect to math, indicating that advanced attainment differentiates students most strongly near the upper end of the achievement distribution. Earnings also increase across most of the distribution, but math and ELA follow markedly different profiles. The math gradient remains steep and continues through the upper end of the distribution, while the ELA gradient is smaller and flattens at higher scores. Appendix Figure A9 broadly reproduces these patterns across the four HS subjects, particularly over the center and upper half of the score distribution. The less regular relationships in the lower tail counsel against strong interpretations of outcomes at the extremes.

Complementarities: We define math and ELA as complementary when the gradient in one subject grows with achievement in the other. We describe them as compensatory when the gradient shrinks as achievement in the other subject rises. Figure 5 plots outcomes across ELA scores at three fixed math levels, while Appendix Figure A11 reverses the comparison by plotting outcomes across math scores at three ELA levels. These two views produce the same central result.

For high school completion and college enrollment, the lines converge as achievement rises. Moving from lower to higher math levels is associated with larger differences in these outcomes among students with low ELA scores, but those differences in outcomes across math levels narrow among students with high ELA scores. Strength in one subject therefore appears to compensate, at least partially, for weakness in the other when students progress through early educational thresholds.

BA and STEM BA completion show the opposite pattern. Differences across math levels widen as ELA rises. Joint strength in both subjects therefore has a stronger relationship with advanced attainment than strength in either subject alone. The positive STEM pattern is especially informative. Although math has the dominant direct relationship with STEM completion, the math gradient grows with ELA achievement. We speculate that STEM readiness may require a combination of quantitative proficiency alongside technical reading, written communication, and the broader academic skills required to complete a four-year degree.

AA attainment again requires a different interpretation. Its negative interaction partly reflects competition between degree pathways: students with stronger math and ELA increasingly complete BA degrees rather than AA degrees. The apparent compensation for AA attainment therefore does not necessarily describe how the two skills contribute to postsecondary success more broadly.

The earnings figures show a strong level difference across math scores and a smaller but positive interaction between the two subjects. Students with higher math scores earn more at nearly every ELA level, consistent with the large math gradient in the baseline model. At the same time, the earnings relationship with math becomes stronger when students also have higher ELA achievement. This pattern is consistent with higher-paying work rewarding combinations of quantitative problem-solving, comprehension, and communication rather than any one skill in isolation.

Appendix Figures A10 and A12 show the same broad progression in the HS sample. Math and ELA appear compensatory for high school completion, enrollment, and AA attainment but complementary for BA completion, STEM BA completion, and earnings.

Figure 6 tests whether own-subject nonlinearities account for these patterns—in other words, non-linearities in general cognitive ability (rather than true cross-skill complementarity) could explain the estimated interactions. The first model includes linear math and ELA scores and their interaction. The second adds quadratic terms, while the third flexibly models each subject with cubic splines and retains the interaction term. We focus on the spline model because it places the fewest restrictions on each subject's relationship with the outcome. The interaction remains negative for high school completion, college enrollment, and AA attainment but positive for BA completion, STEM BA completion, and earnings. Appendix Figure A13 reproduces this ordering in the HS sample. Flexible controls reduce several interaction coefficients, particularly in the HS sample, but do not change their signs. The progression from compensation to complementarity therefore does not simply reflect omitted curvature in the separate math and ELA gradients.

These patterns remain descriptive; for example, ceiling effects, educational sorting, and unobserved dimensions of academic preparation could explain our the estimated interactions. Nevertheless, our results consistently show that the relationship between math and ELA changes across the education-to-work pipeline. Strength in either subject can help students cross early educational thresholds, while advanced degree completion and earnings show stronger gradients

when students possess high achievement in both domains.

4 Conclusion

Researchers and policymakers often use test scores to summarize academic achievement. Our results show that the long-run predictive content of test scores depends on the subject, the competencies measured by the assessment, and the student's broader skill profile. Our primary takeaway is not simply that math achievement predicts earnings more strongly than achievement in other subjects; rather, the relationship between academic skills and later outcomes shifts as students move from broad educational thresholds toward specialization and the labor market.

We encourage researchers to use our baseline estimates as benchmarks for interpreting the long-run significance of short-run achievement gains. Our additional analyses show how researchers can refine those benchmarks to fit the context they study. We highlight that the relevant gradient varies with the subject and grade assessed, the underlying skills captured by the assessment, the outcome of interest, and students' achievement in other domains. We therefore offer not a single test-score conversion, but a set of empirically grounded benchmarks that researchers can match to the intervention and outcome they study.

References

Aucejo, Esteban M. and Jonathan James, "The Path to College Education: The Role of Math and Verbal Skills," *Journal of Political Economy*, 2021, *129* (10), 2905–2946.

Autor, David H, Frank Levy, and Richard J Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333.

Bailey, Martha J. and Susan M. Dynarski, "Inequality in Postsecondary Education," in Greg J. Duncan and Richard J. Murnane, eds., *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, New York, NY: Russell Sage Foundation, 2011, pp. 117–132. Originally published as "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion," NBER Working Paper No. 17633.

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.

Bound, John, Michael F. Lovenheim, and Sarah Turner, "Increasing Time to Baccalaureate Degree in the United States," *Education Finance and Policy*, 10 2012, *7* (4), 375–424.

Bruhn, Jesse, Michael Gilraine, Jens Ludwig, and Sendhil Mullainathan, "Do Test Scores Misrepresent Test Results? An Item-by-Item Analysis," NBER Working Paper 34484, National Bureau of Economic Research November 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*, 2014, *104* (9), 2633–2679.

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

Delaney, Judith M. and Paul J. Devereux, "The Math-Verbal Divide: Unequal Returns to Cognitive Skills in Education and Work," IZA Discussion Paper 18542, IZA@LISER Network April 2026.

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.

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.

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.

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.

Moreno-Medina, Jonathan, Eric R. Nielsen, and Viviana Rodriguez, "The Anatomy of a High-Price Question: Text, Skills, and the Economics of Achievement Measurement," EdWorkingPaper 26-1467, Annenberg Institute at Brown University June 2026.

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," *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.

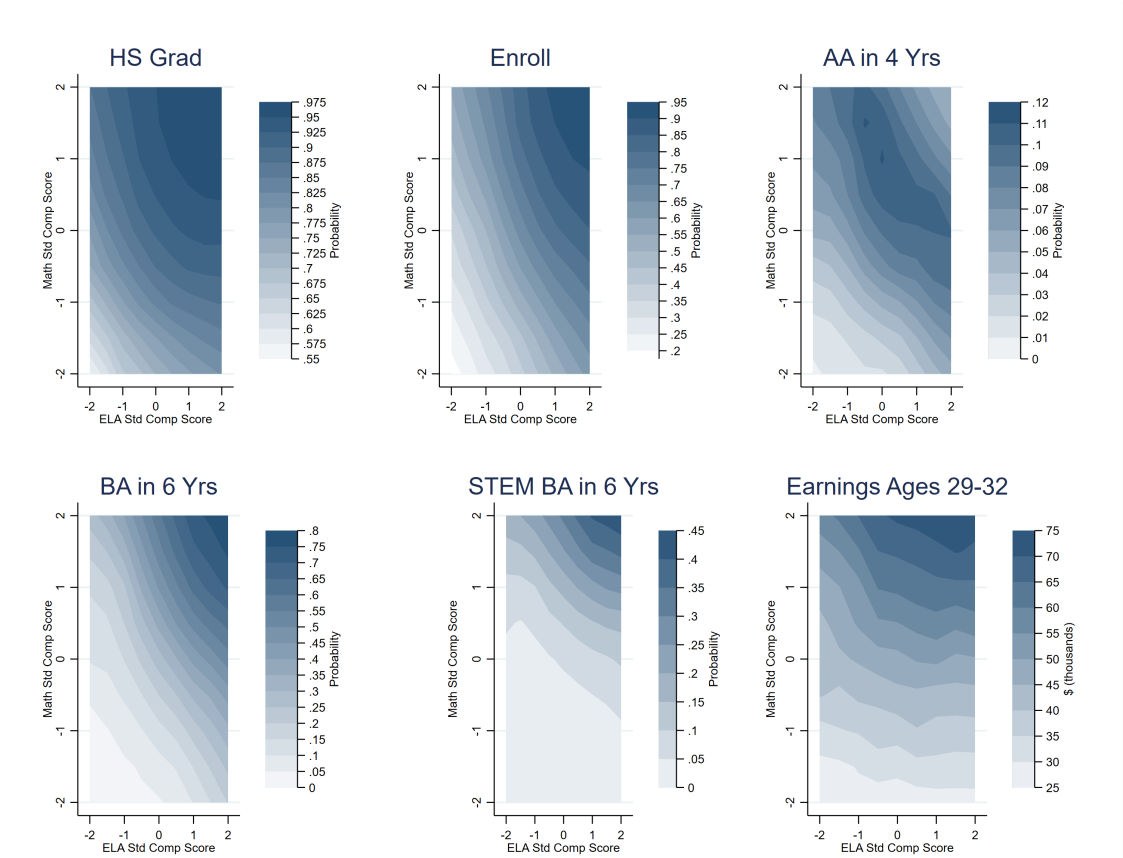
NCES, "The Classification of Instructional Programs (CIP)," https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=56, 2020. Last accessed on July 7, 2025.

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.

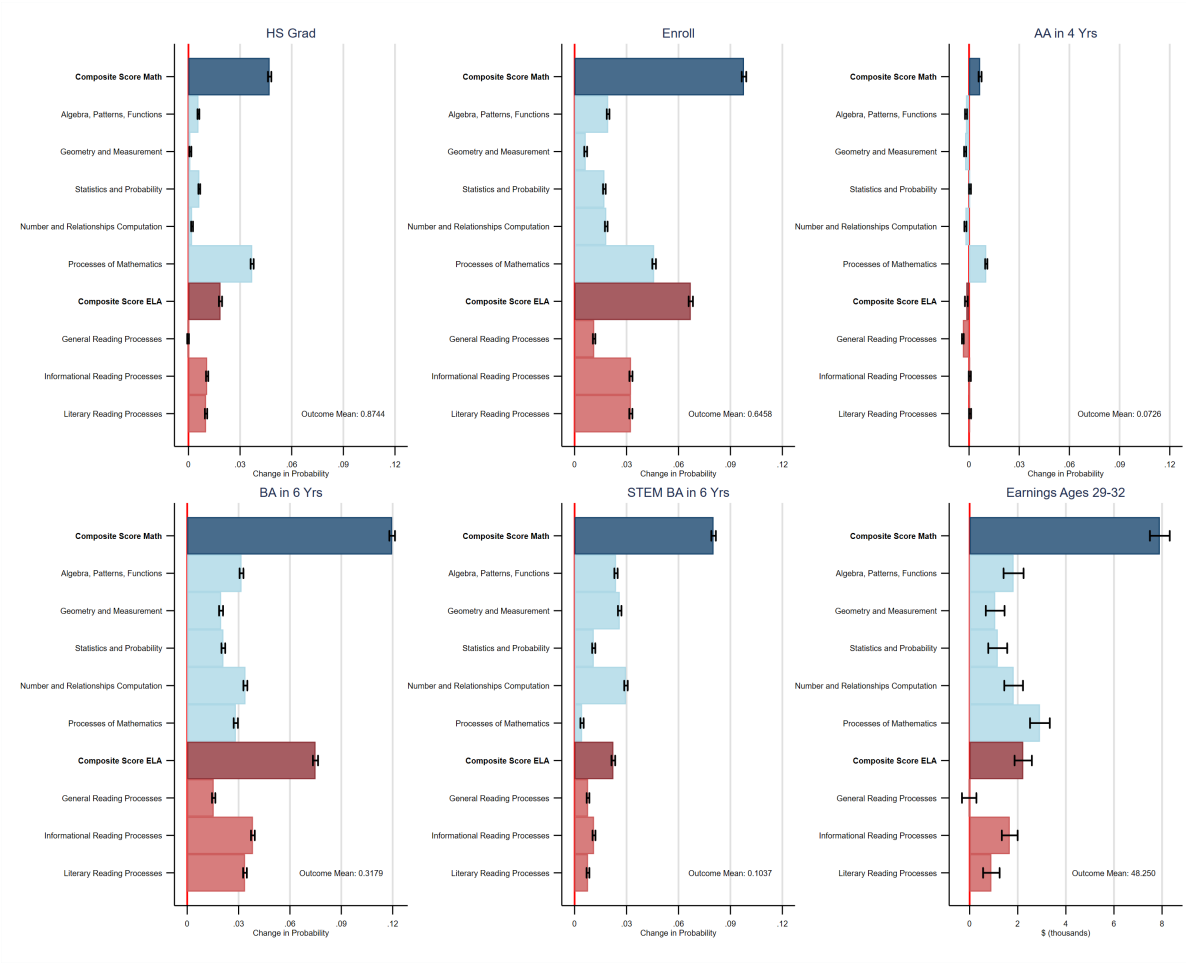
Figures and Tables

Figure 1: Joint Relationship Between Math and ELA Test Scores and Student Outcomes (EMS Sample)



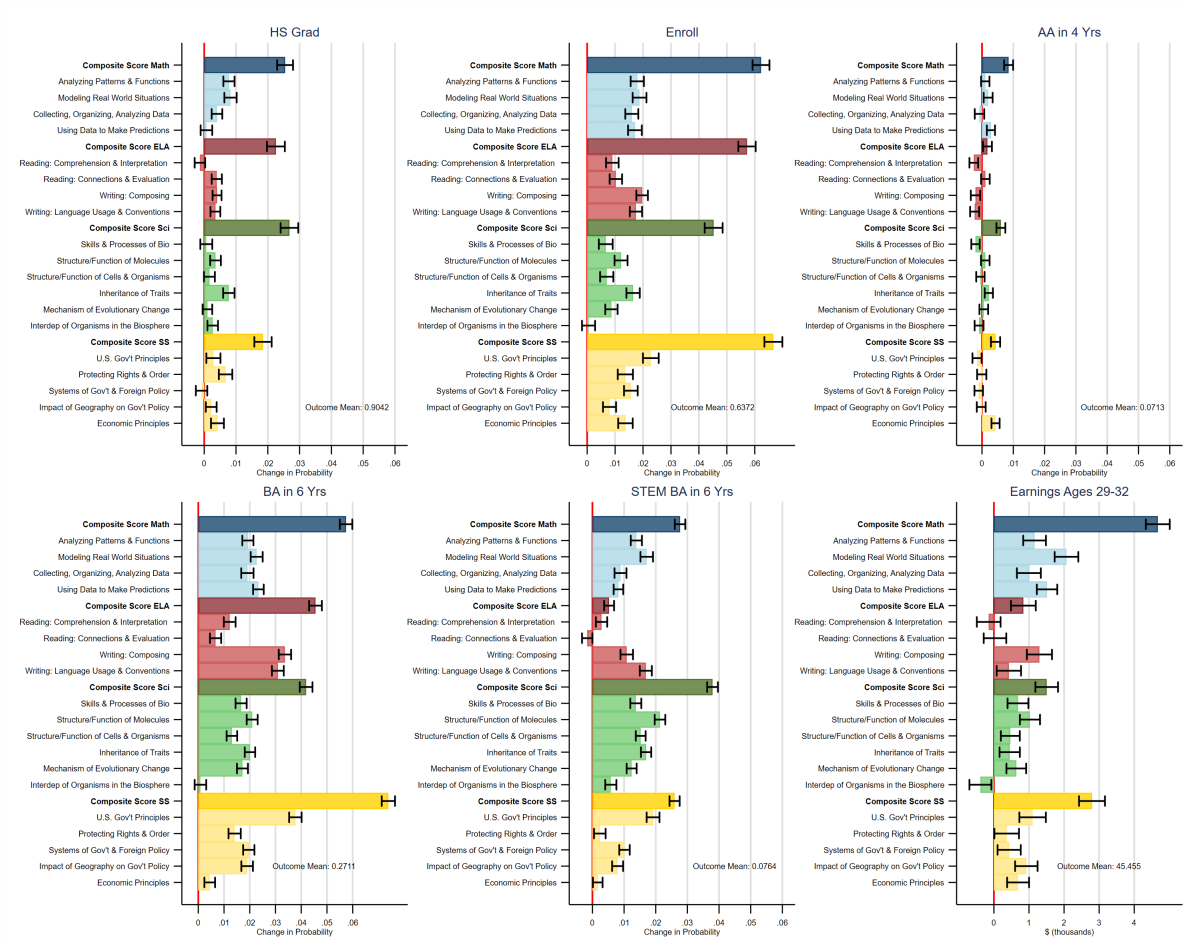
Notes: This figure presents contour plots of mean outcomes by math and ELA test score bins using raw binned data from the elementary and middle school (EMS) sample described in Section 1. Bins are constructed in 0.5-standard-deviation (SD) intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than two SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. High school graduation is measured within four years, college enrollment is defined as on-time relative to 9th grade enrollment, AA receipt is measured within four years, and BA receipt is measured within six years. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure 2: Subscore Decomposition for Educational Attainment and Earnings (EMS Sample)



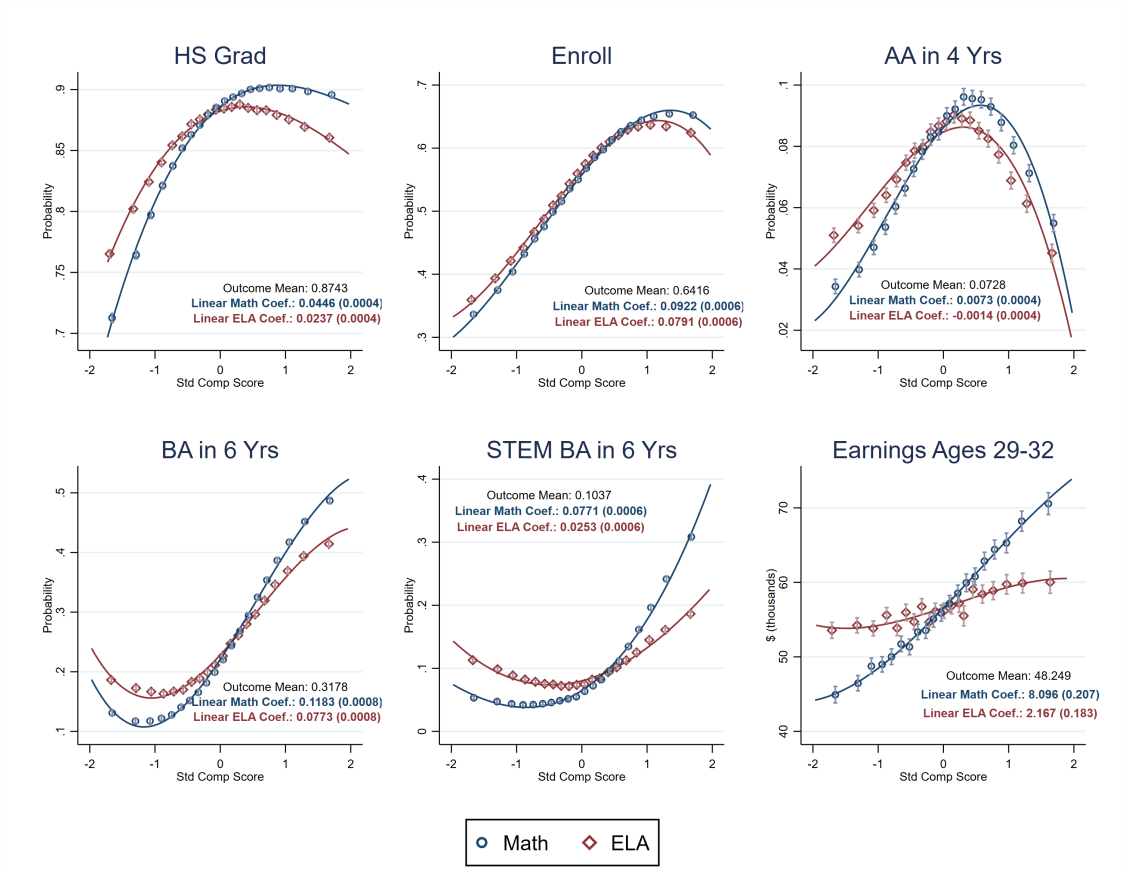
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of educational attainment and earnings on subject achievement measures from our EMS sample. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. Outcome means are shown in the figures. Dark bars report coefficients from the baseline multivariate specification that includes composite (subject-level) math and ELA test scores (Equation 1 in Section 2.1). Light bars report coefficients from specifications that replace one subject's composite score (math or ELA) with the full set of standardized subscores for that subject, while keeping the other subject in composite-score form (Equation 2 in Section 3.1). All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Figure 3: Subscore Decomposition for Educational Attainment and Earnings (HS Sample)



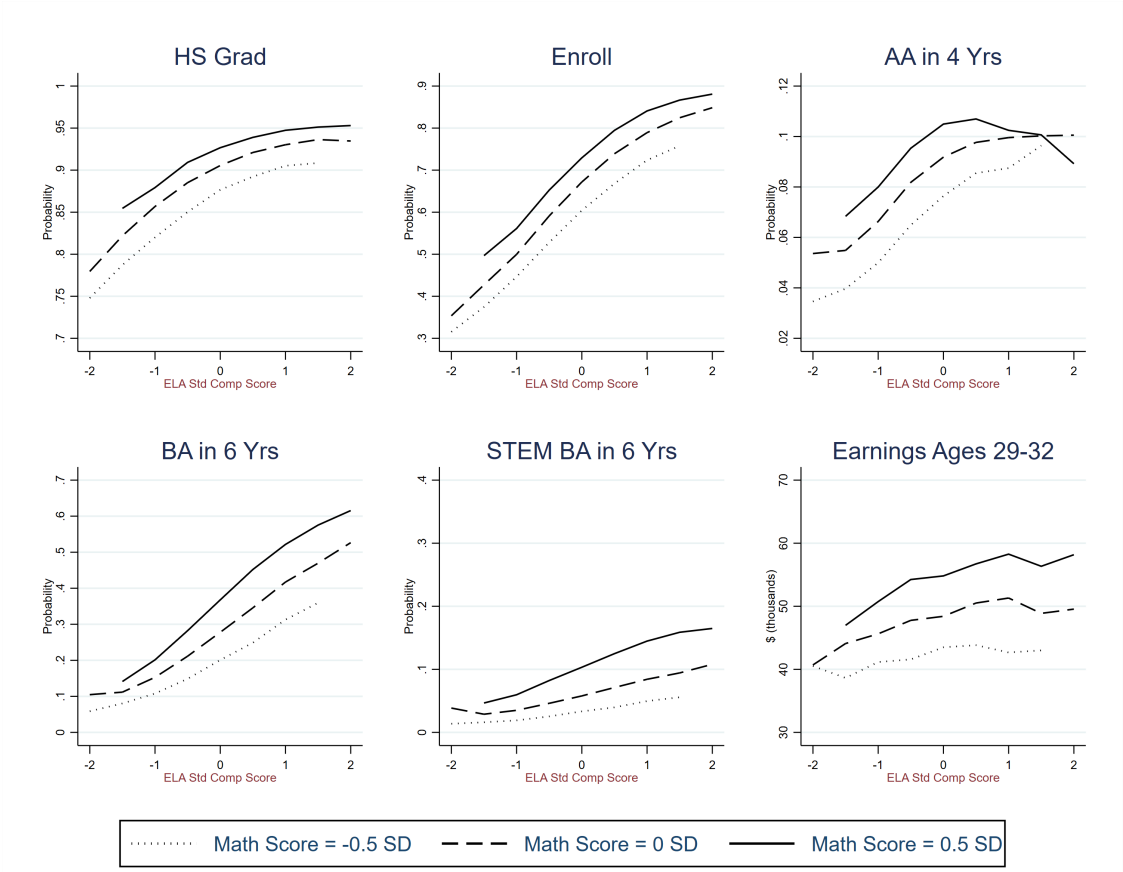
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of educational attainment and earnings on subject achievement measures from our HS sample. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. Outcome means are shown in the figures. Dark bars report coefficients from the baseline multivariate specification that includes composite (subject-level) test scores in Algebra I, English 10, Biology, and U.S. Government (Equation 1 in Section 2.1). Light bars report coefficients from specifications that replace one subject's composite score with the full set of standardized subscores for that subject, while keeping the other subjects in composite-score form (Equation 2 in Section 3.1). All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Figure 4: Non-Linear Relationships Between Test Scores and Outcomes (EMS Sample)



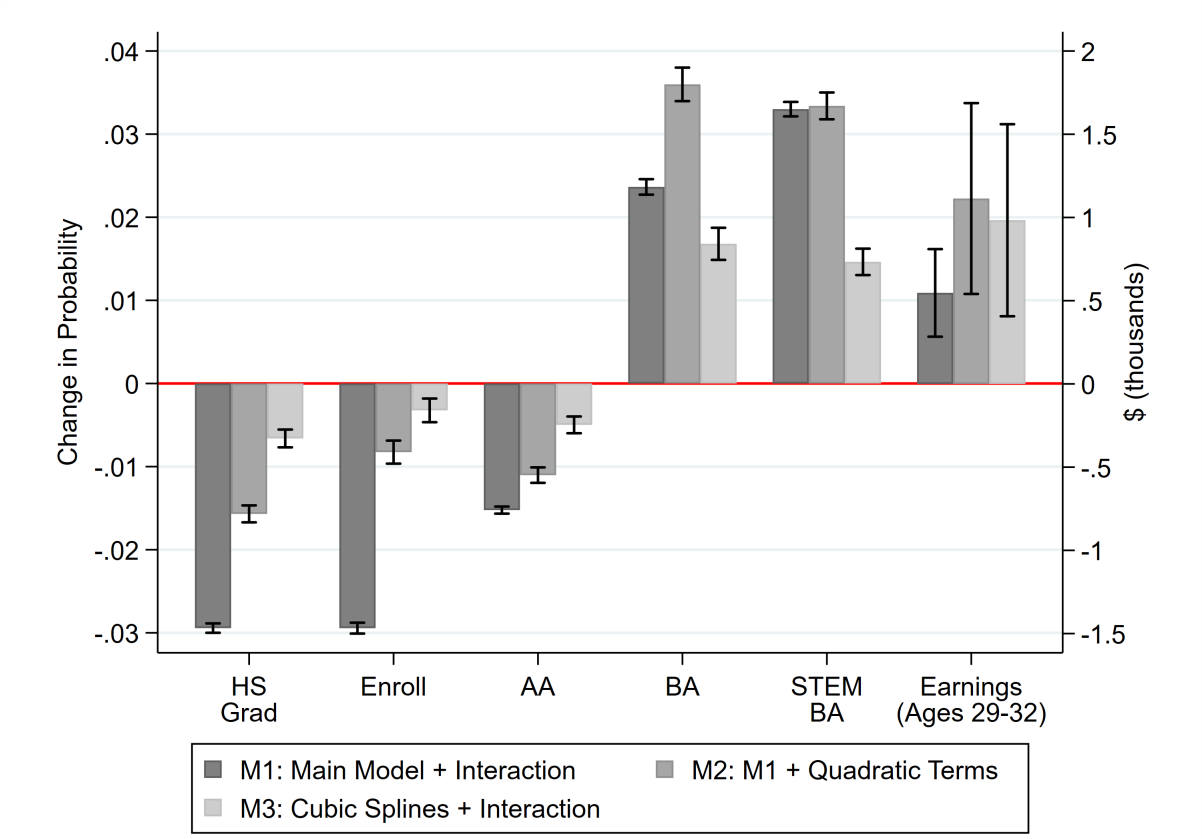
Notes: This figure plots raw binned means of outcomes by math and ELA test scores, with dots representing the means and vertical bars showing 95 percent confidence intervals, using the EMS sample described in Section 1. The sample is restricted to students with standardized test scores within two standard deviations of the mean. The solid curve represents a global cubic polynomial fit. Outcomes include educational attainment and earnings, as described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure 5: Relationship Between ELA Scores and Outcomes Across Math Score Levels (EMS Sample)



Notes: This figure presents line plots of mean outcomes across ELA test score bins for three selected math test score bins: -0.5 SD, 0 SD, and 0.5 SD. The figure is constructed using raw binned data from our EMS sample described in Section 1. Bins are constructed in 0.5-SD intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than two SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure 6: Relationship Between Math-ELA Score Interactions and Student Outcomes (EMS Sample)



Notes: This figure reports coefficients and 95 percent confidence intervals for the interaction between math and ELA test scores from regressions of student outcomes on math and ELA test scores using the EMS sample. The first model (M1) includes math and ELA scores and their interaction. The second model (M2) additionally includes quadratic terms in math and ELA scores to account for non-linear relationships. The third model (M3) uses cubic splines in math and ELA scores to flexibly account for non-linear relationships while retaining the interaction term. The sample and outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Table 1: Summary Statistics for EMS and HS Samples

	EMS Sample			HS Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	SD	N	Mean	SD	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 1. 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: 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 1. 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 2.1, Equation 1, so each column in a panel represents a separate regression. The demographic controls include indicators for gender, race, gender-race interactions, FARMS, ELL, and SPED. 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 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: This table reports estimates from multivariate regressions that include standardized composite test scores and standardized subject GPAs. 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 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 1. The specification is our main multivariate test score model with demographic controls in Section 2.1, 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 5: 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 2.1, 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 4. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 1. Robust standard errors are clustered at the student level.

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

Table 6: Heterogeneity Estimates of the Relationship Between Test Scores and Outcomes for EMS Sample

	(1) Females	(2) Males	(3) FARMS	(4) Non-FARMS	(5) Q1	(6) Q2	(7) Q3
Panel A: HS Grad in 4 Yrs							
Comp Score Math	0.0356***	0.0521***	0.0719***	0.0280***	0.0987***	0.0417***	0.0105***
	(0.0006)	(0.0006)	(0.0008)	(0.0004)	(0.0011)	(0.0009)	(0.0005)
Comp Score ELA	0.0210***	0.0281***	0.0420***	0.0125***	0.0603***	0.0179***	0.0029***
	(0.0005)	(0.0006)	(0.0008)	(0.0004)	(0.0011)	(0.0009)	(0.0004)
Adj R-squared	0.0599	0.0985	0.0672	0.0328	0.0597	0.0272	0.0111
Outcome Mean	[0.8987]	[0.8503]	[0.7935]	[0.9325]	[0.7650]	[0.9028]	[0.9551]
N	1,629,408	1,652,199	1,373,427	1,908,180	1,093,889	1,093,850	1,093,868
Panel B: BA in 6 Yrs							
Comp Score Math	0.1253***	0.1129***	0.0743***	0.1393***	0.0491***	0.1494***	0.1024***
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0010)	(0.0020)	(0.0016)
Comp Score ELA	0.0825***	0.0695***	0.0530***	0.0869***	0.0269***	0.1001***	0.0603***
	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0009)	(0.0020)	(0.0014)
Adj R-squared	0.2610	0.2428	0.1319	0.2093	0.0543	0.0917	0.1094
Outcome Mean	[0.3764]	[0.2597]	[0.1294]	[0.4352]	[0.0828]	[0.2822]	[0.5988]
N	704,202	710,036	542,826	871,412	472,646	484,618	456,974
Panel C: Earnings Ages 29-32							
Comp Score Math	8,804***	7,528***	6,649***	8,680***	5,956***	9,185***	8,503***
	(278)	(304)	(301)	(270)	(332)	(472)	(530)
Comp Score ELA	2,530***	1,696***	2,557***	1,902***	2,367***	2,679***	1,018**
	(234)	(284)	(262)	(243)	(273)	(468)	(463)
Adj R-squared	0.1484	0.1434	0.1095	0.1158	0.0997	0.0717	0.0481
Outcome Mean	[44,957]	[51,609]	[38,384]	[53,857]	[36,008]	[49,028]	[62,490]
N	46,459	45,540	33,340	58,659	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. Estimates by gender are shown in columns (1)-(2), by eligibility for Free and Reduced-price Meals (FARMS) in columns (3)-(4), and by achievement terciles in columns (5)-(7). Achievement terciles are defined based on the average of students' math and ELA test scores, ranging from the lowest achievement tercile (Q1) to the highest achievement tercile (Q3). 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 1. The specification is our main multivariate test score model with demographic controls in Section 2.1, 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 7: 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 value-added (TVA) for students' test scores on educational attainment and earnings. The sample is our student-year level TVA sample described in Section 1, 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 2.5, Equation 7. 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, and SPED; 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.

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

Appendix A: Tables and Figures

Table A1: 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 A2: 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 A3: 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 1. 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 2.1, 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, and SPED. 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 A4: Heterogeneity Estimates of the Relationship Between Test Scores and Outcomes for HS Sample

	(1) Females	(2) Males	(3) FARMS	(4) Non-FARMS	(5) Q1	(6) Q2	(7) Q3
Panel A: HS Grad in 4 Yrs							
Comp Score Math	0.0138***	0.0192***	0.0256***	0.0116***	0.0336***	0.0134***	0.0039***
	(0.0011)	(0.0012)	(0.0015)	(0.0009)	(0.0015)	(0.0014)	(0.0008)
Comp Score ELA	0.0136***	0.0260***	0.0294***	0.0153***	0.0423***	0.0139***	0.0053***
	(0.0012)	(0.0013)	(0.0016)	(0.0010)	(0.0017)	(0.0015)	(0.0008)
Comp Score Sci	0.0171***	0.0201***	0.0234***	0.0152***	0.0300***	0.0180***	0.0042***
	(0.0013)	(0.0014)	(0.0016)	(0.0011)	(0.0015)	(0.0016)	(0.0009)
Comp Score SS	0.0210***	0.0313***	0.0343***	0.0199***	0.0413***	0.0122***	0.0004
	(0.0014)	(0.0014)	(0.0017)	(0.0012)	(0.0016)	(0.0017)	(0.0010)
Adj R-squared	0.0534	0.0896	0.0749	0.0539	0.0610	0.0093	0.0044
Outcome Mean	[0.9363]	[0.8982]	[0.8687]	[0.9452]	[0.8305]	[0.9466]	[0.9751]
N	229,886	227,271	166,298	290,859	152,386	152,386	152,385
Panel B: BA in 6 Yrs							
Comp Score Math	0.0638***	0.0532***	0.0280***	0.0789***	0.0107***	0.1053***	0.0834***
	(0.0016)	(0.0013)	(0.0011)	(0.0016)	(0.0007)	(0.0030)	(0.0030)
Comp Score ELA	0.0724***	0.0561***	0.0412***	0.0732***	0.0126***	0.0754***	0.0752***
	(0.0018)	(0.0014)	(0.0012)	(0.0016)	(0.0008)	(0.0031)	(0.0029)
Comp Score Sci	0.0441***	0.0278***	0.0265***	0.0425***	0.0088***	0.0597***	0.0366***
	(0.0017)	(0.0013)	(0.0011)	(0.0017)	(0.0007)	(0.0036)	(0.0032)
Comp Score SS	0.0885***	0.0553***	0.0439***	0.0880***	0.0173***	0.1216***	0.0876***
	(0.0019)	(0.0013)	(0.0011)	(0.0018)	(0.0007)	(0.0034)	(0.0037)
Adj R-squared	0.2638	0.2163	0.1382	0.2328	0.0342	0.0779	0.1103
Outcome Mean	[0.3471]	[0.2329]	[0.1291]	[0.3817]	[0.0552]	[0.2577]	[0.5857]
N	170,285	169,378	123,087	216,576	117,225	116,450	105,988
Panel C: Earnings Ages 29-32							
Comp Score Math	4,630***	4,640***	3,365***	5,771***	2,777***	7,349***	7,236***
	(241)	(246)	(222)	(257)	(182)	(466)	(662)
Comp Score ELA	1,592***	369	850***	722***	890***	1,138**	-1,049
	(257)	(248)	(232)	(268)	(189)	(522)	(677)
Comp Score Sci	1,802***	1,174***	1,357***	1,633***	798***	1,666***	2,058***
	(219)	(241)	(211)	(248)	(168)	(523)	(782)
Comp Score SS	3,566***	2,176***	2,602***	2,725***	1,165***	4,420***	3,362***
	(262)	(264)	(244)	(278)	(204)	(510)	(730)
Adj R-squared	0.1275	0.1224	0.0973	0.1154	0.0801	0.0579	0.0372
Outcome Mean	[42,219]	[48,746]	[38,155]	[49,613]	[35,159]	[47,784]	[60,107]
N	40,682	39,680	29,251	51,111	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 A5: Heterogeneity Estimates of the Relationship Between Test Scores and STEM BA Receipt

	(1) Females	(2) Males	(3) FARMS	(4) Non-FARMS	(5) Q1	(6) Q2	(7) Q3
Panel A: EMS Sample							
Comp Score Math	0.0699***	0.0832***	0.0333***	0.0995***	0.0122***	0.0638***	0.1300***
	(0.0009)	(0.0009)	(0.0007)	(0.0009)	(0.0005)	(0.0012)	(0.0016)
Comp Score ELA	0.0232***	0.0287***	0.0142***	0.0301***	0.0023***	0.0253***	0.0412***
	(0.0008)	(0.0008)	(0.0006)	(0.0008)	(0.0004)	(0.0011)	(0.0014)
Adj R-squared	0.1234	0.1665	0.0802	0.1405	0.0206	0.0402	0.1000
Outcome Mean	[0.0947]	[0.1126]	[0.0354]	[0.1462]	[0.0131]	[0.0647]	[0.2387]
N	703,994	709,819	542,760	871,053	472,588	484,482	456,743
Panel B: HS Sample							
Comp Score Math	0.0275***	0.0320***	0.0111***	0.0423***	0.0010***	0.0307***	0.0900***
	(0.0011)	(0.0010)	(0.0006)	(0.0012)	(0.0003)	(0.0015)	(0.0028)
Comp Score ELA	0.0163***	0.0224***	0.0120***	0.0209***	0.0015***	0.0086***	0.0352***
	(0.0012)	(0.0010)	(0.0007)	(0.0012)	(0.0003)	(0.0016)	(0.0026)
Comp Score Sci	0.0386***	0.0358***	0.0137***	0.0565***	0.0022***	0.0378***	0.1268***
	(0.0011)	(0.0010)	(0.0006)	(0.0013)	(0.0002)	(0.0018)	(0.0033)
Comp Score SS	0.0223***	0.0199***	0.0131***	0.0246***	0.0021***	0.0257***	0.0494***
	(0.0011)	(0.0009)	(0.0006)	(0.0012)	(0.0003)	(0.0017)	(0.0032)
Adj R-squared	0.1147	0.1406	0.0743	0.1375	0.0071	0.0271	0.1139
Outcome Mean	[0.0798]	[0.0942]	[0.0343]	[0.1169]	[0.0062]	[0.0450]	[0.2225]
N	169,969	169,132	122,995	216,106	117,137	116,228	105,736

Notes: This table shows estimates for the multivariate relationship between test scores and STEM BA receipt within six years across subgroups for the EMS sample (Panel A) and the HS sample (Panel B). Estimates by gender are shown in columns (1)-(2), by eligibility for Free and Reduced-price Meals (FARMS) in columns (3)-(4), and by achievement terciles in columns (5)-(7). Achievement terciles are defined based on the average of students' observable composite test scores in each respective sample, ranging from the lowest achievement tercile (Q1) to the highest achievement tercile (Q3). The sample and the outcome are described in Section 1. The specification is our main multivariate test score model with demographic controls in Section 2.1, 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 A6: 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

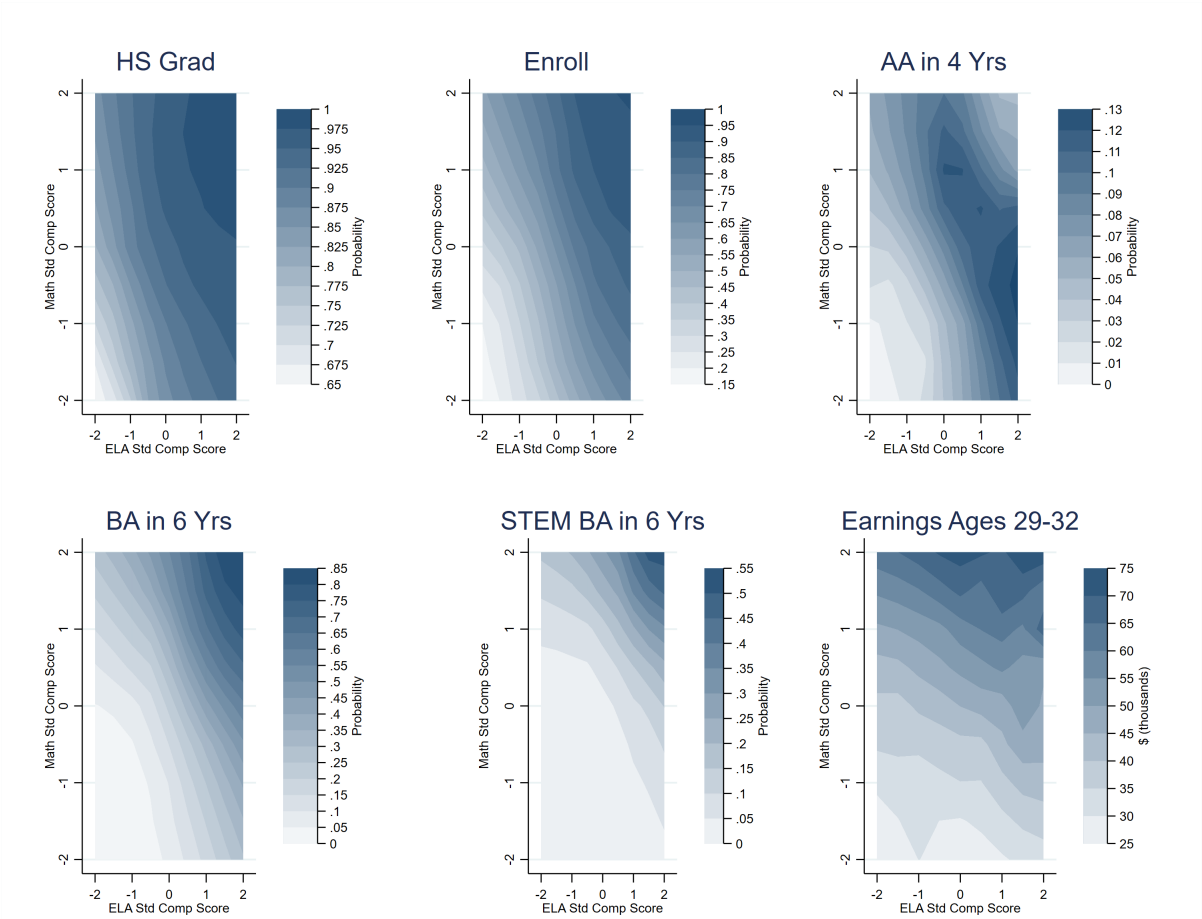
Notes: This table reports the effect of teacher value-added (TVA) on current-year student test scores, serving as a validity check of the TVA identification strategy. The sample includes student-year observations from the TVA sample (Section 1) with non-missing leave-out TVA measures, where TVA is scaled in student-level test score standard deviations and estimated using data from other years taught by the same teacher (Section 2.5). Each column shows the effect of subject-specific TVA on the corresponding standardized test score. The specification follows our main TVA model (Equation 7), 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, and SPED); 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 A7: Multivariate Relationship Between Test Scores and Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Emp 22-24	Emp 25-28	Emp 29-32	Earn 22-24 Uncond	Earn 25-28 Uncond	Earn 29-32 Uncond	Earn 22-24 Cond	Earn 25-28 Cond	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

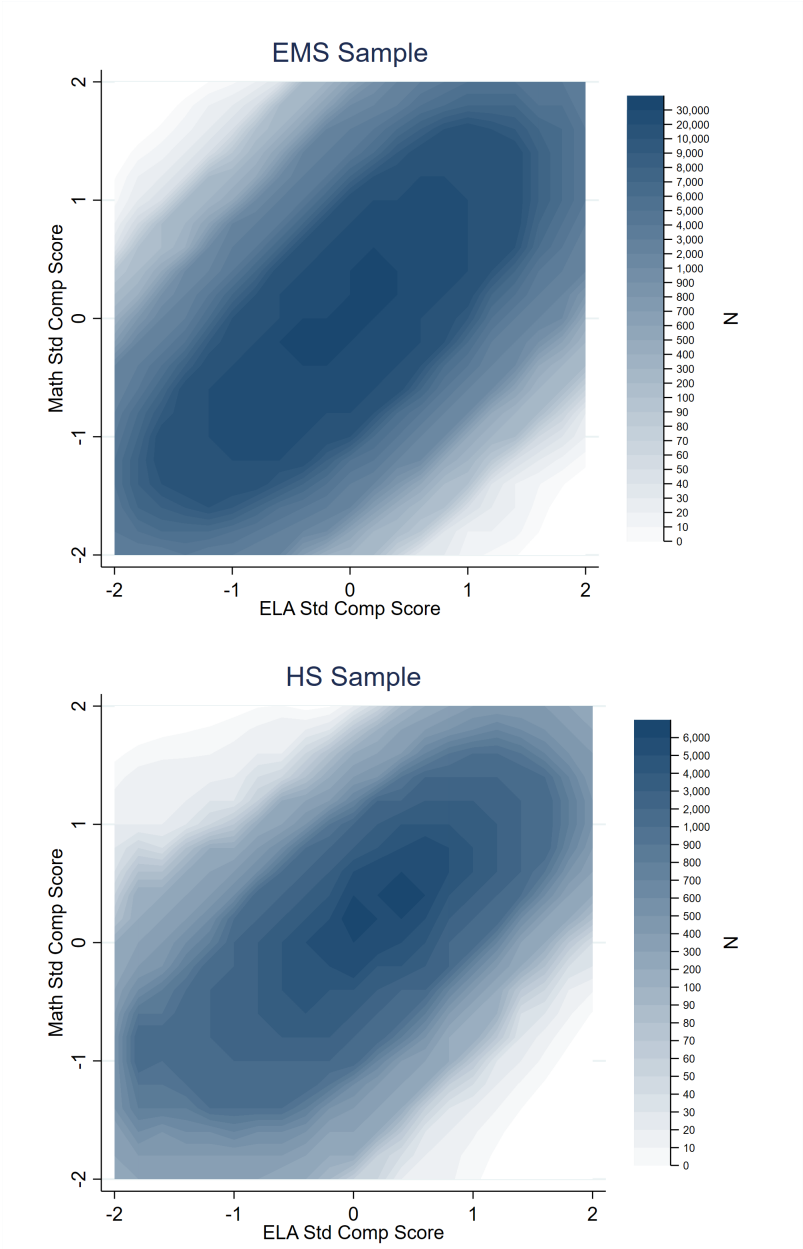
Notes: This table shows estimates for the multivariate relationship between test scores and employment and earnings. Panel A shows results for the EMS sample and Panel B shows results for the HS sample. The samples are described in Section 1. Columns (1)-(3) show estimates for employment at ages 22-24, 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 2, columns (6)-(8). The specification is our main multivariate test score model with demographic controls in Section 2.1, 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

Figure A1: Joint Relationship Between Math and ELA Test Scores and Student Outcomes (HS Sample)



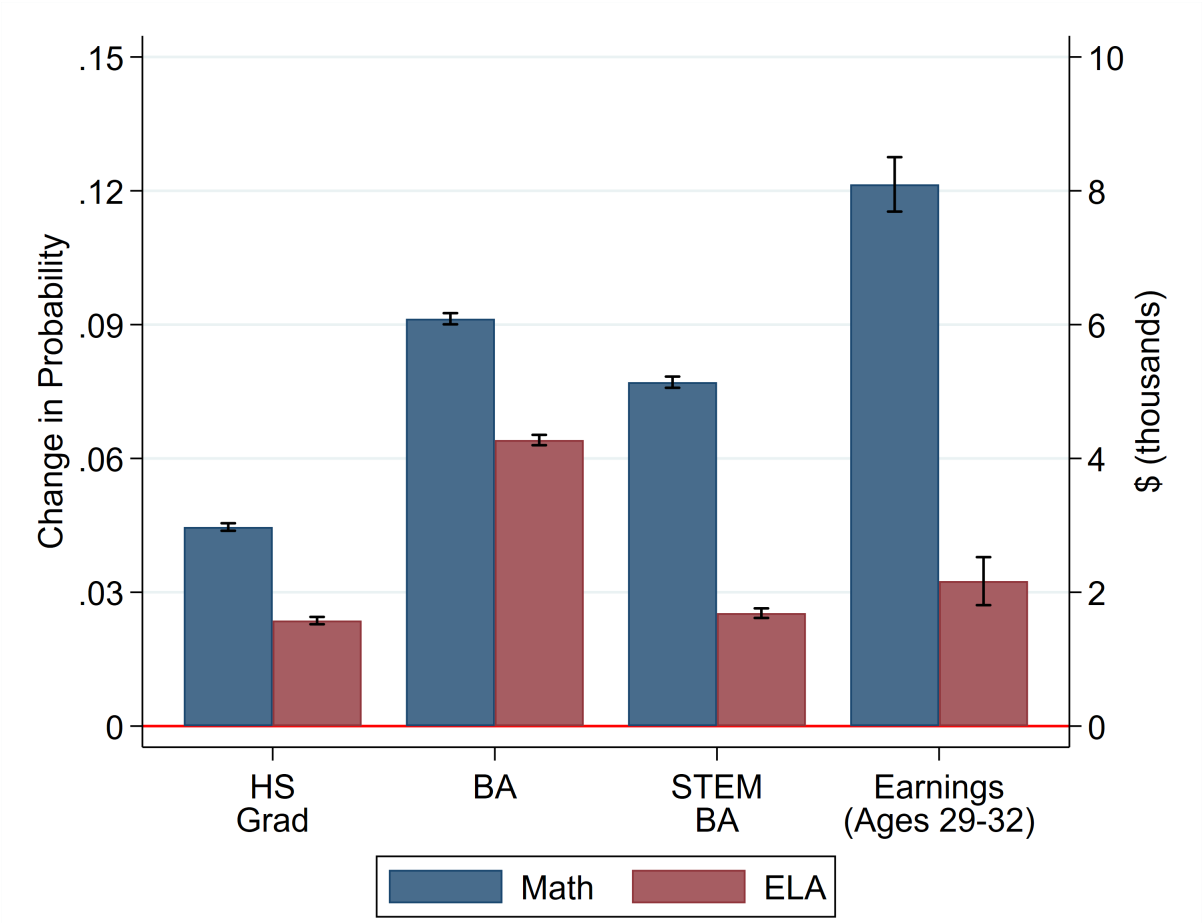
Notes: This figure presents contour plots of mean outcomes by math and ELA test score bins using raw binned data from the high school (HS) sample described in Section 1. Bins are constructed in 0.5-standard-deviation (SD) intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than 1.5 SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. High school graduation is measured within four years, college enrollment is defined as on-time relative to 9th grade enrollment, AA receipt is measured within four years, and BA receipt is measured within six years. Average annual earnings are conditional on positive observed earnings and are reported in 2024 dollars.

Figure A2: Joint Distribution of Math and ELA Scores



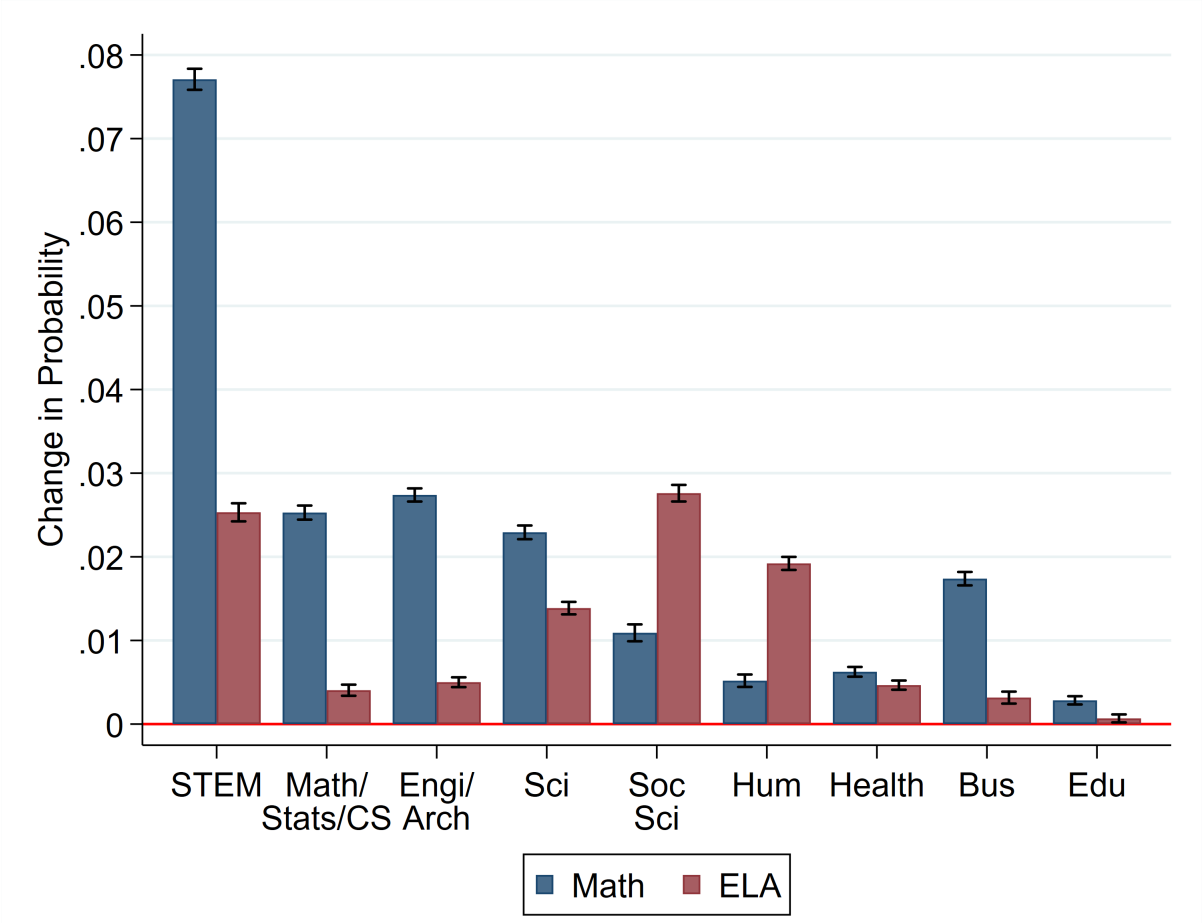
Notes: This figure presents contour plots based on raw binned data from the EMS and HS samples. Bins are constructed in 0.2-SD intervals for math and ELA test scores, and the number of observations is calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two standard deviations of the mean. Bins with 10 or fewer observations are omitted to satisfy minimum cell-size requirements.

Figure A3: Multivariate Relationship between Test Scores and Outcomes (EMS Sample)



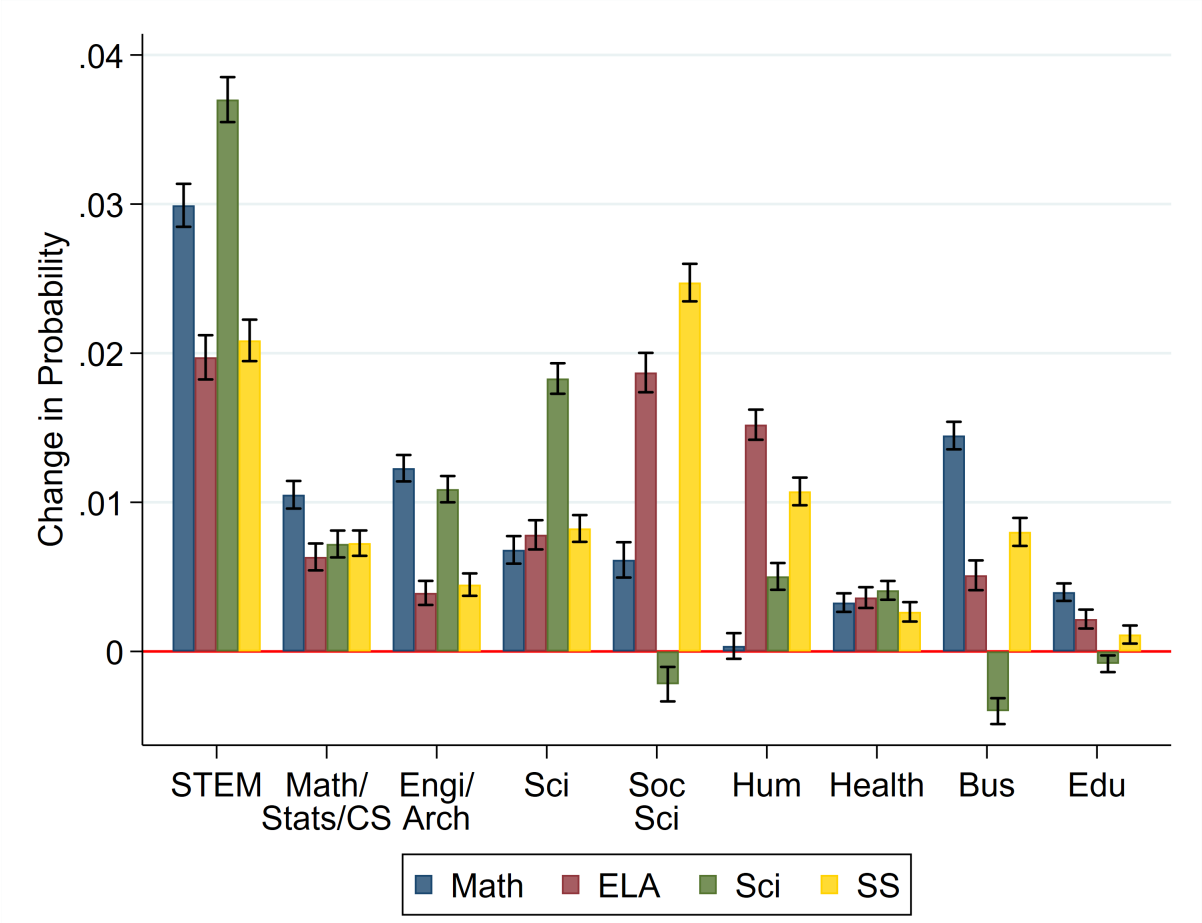
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of educational attainment and earnings on subject achievement measures in our EMS sample. This plot shows a subset of our coefficients from Table 2. The sample and outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. The specification is our main multivariate test score model with demographic controls in Section 2.1, Equation 1. The demographic controls include indicators for gender, race, gender-race interactions, FARMS, ELL, and SPED.

Figure A4: Multivariate Relationship between Test Scores and BA Field in EMS Sample



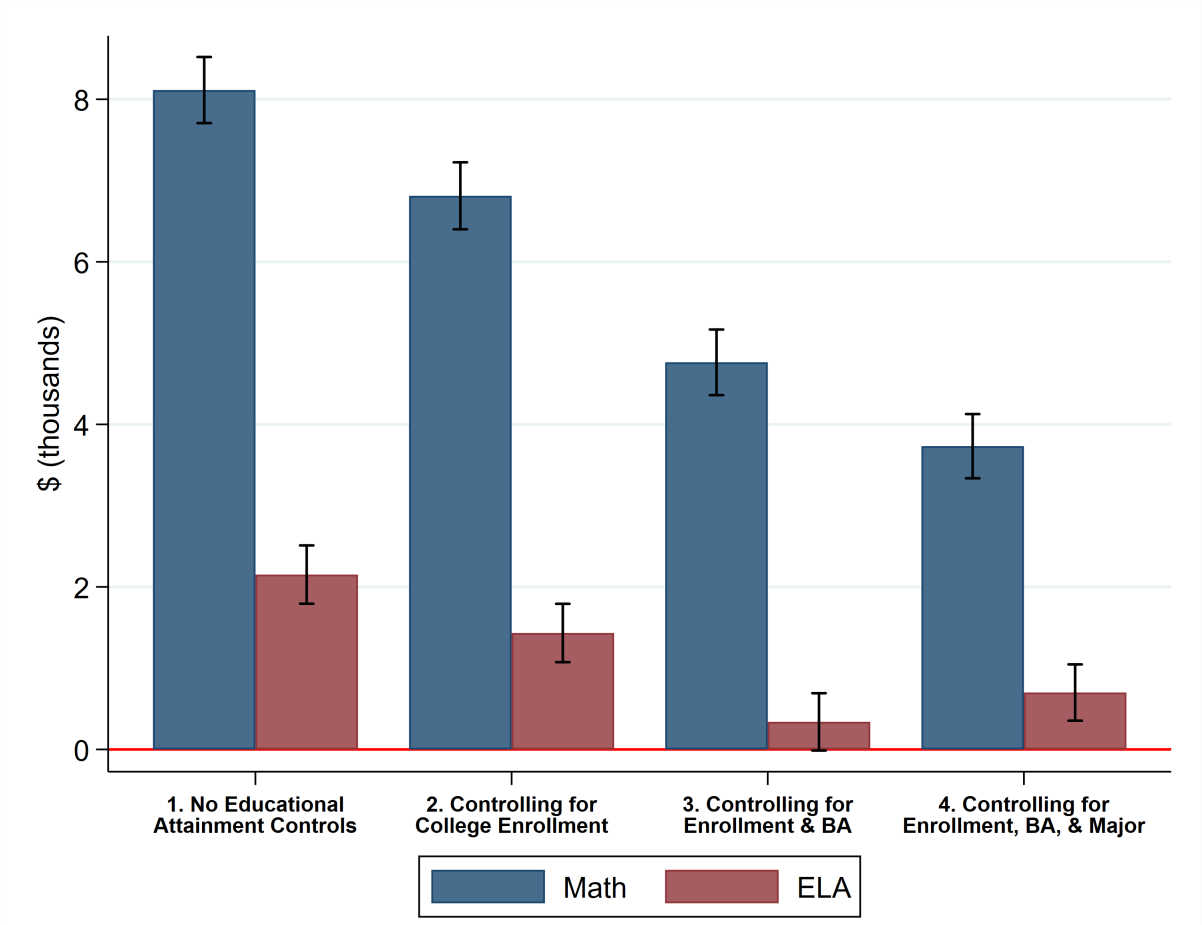
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions showing the relationship between test scores and BA degree field outcomes within six years in our EMS sample. This plot shows a subset of our coefficients from Table 4. The sample and outcomes are described in Section 1. The specification is our main multivariate test score model with demographic controls in Section 2.1, Equation 1.

Figure A5: Multivariate Relationship between Test Scores and BA Field in HS Sample



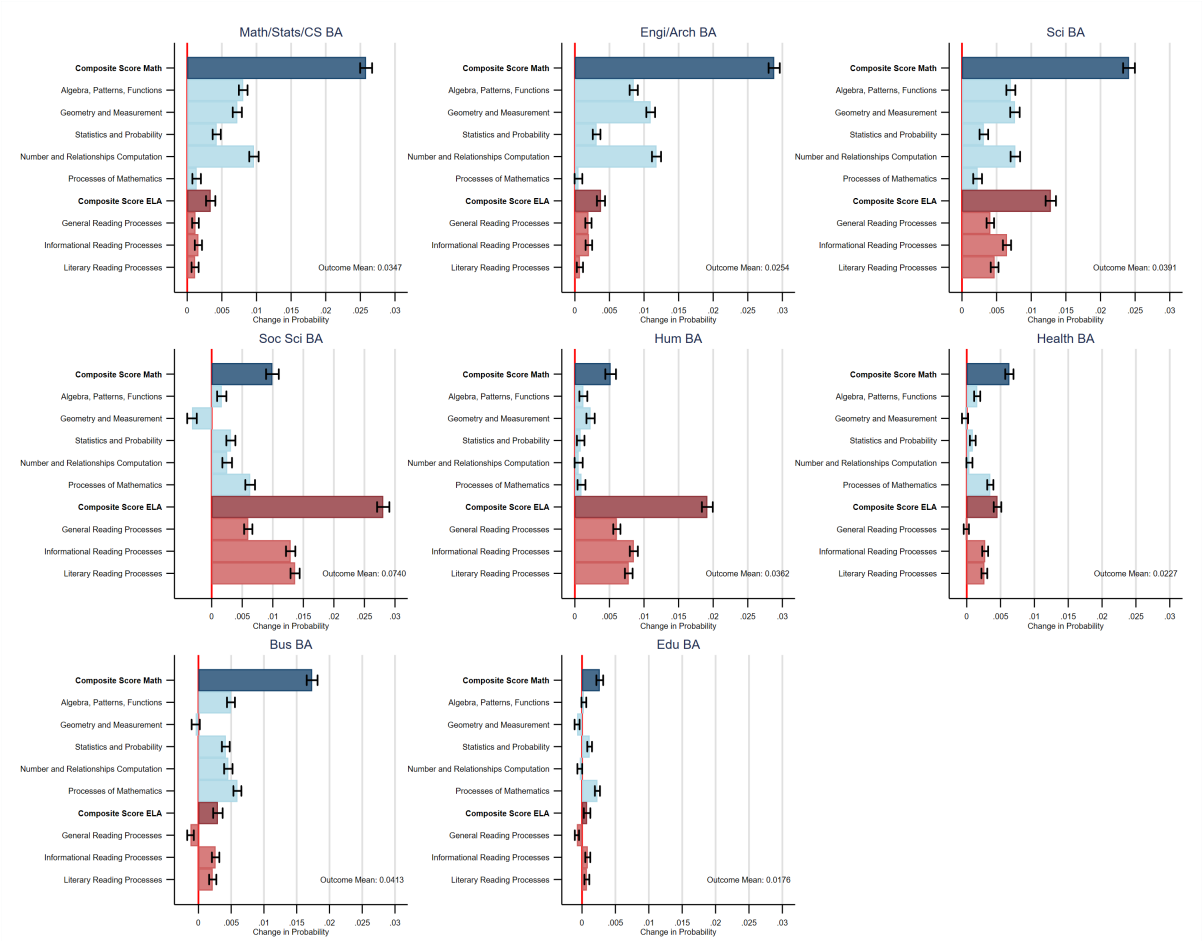
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions showing the relationship between test scores and BA degree field outcomes within six years in our HS sample. This plot shows a subset of our coefficients from Table 4. The sample and outcomes are described in Section 1. The specification is our main multivariate test score model with demographic controls in Section 2.1, Equation 1.

Figure A6: Test Scores and Earnings: Mediation Analysis Across Specifications in EMS Sample



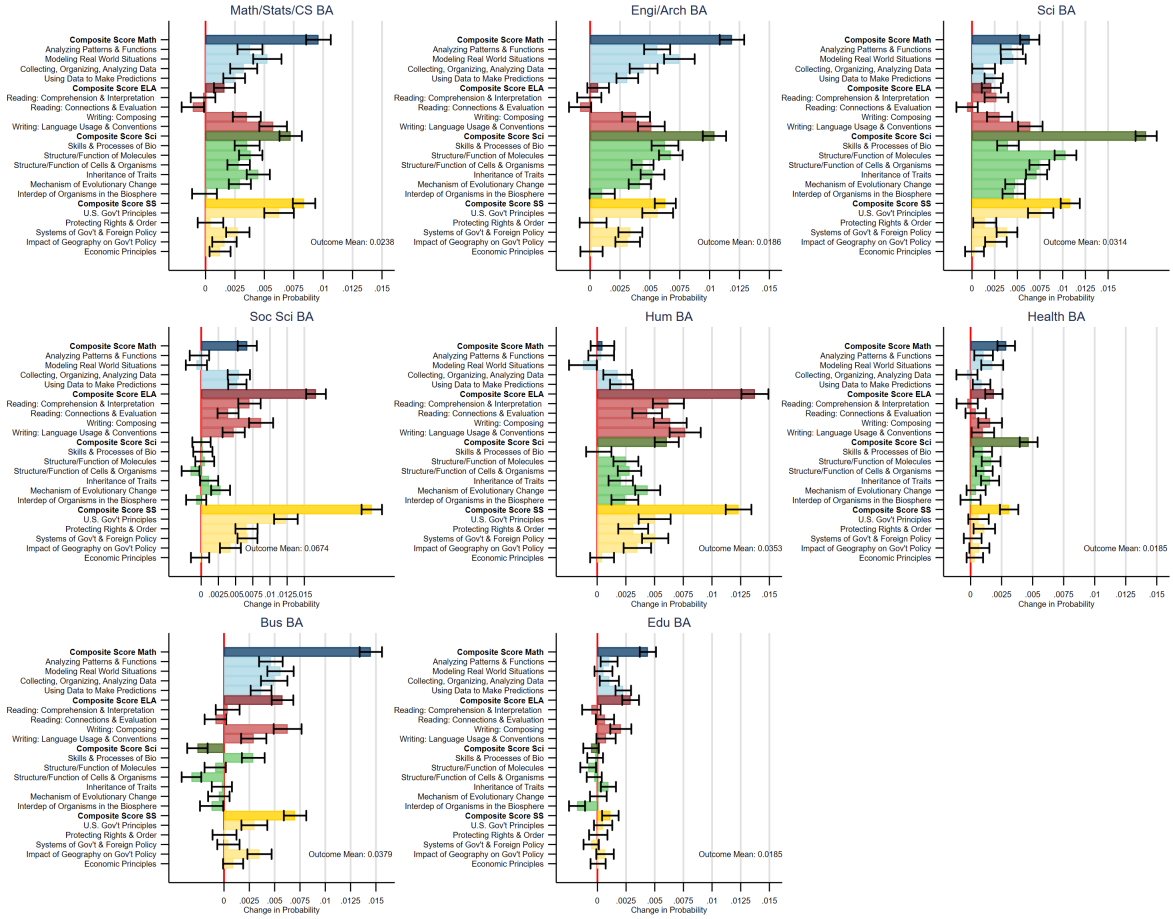
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions showing the relationship between test scores and earnings using different specifications in our EMS sample. This plot shows a subset of our coefficients from Table 5. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. Column (1) shows our main multivariate test score model specification from Section 2.1, 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 4. The sample is described in Section 1.

Figure A7: Subscore Decomposition for BA Major Outcomes (EMS Sample)



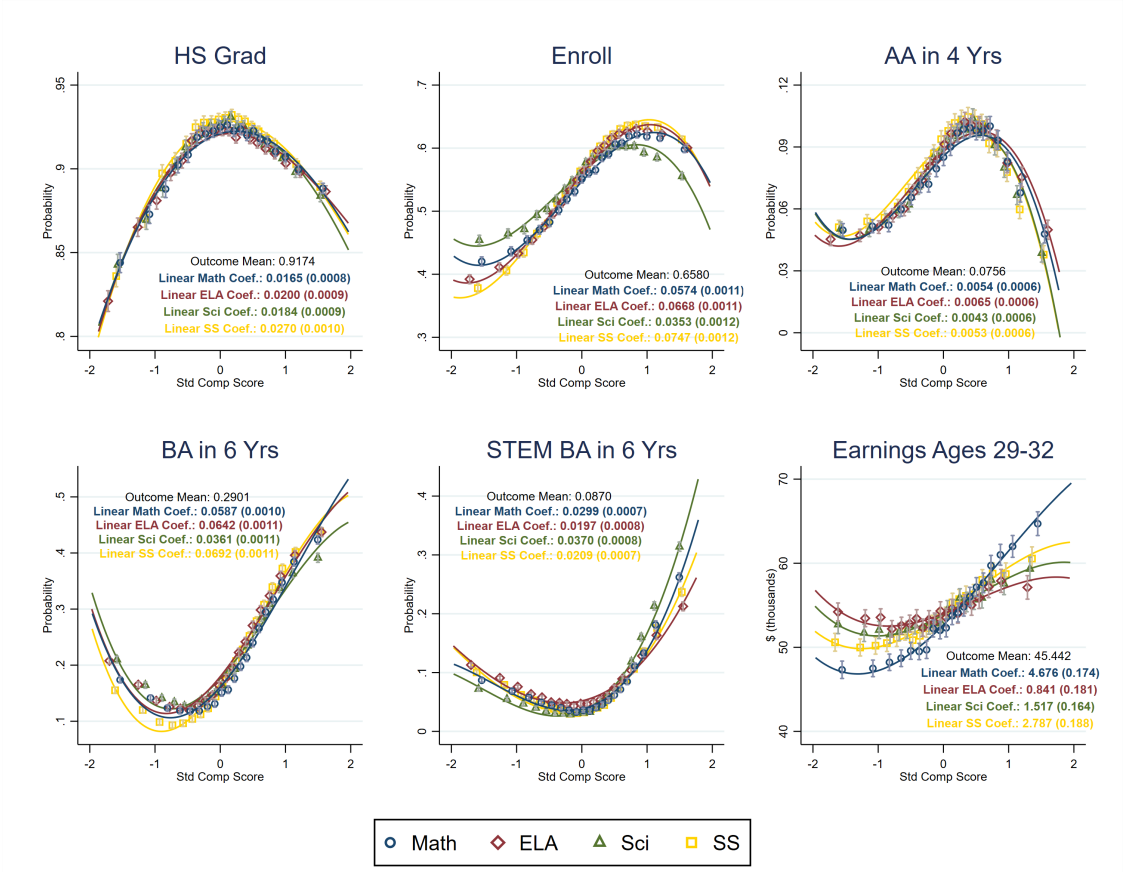
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of BA degree field within six years on subject achievement measures in our EMS sample. All outcomes are described in Section 1 and outcome means are shown in the figures. Dark bars report coefficients from the baseline multivariate specification that includes composite (subject-level) math and ELA test scores (Equation 1 in Section 2.1). Light bars report coefficients from specifications that replace one subject's composite score (math or ELA) with the full set of standardized subscores for that subject, while keeping the other subject in composite-score form (Equation 2 in Section 3.1). All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Figure A8: Subscore Decomposition for BA Major Outcomes (HS Sample)



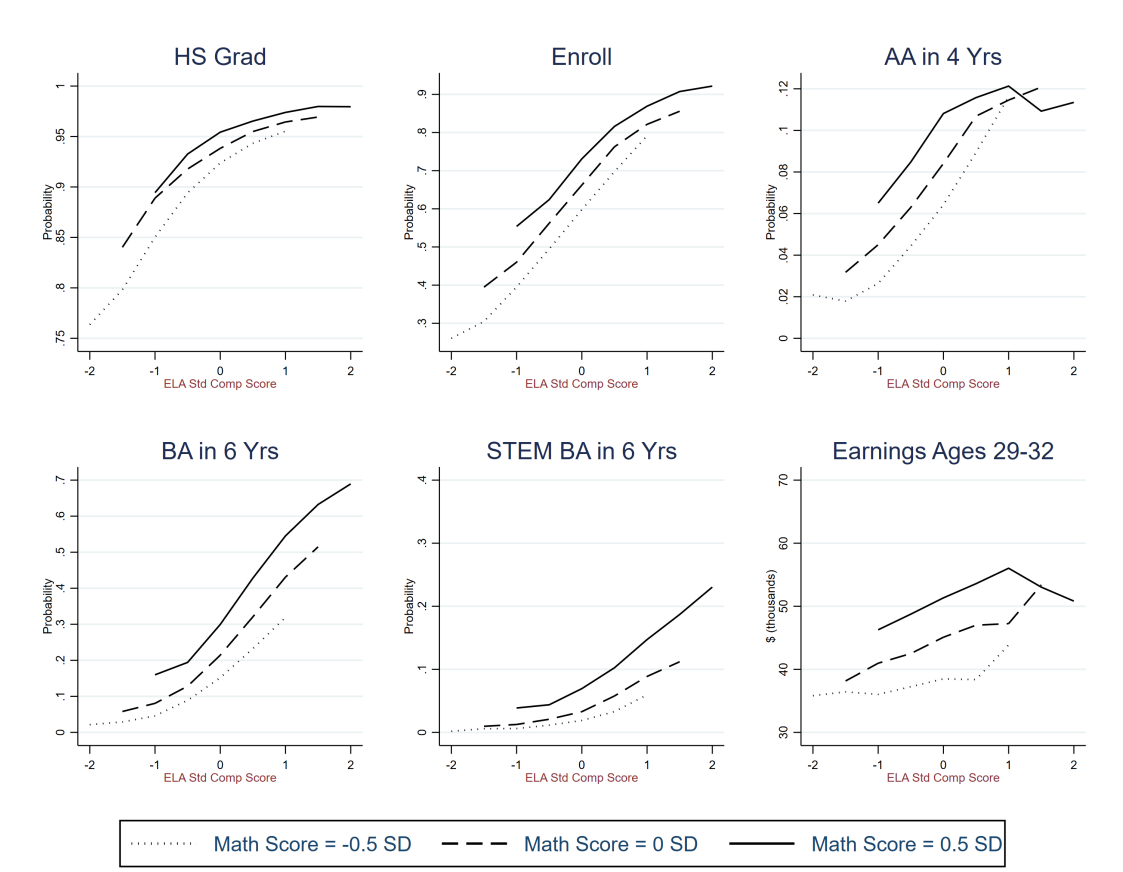
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of BA degree field within six years on subject achievement measures in our HS sample. All outcomes are described in Section 1 and outcome means are shown in the figures. Dark bars report coefficients from the baseline multivariate specification that includes composite (subject-level) test scores in Algebra I, English 10, Biology, and U.S. Government (Equation 1 in Section 2.1). Light bars report coefficients from specifications that replace one subject's composite score with the full set of standardized subscores for that subject, while keeping the other subjects in composite-score form (Equation 2 in Section 3.1). All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Figure A9: Non-Linear Relationships Between Test Scores and Outcomes (HS Sample)



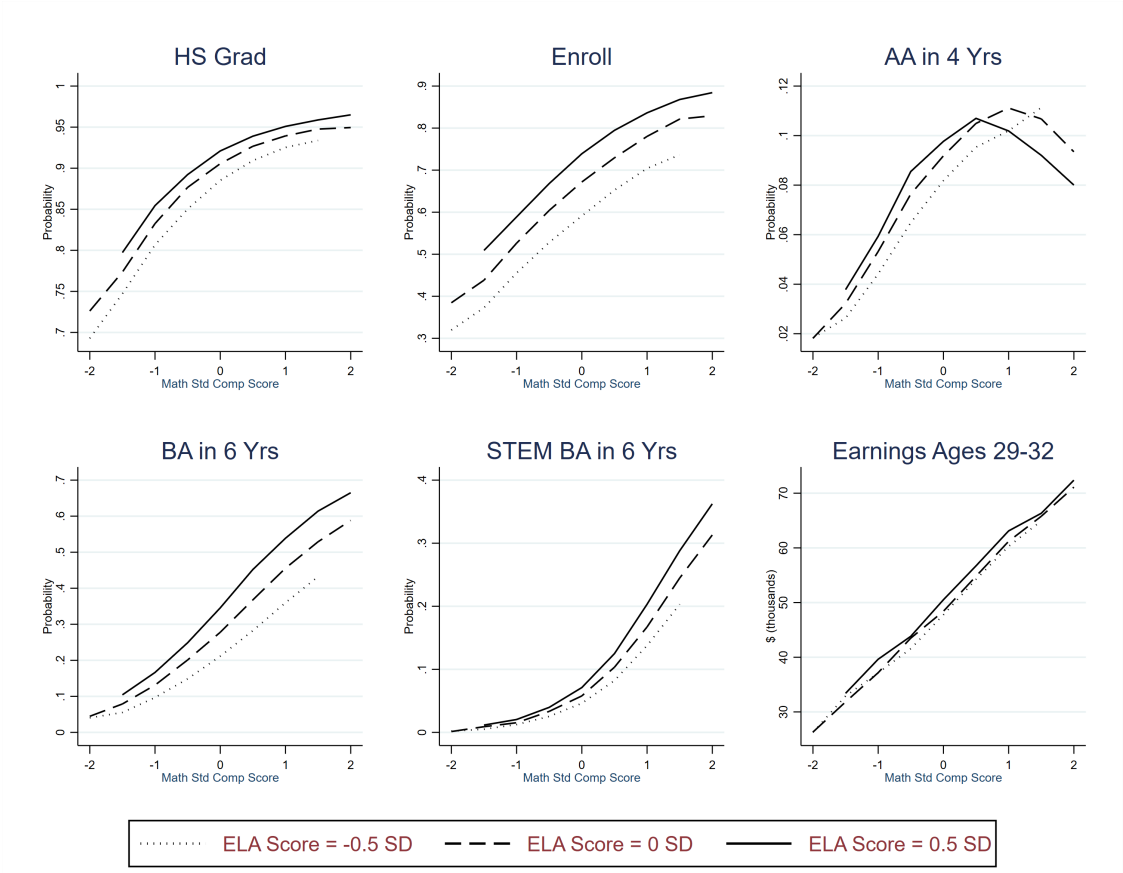
Notes: This figure plots raw binned means of outcomes by Algebra I, English 10, Biology, and U.S. Government test scores, with dots representing the means and vertical bars showing 95 percent confidence intervals, using the HS sample described in Section 1. The sample is restricted to students with standardized test scores within two standard deviations of the mean. The solid curve represents a global cubic polynomial fit. Outcomes include educational attainment and earnings, as described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure A10: Relationship Between ELA Scores and Outcomes Across Math Score Levels (HS Sample)



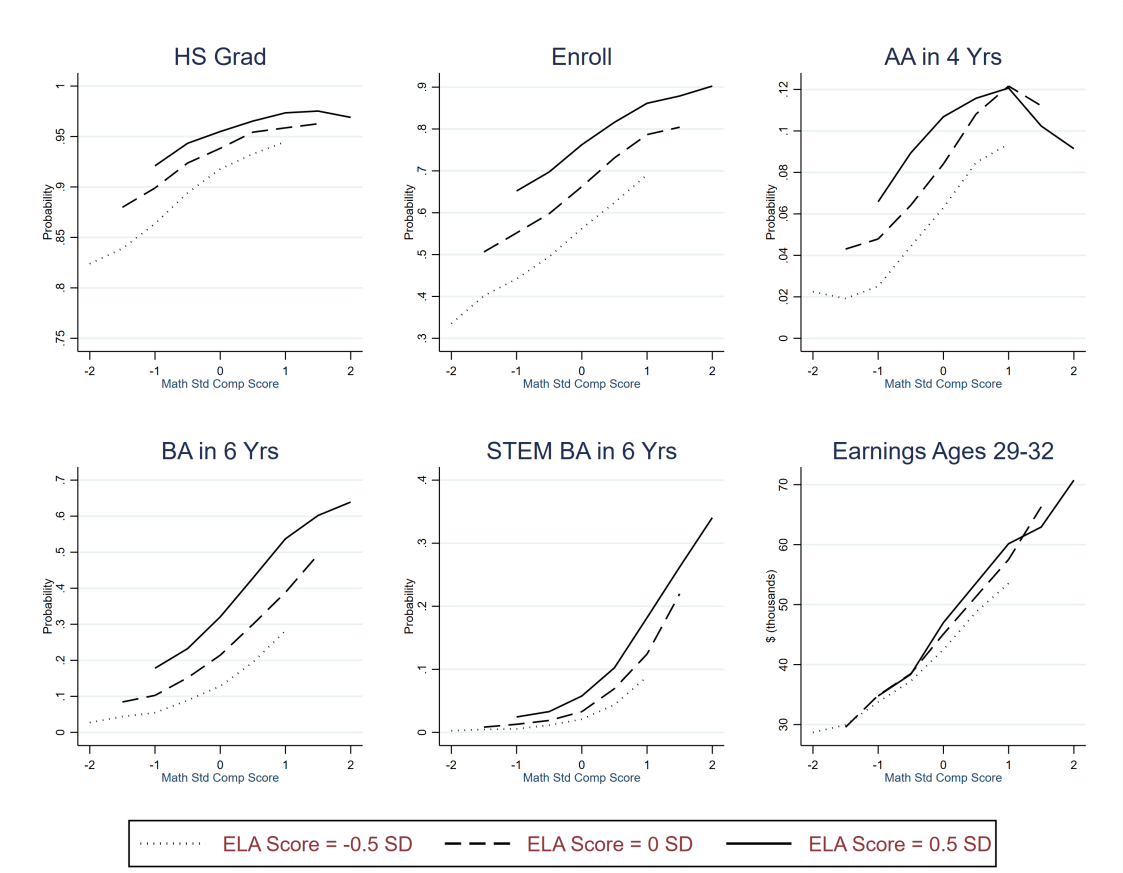
Notes: This figure presents line plots of mean outcomes across ELA test score bins for three selected math test score bins: -0.5 SD, 0 SD, and 0.5 SD. The figure is constructed using raw binned data from our HS sample described in Section 1. Bins are constructed in 0.5-SD intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than 1.5 SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure A11: Relationship Between Math Scores and Outcomes Across ELA Score Levels (EMS Sample)



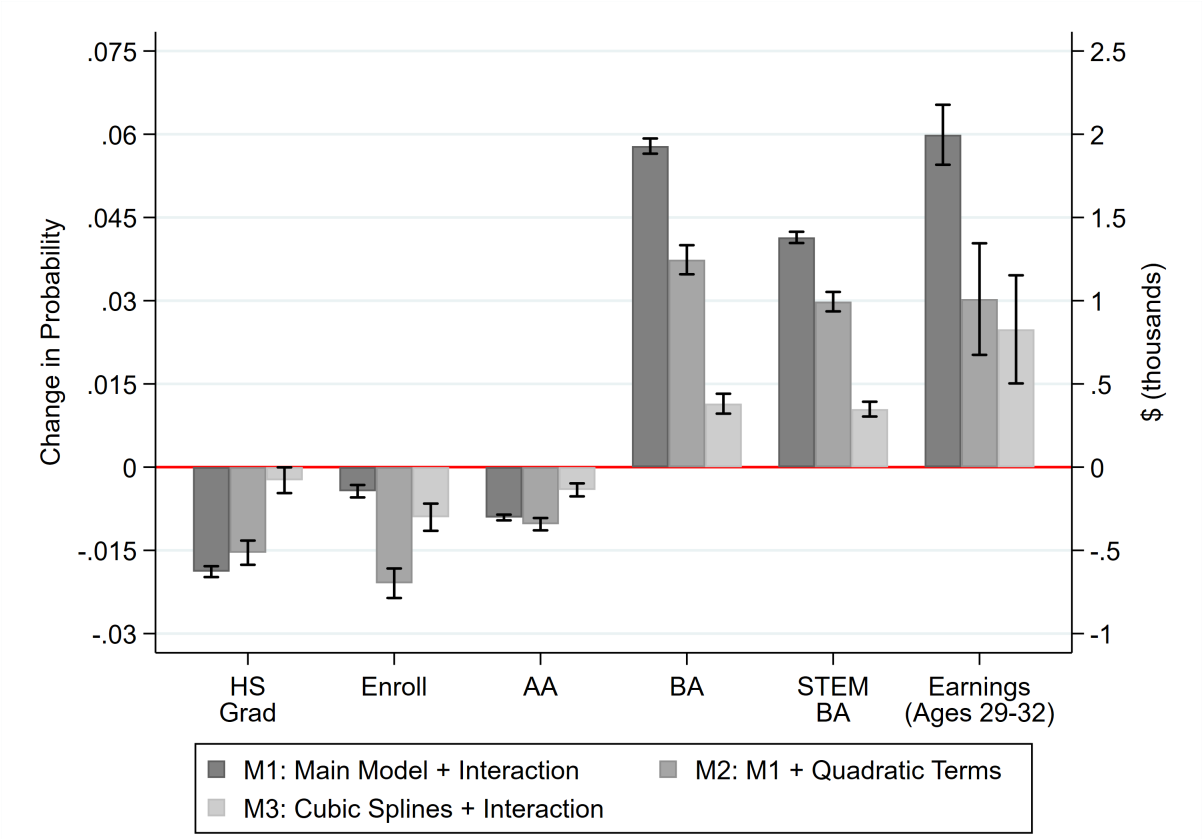
Notes: This figure presents line plots of mean outcomes across math test score bins for three selected ELA test score bins: -0.5 SD, 0 SD, and 0.5 SD. The figure is constructed using raw binned data from our EMS sample described in Section 1. Bins are constructed in 0.5-SD intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than two SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure A12: Relationship Between Math Scores and Outcomes Across ELA Score Levels (HS Sample)



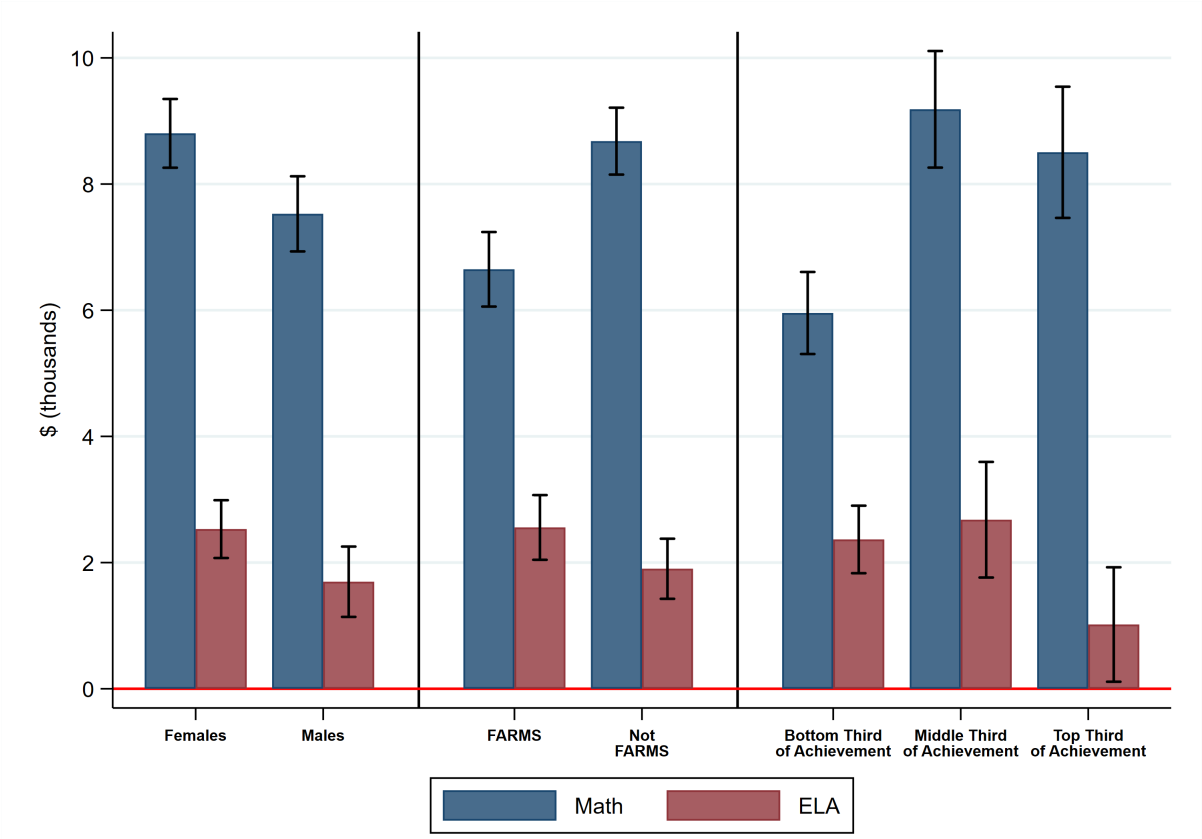
Notes: This figure presents line plots of mean outcomes across math test score bins for three selected ELA test score bins: -0.5 SD, 0 SD, and 0.5 SD. The figure is constructed using raw binned data from our HS sample described in Section 1. Bins are constructed in 0.5-SD intervals for math and ELA test scores, and mean outcomes are calculated for each combination of math and ELA score bins. The sample is restricted to observations with both math and ELA scores within two SDs of the mean. We further restrict the sample to observations for which the absolute difference between math and ELA scores is no more than 1.5 SDs, which ensures that all displayed bins satisfy the minimum cell-size requirement. All outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars.

Figure A13: The Relationship Between the Interaction of Math and ELA Scores and Outcomes (HS Sample)



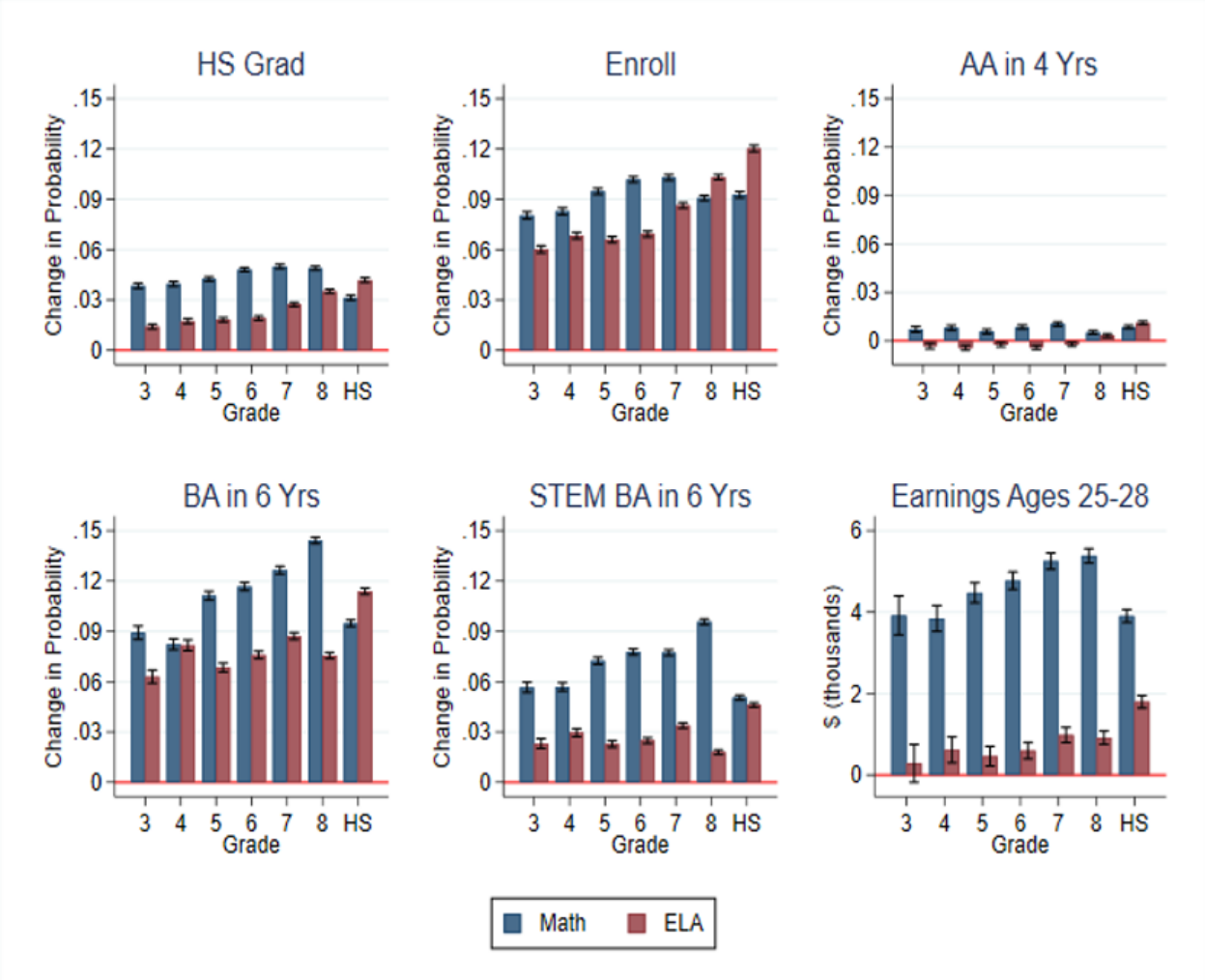
Notes: This figure reports coefficients and 95 percent confidence intervals for the interaction between math and ELA test scores from regressions of student outcomes on math and ELA test scores using the HS sample. The first model (M1) includes math and ELA scores and their interaction. The second model (M2) additionally includes quadratic terms in math and ELA scores to account for non-linear relationships. The third model (M3) uses cubic splines in math and ELA scores to flexibly account for non-linear relationships while retaining the interaction term. The sample and outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. All test measures are standardized, and all regressions include the full set of demographic controls used in the main specification.

Figure A14: Heterogeneity Estimates of the Relationship Between Test Scores and Earnings (EMS Sample)



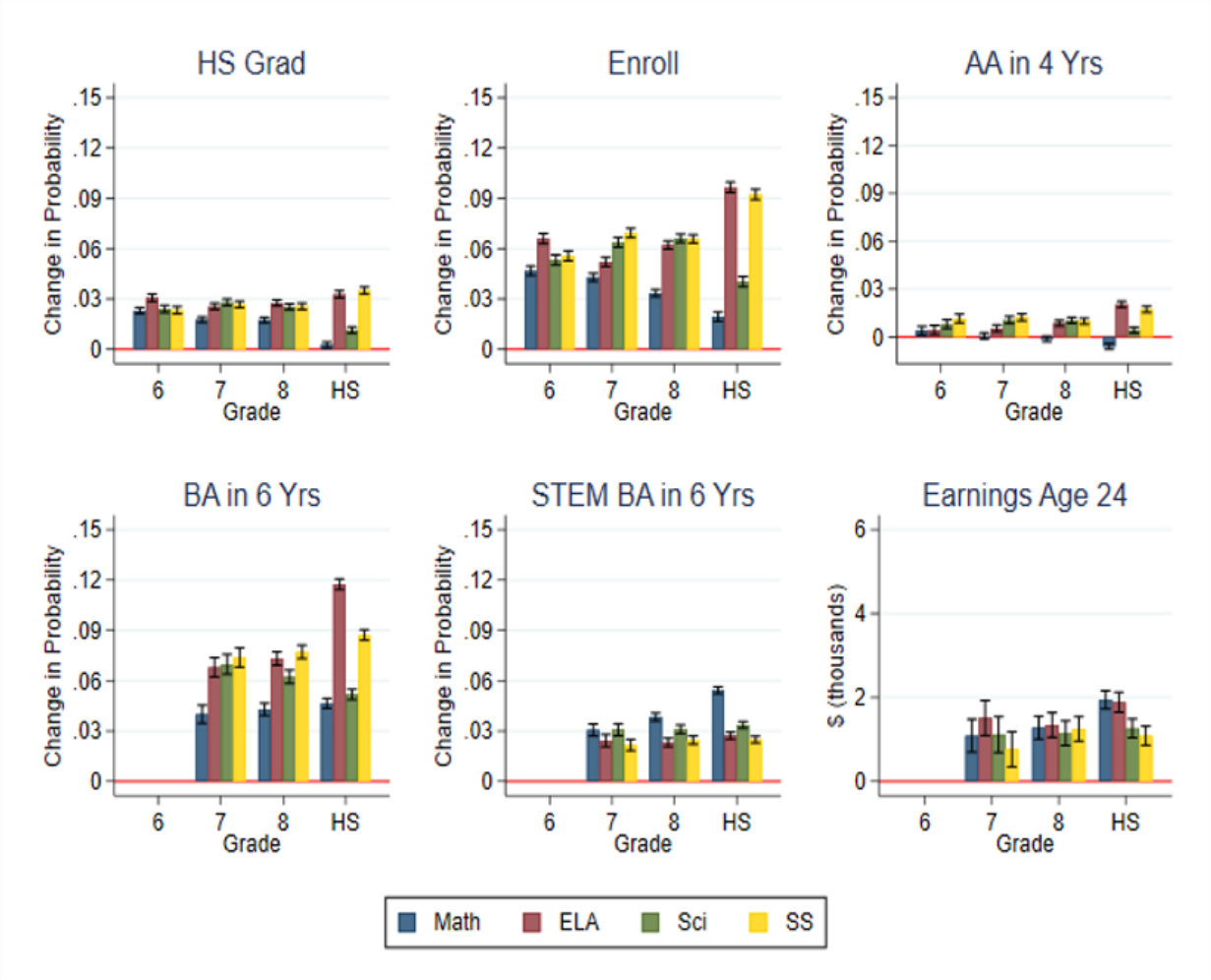
Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions of earnings on subject achievement measures across subgroups in our EMS sample. This plot shows a subset of our coefficients from Table 6. Estimates by gender are shown in columns (1)-(2), by eligibility for Free and Reduced-price Meals (FARMS) in columns (3)-(4), and by achievement terciles in columns (5)-(7). Achievement terciles are defined based on the average of students' math and ELA test scores, ranging from the lowest achievement tercile (Q1) to the highest achievement tercile (Q3). The sample and outcomes are described in Section 1. Average annual earnings are conditional on positive observed earnings and are reported in thousands of 2024 dollars. The specification is our main multivariate test score model with demographic controls in Section 2.1, Equation 1.

Figure A15: Relationship Between Test Scores and Outcomes by Grade



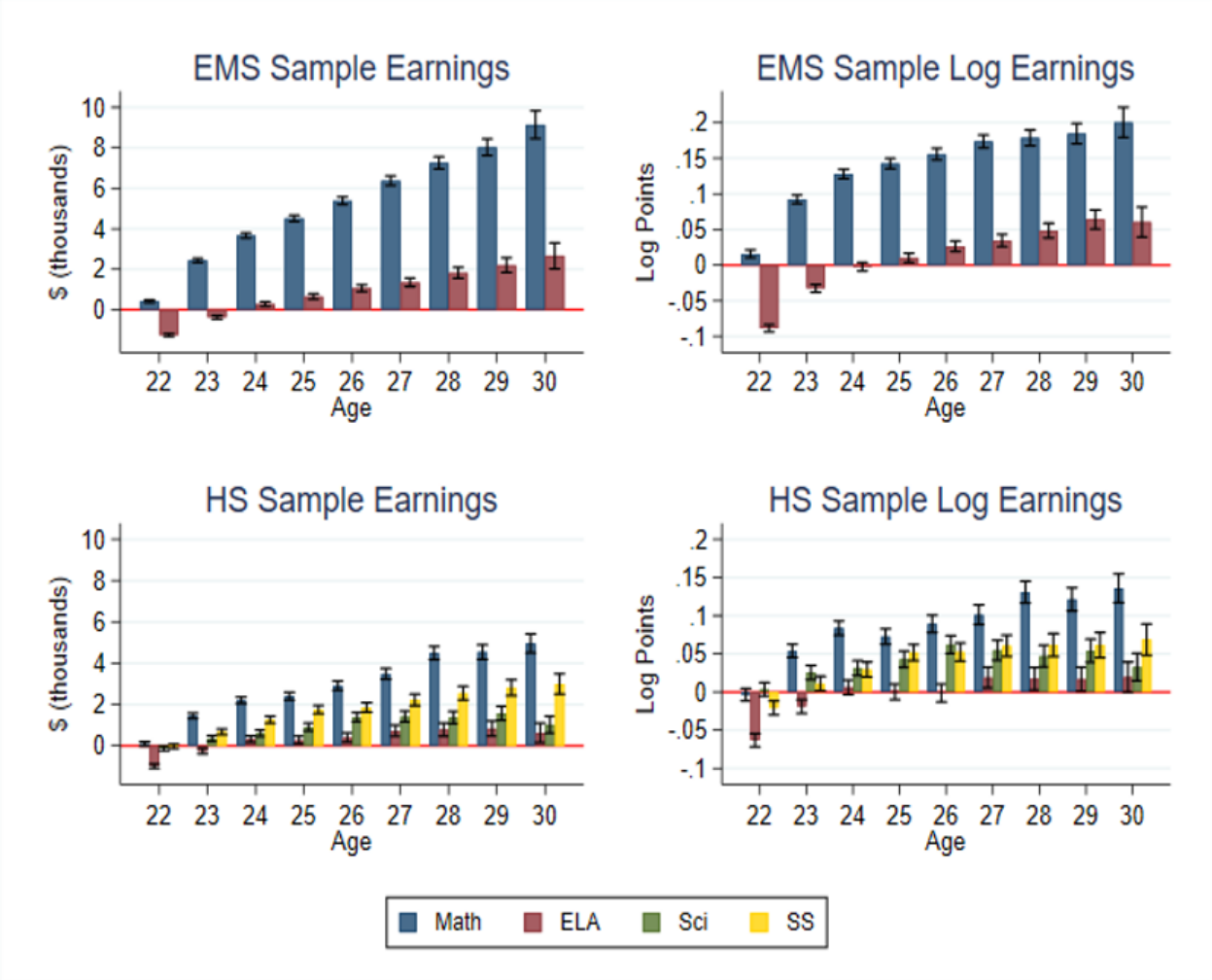
Notes: This figure reports coefficient estimates and 95 percent confidence intervals for the multivariate relationship between test scores and long-term outcomes by grade level. We use the EMS sample to construct estimates for grades 3-8 and the HS sample to construct estimates for HS. The samples are described in Section 1. 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 2.1, Equation 1, with only standardized math and ELA test scores as the main explanatory variables. Average annual earnings from ages 25-28 are conditional on having positive observed earnings and measured in thousands of 2024 dollars.

Figure A16: Relationship Between GPAs and Outcomes by Grade



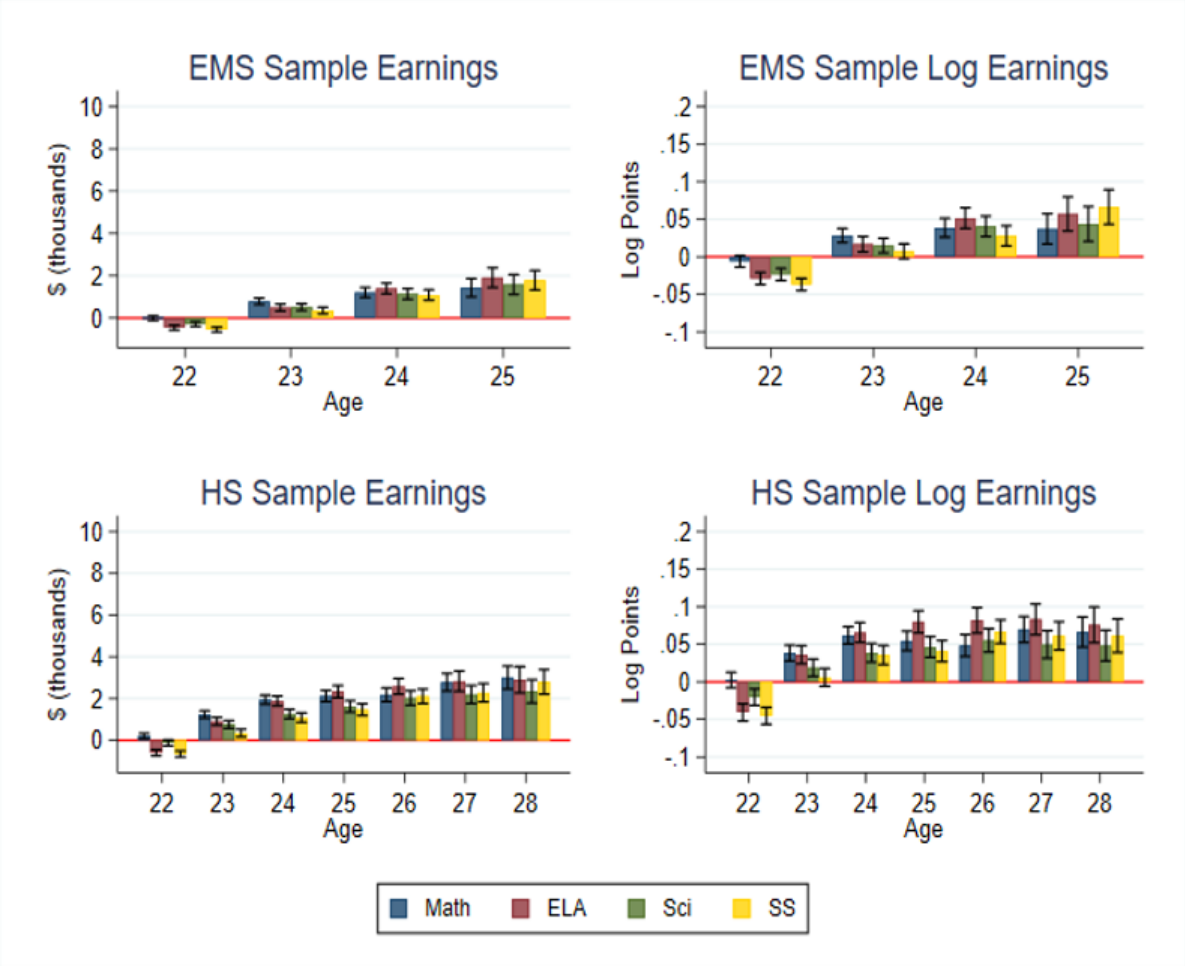
Notes: This figure reports coefficient estimates and 95 percent confidence intervals for the multivariate relationship between GPAs and long-term outcomes by grade level. We use the EMS sample to construct estimates for grades 6-8 and the HS sample to construct estimates for HS. The samples are described in Section 1. 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. Annual earnings at age 24 are conditional on having positive observed earnings and measured in thousands of 2024 dollars.

Figure A17: Relationship Between Test Scores and Earnings by Age



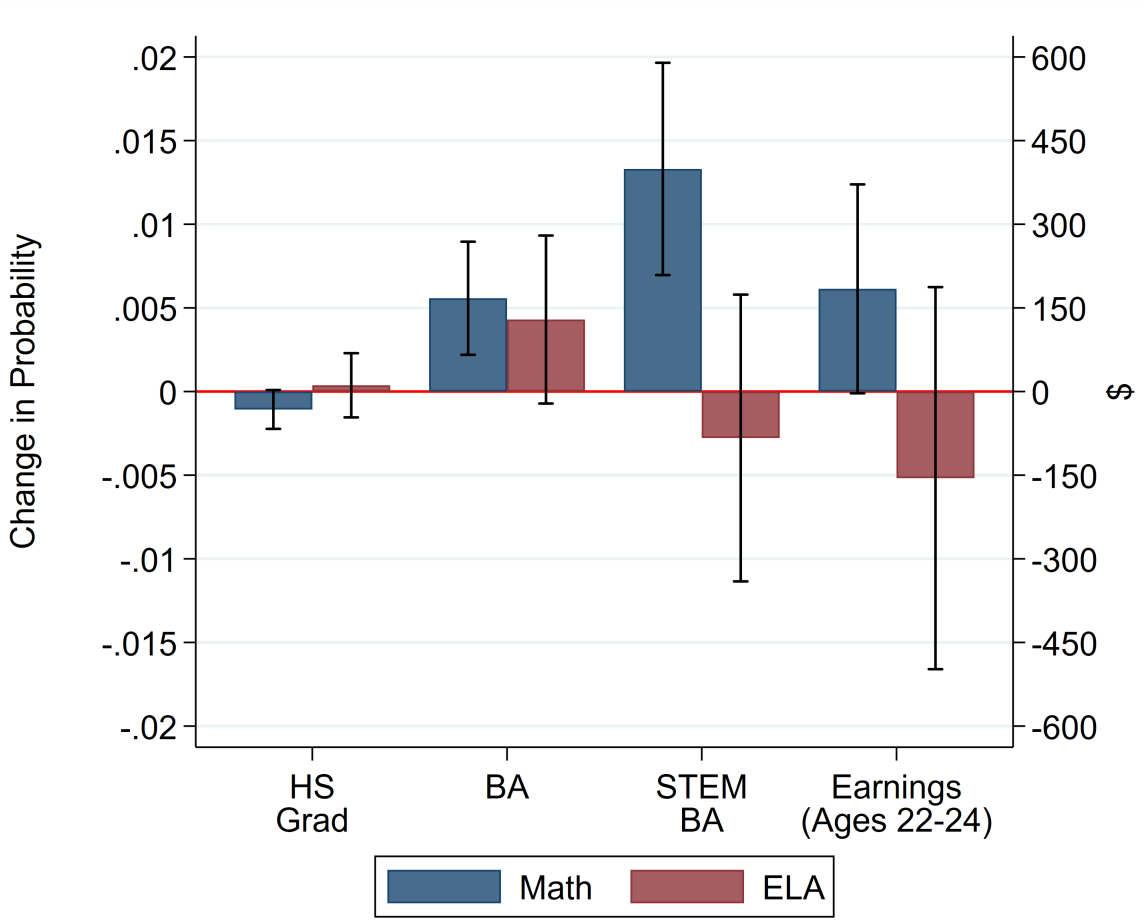
Notes: This figure reports coefficient estimates and 95 percent confidence intervals for the multivariate relationship between test scores and earnings by the age when earnings are measured. Estimates for the EMS sample are shown in the top two sub-figures and estimates for the HS sample are shown in the bottom two sub-figures. The samples are described in Section 1. 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 2.1, Equation 1.

Figure A18: Relationship Between GPAs and Earnings by Age



Notes: This figure reports coefficient estimates and 95 percent confidence intervals 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 1. 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.

Figure A19: Effect of a One SD Increaese in Teacher Test Score Value-Added on Educational Attainment and Earnings (EMS Sample)



Notes: This figure reports coefficient estimates and 95 percent confidence intervals from multivariate regressions estimating the effect of teacher value-added (TVA) for students' test scores on educational attainment and earnings. This plot shows a subset of our coefficients from Table 7, Panel B. The sample is our student-year level TVA sample described in Section 1, restricted to observations with a non-missing leave-out TVA measure. This figure reports the effect of a one SD increase in test score TVA on outcomes. The specification is our main TVA model with TVA controls in Section 2.5, Equation 7. 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, and SPED; 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. Average annual earnings from ages 22-24 are conditional on having positive observed earnings and measured in 2024 dollars. Robust standard errors used to compute confidence intervals are clustered at the school level.

Appendix B: Maryland Assessment 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.[15] The HSA tests continued to be administered through the mid- to late-2010s 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 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 Assess-

[15] Data collection from HSA exams 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 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 Structure, Functions, and Principles) (MSDE, 2012).

ments (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 were first administered to students in grades 5 and 8 in 2007.[16] 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 known as the Common Core State Standards (CCSS) Initiative began with the aim of strengthening the rigor and consistency of state academic standards 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 Assessment 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).[17]

Continuing the trend of implementing more rigorous standards and tests, the Maryland State Board of Education also adopted the Next Generation Science Standards (NGSS) in 2013. Similar to CCSS, these standards 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.

[16] Data collection from MSA tests 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 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).

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

Appendix C: Implementation of Teacher Value-Added Methods

Estimating Test Score TVA. Our TVA strategy proceeds in three steps. First, we construct subject-specific TVA measures following Chetty et al. (2014a). Second, we estimate how these TVA measures affect educational attainment and earnings. Third, we report forecast bias diagnostics to assess the validity of our research design.

We construct subject-level TVA in two steps: i) residualize student test scores on a rich set of controls and ii) compute leave-year-out predictions of teacher effects using residuals from surrounding years. First, we residualize test score S_{ijt}^k by regressing it on a vector of controls X_{ijt} for student i taught by teacher j in academic year t for subject k. The control vector includes flexible functions of 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 demographic controls used in our descriptive test score model; and iv) class- and grade-level means of those demographic variables. We interact each of these covariates with grade fixed effects, and we further include a control for class size.[18] We compute residualized test scores ν_{ijt}^k via following equation:

$$\nu_{ijt}^k = S_{ijt}^k - \hat{\Gamma} X_{ijt}. \tag{3}$$

This residualization partials out observed determinants of S_{ijt}^k, including prior achievement, demographics, grade level, and classroom composition. Next, we compute the mean residual across students by year for each teacher j:

$$\bar{\nu}_{jt}^k = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \nu_{ijt}^k. \tag{4}$$

where N_{jt} is the number of students taught by teacher j in year t. We then predict test score TVA in year t with mean teacher test score residuals from surrounding years:

$$\hat{\nu}_{jt}^k = \sum_{s=t-a}^{t+a} \hat{\psi}_s \bar{\nu}_{js}^k \mathbb{1}[s \neq t]. \tag{5}$$

where $a = 6$ in our setting. Using data from surrounding years reduces exposure to transitory unobservables and shocks in year t that could affect both the mean teacher test score residual $\bar{\nu}_{jt}^k$ and test scores in year t. The weights $\hat{\psi}_s$, which vary by the number of years before or after t, are chosen to minimize the mean-squared prediction error of $\bar{\nu}_{jt}^k$ using residuals from surrounding years taught by the same teacher:

[18] 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.

$$\Psi = \underset{\{\psi_{t-a},\ldots,\psi_{t+a}\}}{\arg\min} \sum_{j=1}^{J} (\bar{\nu}_{jt}^{k} - \sum_{s=t-a}^{t+a} \psi_s \bar{\nu}_{js}^{k} \mathbb{1}[s \neq t])^2 \tag{6}$$

This procedure yields leave-year-out jackknife TVA predictions which allow teacher quality to drift over time and incorporate Bayesian shrinkage toward the mean (Chetty et al., 2014a).

Relating TVA to Long-Run Outcomes. We use the leave-year-out TVA measures $\hat{\nu}_{jt}^{k}$ to estimate how assignment to a higher-TVA teacher relates to long-run educational attainment and earnings. Our main specification is a multivariate TVA model which includes TVA from multiple subjects as follows:

$$Y_i = \alpha + \sum_{k=1}^{K} \delta^k \hat{\nu}_{jt}^{k} + \Gamma X_{ijt} + \eta_{ijt}, \tag{7}$$

where Y_i is a long-run outcome for student i, $\hat{\nu}_{jt}^{k}$ is our TVA measure for teacher j in academic year t and subject k, X_{ijt} is the same vector of controls from Equation (3),[19] and η_{ijt} is the error term. We cluster standard errors at the school level. δ^k is the coefficient of interest and represents how outcomes vary with a one standard deviation increase in TVA in subject k, holding fixed our control vector as well as TVA in other subjects. In practice, our main multivariate specification includes just two TVA measures: math TVA $\hat{\nu}_{jt}^{m}$ and ELA TVA $\hat{\nu}_{jt}^{e}$. We also report univariate models with $K = 1$.

Identification requires that, conditional on X_{ijt}, students are not systematically sorted to teachers on unobserved determinants of Y_i. The primary threat is residual non-random assignment (e.g., sorting based on unobserved ability, motivation, or family resources) that remains correlated with $\hat{\nu}_{jt}^{k}$ even after conditioning on prior achievement and classroom composition. Prior research suggests 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 using forecast bias as a diagnostic.

Estimating Forecast Bias. We follow Chetty et al. (2014a) and estimate the relationship between residualized test scores ν_{ijt}^{k} and estimated 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.

[19] 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) student demographics (gender, race, gender-race interactions, FARMS, ELL, and SPED); 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.