The Content of Grades: Inflation, Information, and Inequality in High School Performance Measures*

Arnav Dharmagadda† Sarah Turner‡

Version 4.5b — September 6, 2026

Preliminary: Please do not circulate or cite without permission

Abstract

Students' high school grades and test scores serve as signals of skill for students, parents, post-secondary admissions and placement, and employers. Grade inflation, defined as differences in grades at a given level of test performance, appears both over time and across demographic groups. We use administrative data from the Virginia Longitudinal Data System (2014–2025) to document the magnitude, structure, and information consequences of the changing relationship between high school grades and test scores. The temporal component is substantial: conditional on test performance, the probability of earning an A rose 3–11 percentage points across subjects over the pre-COVID window, with further increases thereafter. The bulk of the rise in grades is attributable to changes in grading standards rather than shifts in student composition or test performance. To understand the changing correspondence between grades and test scores, we estimate models that recover student-specific differences in grade performance or "grade skill" which encompass skills and behaviors affecting grades distinct from the content knowledge measured by test-based assessments. This component predicts grades in untested courses, indicating that it is not idiosyncratic noise.

*For helpful comments and discussions, we thank Jeff Denning, Josh Goodman, Andrew Simon, and participants at SOLE.

†University of Virginia. qbd8dd@virginia.edu

‡University of Virginia and NBER. sturner@virginia.edu

1 Introduction

High school grades and test scores serve as key indicators of the student preparation. These indicators and the skills they measure impact credential attainment and graduation, postsecondary selection, performance in college, and labor market outcomes. For students, these measures guide decisions about course selection, college applications, and career paths. For colleges and employers, they provide signals to identify well-matched candidates from different backgrounds and school environments. Over the last two decades, grades have risen markedly while standardized test scores have plateaued and, more recently, declined. This divergence raises questions about how high school performance measures convey information about students' preparation for college and the workforce.

Information provided by grades and test scores may contain noise or be difficult for stakeholders to interpret, and measurement itself may differ across schools and change over time. Course grades and standardized tests do not provide identical measures of skills or class performance: grades capture both academic mastery and a broader set of behaviors and dispositions (Borghans et al., 2016; Heckman and Kautz, 2012). Standardized tests, by contrast, aim to provide a common yardstick for comparing content knowledge across students and schools, though they are also incomplete measures of the skills that predict long-run success (Jackson, 2018). These measures need not agree: a student may demonstrate strong content mastery on a standardized exam while receiving a lower course grade. Grades may incorporate broader (or different) measures of content mastery, they may reflect other types of behavioral skills such as including effort, participation, and homework completion or grades may reflect subjective elements not captured by the exam. Understanding the relationship between grades and test scores, the extent to which they provide similar or different measures of skill, and how this alignment shapes the reliability of high school performance indicators motivates this research.

Rising high school grades and flat scores on test-based assessments are found nationally as well as in the state of Virginia. According to the NAEP High School Transcript Study, average high school GPA increased from 2.68 in 1990 to 3.11 in 2019, while 12th-grade NAEP mathematics and reading scores showed no corresponding gains (U.S. Department of Education, National

Center for Education Statistics, 2022; Sanchez and Moore, 2022). Researchers from the College Board found similar results in their examination of grades relative to SAT scores (Hurwitz and Lee, 2018), and ACT research documents analogous patterns, showing further that grade inflation reduces the predictive validity of high school GPA for first-year college performance (Sanchez, 2025). This divergence between grades and test performance raises questions about what each measure captures, how their relationship has changed over time, and whether the "signal" about student preparation and skills in high school grades has changed.

This paper uses administrative data from the Virginia Longitudinal Data System (VLDS) to examine how high school grades have changed over time in relation to test scores, the implications for different demographic groups, and the consequences for understanding the information content of grades. Our empirical strategy makes use of a key feature of Virginia's education system: mandatory, criterion-referenced Standards of Learning (SOL) end-of-course examinations that are linked to course-level grades in the administrative record. Because SOL tests measure mastery of specific content standards, are required for essentially all students in tested courses, and are not used in college admissions decisions, they provide a benchmark against which to assess what course grades capture beyond content knowledge. Many students also take exams that carry external currency in college admissions such as the SAT and end-of-course Advanced Placement (AP) exams, which are aligned with college performance.

We document three sets of findings. First, we show substantial grade inflation over the 2014–2025 period that is concentrated at the top of the grade distribution: the share of students receiving A grades in SOL-linked and AP courses has risen sharply while associated test scores (SOL and AP) have remained flat or declined. Oaxaca–Blinder decompositions attribute the bulk of the rise to changes in grading standards rather than shifts in student composition. The grade–test-score relationship also varies substantially by race, gender, and economic status: women tend to receive higher grades than men conditional on test scores, while Black students and economically disadvantaged students receive lower grades at the high end of the grade distribution. Second, exploiting the panel structure of student transcripts, we estimate student fixed effects across tested courses to identify a stable behavioral component of grades—what we term "grade skill," the student-specific propensity to earn higher grades than test performance alone would predict

(Section 5). Net of estimation noise, the within-school variance of grade skill grew 42 percent between the 2014 and 2024 cohorts. This component predicts grades in non-tested courses and on externally assessed AP examinations, confirming that it captures a real, transferable student trait whose weight in the grading function has expanded. These findings speak directly to contemporary debates over test-optional admissions, grading reform, and the reliability of high school transcripts as indicators of student preparation.

The remainder of the paper proceeds as follows. Section 2 provides research context on the role of grades and tests and defines grade inflation. Section 3 describes the Virginia education system, data, and descriptive evidence. Section 4 measures grade inflation using regression analysis and decomposition methods. Section 5 decomposes the content of grade inflation using student fixed effects estimated across courses to isolate a stable behavioral component of grades—"grade skill"—and traces its expansion across cohorts. Section 6 discusses implications and concludes.

2 Research Context

2.1 The Role of Grades and Test Scores

Questions about the information content of high school grades are not new. In 1913, educational psychologist Guy Montrose Whipple described the marking system as "an absolutely uncalibrated instrument" (Whipple, 1913). More than a century later, the critique remains apt. There is no universal standard governing what a given grade means, and there is considerable variation across districts and among teachers within schools in the relative weight given to test performance, homework completion, and behavioral factors such as attendance and class participation in determining course grades (Borghans et al., 2016; Franco and Eberhardt, 2021).

Measuring acquired knowledge is an objective of both grades and standardized tests, but they do so in fundamentally different ways. The aim of standardized achievement tests is to provide objective, comparable measures of content mastery across students and schools. Yet achievement tests are necessarily incomplete: they do not measure everything covered in a class, and concerns

about socio-demographic biases in performance are well documented (Heckman and Kautz, 2012). Course grades, by contrast, are broader measures that reflect not only mastery of subject content but also engagement with the learning process including class participation, homework completion, and other demonstrations of effort and consistency. Borghans et al. (2016) provide direct evidence that grades and achievement tests measure different constructs: while both are correlated with cognitive ability, grades are substantially more sensitive to personality traits, particularly conscientiousness. Heckman and Kautz (2012) show more broadly that achievement tests miss important "soft skills" such as personality traits, goals, motivations, and preferences that are valued in the labor market and in many other domains.

To the extent that grades capture productive non-cognitive skills alongside content knowledge, they provide richer information than test scores alone. But because grades bundle these dimensions into a single measure, stakeholders cannot easily distinguish whether a high grade reflects deep content mastery, strong behavioral engagement, lenient grading standards, or some combination. Grades also carry a substantial subjective dimension that test-based measures do not: teacher discretion over how to weight effort, participation, and conduct gives them an element of judgment that a standardized exam does not allow.

2.2 Grade Inflation: Definitions and Evidence

The term "grade inflation" is ubiquitous, yet its meaning is often unclear. Tyner and Gershenson (2020) provide a useful typology distinguishing three forms. *Static grade inflation* refers to a cross-sectional disconnect between grades and measured achievement; for example, when students in some schools receive systematically higher grades than students in other schools with the same test scores. *Dynamic grade inflation* refers to grades rising over time without corresponding gains in achievement. *Differential grade inflation* occurs when the mapping between grades and achievement varies across student groups or school types, and when these differentials change over time. These distinctions matter because the consequences of each type are different: static inflation may reflect stable differences in grading norms, while dynamic inflation actively degrades the signal value of grades, and differential inflation can exacerbate or mask inequality.

The national evidence for dynamic grade inflation is substantial. Gershenson (2018) documents that average high school GPAs rose from 2.68 to 3.11 between 1990 and 2016, even as NAEP scores plateaued and later declined. Sanchez and Moore (2022) extend this evidence using ACT-tested students, finding that the share of students with an A average increased from 39% in 2010 to 53% by 2021 while ACT scores declined.

Measured achievement did more than plateau. Wyckoff (2026) documents that NAEP gains accumulated through about 2013 and then reversed: eighth-grade mathematics scores at the tenth percentile fell by 17 percent of a student-level standard deviation between 2013 and 2019, before any pandemic disruption, and simple extrapolation attributes roughly 40 percent of the post-2020 learning loss to influences already in place. Grades and measured achievement have therefore moved in opposite directions over the period in which grades rose fastest.

The pattern is consistent across data sources: grades have risen markedly at the top of the distribution while standardized measures of achievement have not.

What explains the rise in grades? Several forces likely contribute, though disentangling their relative importance remains an open empirical question. One set of factors involves deliberate changes to grading policies. In recent decades, there has been a shift toward "standards-based grading," with evaluations tied to mastery of specific learning objectives rather than averaging scores across assignments (Guskey and Bailey, 2015). In this approach, students have multiple opportunities to demonstrate proficiency, including reassessment and the elimination of zeros, which proponents argue better reflect actual learning (Feldman, 2019). Concurrently, the "grading for equity" movement has advocated for practices that focus on measuring content mastery while eliminating penalties for behaviors such as late work, attendance, or homework completion (Feldman, 2019; Romano, 2022). Critics raise concerns about the limited independent evidence supporting these claims (Coffey, 2023; Hess, 2024). Some associated practices such as minimum grading floors, unlimited retakes may contribute to grade inflation. In Virginia, these reforms have been adopted unevenly: Fairfax County introduced a 50% grade floor in 2021 and a 70-30 summative-formative split with retakes in 2022; Albemarle County adopted an equitable grading policy in 2020 that prevented homework, extra credit, and practice from affecting final grades;

Loudoun County implemented a 50% floor and retake policy in 2019 (Franco and Eberhardt, 2021).[1] Beyond formal policy changes, increasing competitive pressure in college admissions may also create incentives for grade leniency at both the teacher and institutional level. The accountability environment also shifted. The Every Student Succeeds Act, enacted in 2015, retained federal testing and reporting requirements while devolving decisions about consequences for low-performing schools to the states, and the subsequent retreat from consequential accountability in many states is one candidate explanation for the achievement decline (Wyckoff, 2026). Whether weaker accountability also changed how schools grade, as distinct from how students learn, has not been established.

The COVID-19 pandemic disrupted secondary education, suspending assessments and reshaping grading policies. Many school districts implemented emergency policies such as minimum grading floors, "no-harm" grading that prevented student grades from declining, and pass/fail options (Arundel, 2020; Goldhaber and Young, 2023). While intended as temporary measures, many of these policies have persisted beyond the emergency period. Goldhaber and Young (2023), using Washington State administrative data, show that the relationship between grades and test scores weakened substantially during the pandemic and that elevated grading persisted even after the return to in-person instruction. This pattern is not symmetric: grade inflation is concentrated at the top of the distribution, with A grades rising sharply while D and F rates remain relatively stable. This is a pattern Tyner and Gershenson (2020) describe as "top-only" inflation.

Two recent papers use teacher-level variation in grading standards to identify the consequences of grading leniency on students. Denning et al. (2026) construct and validate two distinct teacher-level measures of grade inflation using high school administrative data from the Los Angeles Unified School District and from Maryland—"mean grade inflation," which captures the extent to which a teacher assigns grades higher than students' standardized test scores would predict, and "passing grade inflation," which captures a teacher's propensity to lift the marginal student from a failing to a passing grade. The two measures are correlated but distinct, and have systematically different consequences. Higher mean grade inflation reduces students' next-year test scores,

[1] A systematic analysis of the 132 Virginia school divisions' grading policies and their effects on grade distributions is the subject of ongoing work.

high school graduation, SAT-taking, postsecondary enrollment, and earnings up to six years after expected graduation. On the other hand, higher passing grade inflation lowers grade repetition and raises five-year high school graduation rates. Gershenson et al. (2024), using North Carolina end-of-course data linked to Algebra I teachers, find that exposure to high teacher grading standards raises contemporaneous Algebra I achievement and performance in subsequent math courses. These analyses build on Figlio and Lucas (2004) and Betts and Grogger (2003), who document positive effects of high grading standards on achievement and attainment, from within-school variation across teachers and the measurement of those effects on downstream student outcomes.

It is worth noting that grade inflation does not necessarily imply a loss of information to the market. A theoretical literature on strategic grade disclosure suggests conditions under which coarser or inflated grading can be informationally neutral or even beneficial. Chan et al. (2007) model grade inflation as an equilibrium phenomenon in which schools strategically adjust standards in competition for student placements; employers can partially offset inflation by conditioning on school reputation and grade distributions. Harbaugh and Rasmusen (2018) show that coarse grading schemes such as pass/fail or honors/pass/fail can generate more information than exact scores when they induce broader participation in certification. Ostrovsky and Schwarz (2010) demonstrate that the equilibrium level of information disclosure in matching markets depends on market structure, not simply on the precision of individual signals. Whether grade inflation in practice degrades the information available to colleges and employers is therefore an empirical question.

2.3 Grades, Test Scores, and Predicting Postsecondary Success

The introduction of standardized tests to assess college preparation in the first part of the twentieth century—the SAT in 1926 and the ACT in 1959—offered instruments to compare students from heterogeneous high schools on a common scale. Not only did these instruments predict performance in the first year of college, but the innovation of standardized measures increased geographic integration in the higher education market (Hoxby, 2009). Of course, colleges have

also relied on high school grades and transcripts to measure achievement and preparation. Some analysts suggest that these measures convey information about skills such as persistence, self-control, and sustained classroom engagement that predict college success and are not well captured by a single standardized exam. Others have raised concerns that standardized tests may disadvantage low-income and minority applicants relative to grades (University of California Academic Senate, 2020).

The relative predictive power of grades and tests likely depends on the institutional context. Beyond content knowledge, grades may reflect skills like perseverance and sustained engagement that plausibly predict completion and success in college coursework. Studying large public universities where students vary widely in academic preparation, Bowen et al. (2009) find that high school GPA is a stronger predictor of college completion than SAT scores. In contrast, Friedman et al. (2025) study academic outcomes at Ivy Plus institutions and find that SAT and ACT scores predict college grades roughly four times more strongly than high school GPA, conditional on demographics and family background. The discrepancy likely reflects selectivity: at highly selective institutions, matriculants exhibit relatively little variation in either high school grades or college completion rates, leaving relatively more predictive work for standardized tests. Kurlaender and Dykeman (2019) document a parallel pattern in California's public university systems—the SAT and the state-mandated SBAC accountability exam each explain more variance in college outcomes at the more selective University of California campuses than at the California State University campuses, where high school grades carry comparatively more weight. Two patterns emerge. First, grades and test scores capture overlapping but distinct dimensions of preparation. Second, the marginal predictive value of each measure depends on the variance of the other in the relevant applicant pool.

The COVID-19 pandemic sharply increased the emphasis on high school transcripts in college admissions.[2] Grade inflation has direct consequences for the predictive content of those tran-

[2] McManus et al. (2023) show that applicants with scores below an institution's median tend to withhold them. Sacerdote et al. (2025) document that high-achieving applicants from disadvantaged backgrounds are often made worse off by test-optional policies, because they would have benefited from submitting strong scores that conveyed preparation beyond what their high school grades alone would suggest. By 2025, Harvard, Yale, Dartmouth, Brown, Cornell, Penn, MIT, Stanford, and Caltech had reinstated testing requirements, while several public universities—including the University of Virginia and the College of William & Mary—and a small number of selective privates retain test-optional policies.

scripts. As the share of students receiving A's grows, the GPA distribution compresses at the top, potentially eroding the ability of grades to distinguish among applicants (Pattison et al., 2013). Sanchez (2025), using ACT data linked to college outcomes, shows that school-level grade inflation reduces the predictive validity of high school GPA for first-year college performance: among students with a high school GPA of 2.5, those at high-inflation schools earn predicted college GPAs nearly 0.4 points lower than peers at low-inflation schools—a gap that ACT scores, unaffected by local grading norms, do not exhibit.

3 Measuring High School Performance in Virginia

3.1 Data and Institutional Context

Virginia's public school system enrolls approximately 1.3 million students across 132 school divisions and roughly 330 public high schools, graduating over 100,000 students annually.[3] The student population is demographically diverse and has shifted substantially over the study period: between the 2014 and 2025 graduation cohorts, the Hispanic share of graduates rose from 11% to 20%, the White share fell from 55% to 45%, and the share classified as economically disadvantaged grew from 45% to 55% (Table 1). Approximately 89 percent of Virginia public school students earn an on-time high school diploma (Virginia Department of Education, 2025), and 64 percent of graduates enroll in a postsecondary institution within 16 months of graduation—34 percent at a four-year public institution, 10 percent at a four-year private institution, and 21 percent at a two-year institution (Virginia Department of Education, 2024).

This study uses administrative records from the Virginia Longitudinal Data System (VLDS), a federated data system that links individual-level records across 13 state agencies, including the Virginia Department of Education and the State Council of Higher Education for Virginia (SCHEV).[4] For this paper, we draw on VLDS K–12 records maintained by the Virginia Depart-

[3] School divisions in Virginia are the equivalent of school districts in other states.

[4] VLDS was developed under the federal Statewide Longitudinal Data Systems (SLDS) grant program and uses a double-deidentification architecture: participating agencies retain personally identifiable information within their own systems while contributing hashed records that can be linked across agencies without exposing individual identities. See https://vlds.virginia.gov for details on governance and participating agencies.

ment of Education: course-level grades, end-of-course standardized test scores, SAT and ACT scores, attendance records, and student demographics for the universe of Virginia public school students.

Our analytic cohorts encompass Virginia high school graduates from the 2014–2025 graduation cohorts. We assign students to graduation cohorts based on expected graduation year; many of the courses and assessments we study—Algebra I, Biology, Earth Science—are typically taken in the ninth or tenth grade, well before graduation. For SOL-linked courses, we observe course grades matched to end-of-course examinations, and we focus our analysis on seven subjects: Algebra I, Algebra II, Geometry, Biology, Chemistry, Earth Science, and English Reading. For AP courses, we observe course grades matched to AP examination scores, and we focus on six subjects: Calculus AB, Biology, English Language, English Literature, U.S. Government, and U.S. History.

We observe letter grades (A through F) for each course enrollment and construct three grade outcome measures: an indicator for receiving an A, an indicator for receiving a B or above, and the numeric GPA equivalent (A=4, B=3, C=2, D=1, F=0).[5] Annual absence rates capture days absent as a share of enrolled days; we use absences both as a control variable and as a mediator in understanding what grades capture beyond test performance. Student-level demographic indicators include race/ethnicity (Black, Hispanic, Asian, White, and other), gender, and economic disadvantage status (qualifying for free or reduced-price lunch). Table 1 summarizes these characteristics for the early (2014–2015), middle (2018–2019), and late (2024–2025) graduation cohorts.

3.2 Testing Landscape in Virginia

Virginia adopted the Standards of Learning (SOL) for mathematics, English and language arts, science, and social studies in 1995—predating the federal No Child Left Behind Act by seven years. The first SOL examinations were administered in spring 1998 as part of a broader accountability system that tied school accreditation to student pass rates. At the high school level, SOL

[5]Virginia does not mandate a uniform grading scale (Virginia Administrative Code 8 VAC 20-160-50), but the vast majority of divisions use a standard 10-point scale.

assessments are end-of-course exams directly linked to the content of specific courses: a student completing Algebra I takes the Algebra I SOL, a student completing Biology takes the Biology SOL, and so forth. Scores range from 0 to 600, with 400 as the pass threshold and 500 indicating advanced proficiency. The SOL system also underpins Virginia's verified credit graduation requirements: students must earn a specified number of verified credits, typically obtained by passing both the course and the corresponding SOL exam.

Three features of the SOL system are central to our empirical strategy. First, SOL end-of-course exams were mandatory for all students enrolled in tested subjects through the 2020 graduation cohort, generating near-universal participation in courses like Algebra I (91% tested in 2014, averaging 90% over 2014–2019), Biology (91% in 2014, averaging 89%), and English Reading (87% in 2014, falling to 80% by 2019) (Appendix Table A1). Second, SOL scores are not used in college admissions decisions and are primarily a school accountability tool, limiting strategic incentives for students to adjust effort on one assessment in response to performance on the other.[6] Third, the criterion-referenced design provides scores with a stable interpretation across cohorts: a score of 450 in Algebra I reflects the same content mastery in 2014 as in 2024. Together, these features make SOL scores a benchmark against which to assess changes in grading standards.

The SOL testing regime changed substantially after 2021. Virginia reduced the number of required end-of-course assessments, and the resulting decline in test-taking varied sharply across subjects. Biology and Algebra I testing rates recovered to roughly 73–89% by the 2025 cohort, but Chemistry testing collapsed from over 60% to under 5%, and Algebra II fell from 69% to 15% (Appendix Table A1). This differential attrition is not random: among post-2021 cohorts, students who continue to take the SOL exam have higher course GPAs, lower absence rates, and are less likely to be from minority or economically disadvantaged backgrounds than their non-tested peers. We address this selection directly in our analysis and flag post-2021 results for subjects with severe test-taking declines. Appendix A provides further detail on SOL test versions and policy changes over the study period.

[6] SOL scores are typically available to students before they receive final course grades.

Virginia students also take several standardized assessments that carry external currency. Advanced Placement (AP) examinations are nationally administered, subject-specific tests scored on a 1–5 scale; students take them voluntarily after completing AP coursework. AP enrollment and test-taking rates vary across schools and student backgrounds but have been relatively stable over our study period, with 5–15% of students taking AP exams in any given subject. The SAT serves a different function as a college entrance examination. SAT and ACT testing was severely disrupted between spring 2020 and early 2021 when test centers closed due to the COVID-19 pandemic, and Virginia's flagship institutions—including the University of Virginia and the College of William & Mary—adopted test-optional admissions policies beginning with the fall 2021 matriculation cycle. SAT take-up rates have not fully recovered. We observe SAT scores for graduation cohorts 2017–2024, following the 2016 SAT redesign to the 400–1600 scale.

3.3 Descriptive Evidence

The central empirical regularity motivating this paper is visible in the raw data. Figure 1 compares the full grade distribution for each SOL-linked course across three graduation cohorts: 2015, 2018, and 2024. The dominant pattern is a shift of mass from B to A grades. In Biology, for example, the share of students in the top SOL score quintile receiving an A rose from 66% in 2015 to 78% in 2024; among students in the bottom quintile, the share receiving an A increased from 7% to 12%. The share of C grades declined modestly, and D/F rates remained largely stable. This "top-only" pattern—rising A rates without corresponding declines in failure rates—is consistent across all seven SOL subjects and the six AP subjects examined (Figure 2).

The upward shift in course grades is mirrored in overall academic performance measures. Mean overall HS GPA rose from 2.78 for the 2014 cohort to 2.92 by 2021–2023 before easing slightly to 2.90 by 2025 (Appendix Table D1). The share of students with a GPA of 3.5 or above grew from 22.8% to 30.6% over the 2014–2021 period. The 75th percentile rose from 3.43 to 3.64 by 2025, while the standard deviation is roughly flat through the pre-COVID period (0.84 in 2014, 0.81 in 2016–2019) before rising to 0.89 by 2025. The share of students at or above 4.0 grew from 4.4% to 6.5% by 2025. Figure 3 displays the kernel density of overall HSGPA for selected cohorts, illus-

trating both the rightward shift and the increasing mass at the 4.0 ceiling.

Figure 4 traces the share of students earning an A and the share earning B or above in each SOL-linked course over the full 2014–2025 window. The share earning an A rises steadily across cohorts, while the share earning B or better is more muted and the share with a D or F (not shown) remains flat. For example, in Algebra I, P(A) roughly doubled over the period, from 16.7% for the 2014 cohort to 33.8% by 2024. The stability of the bottom tail suggests that grade inflation in Virginia operates primarily through reclassification of B-level performance as A-level performance, rather than through broad-based leniency that lifts all grades. Figure 4 also presents these trends alongside SOL test scores, which show no corresponding trend. The same divergence holds in AP courses (not shown). This disconnect—rising grades against flat or declining test performance—is the motivating fact for the remainder of the paper.

4 Measuring Grade Inflation

The descriptive evidence in Section 3 shows rising grades alongside flat or declining test performance, but this evidence is only suggestive of grade inflation: other student characteristics may be changing in ways that affect the grade distribution. We address this question with a regression framework that progressively controls for student composition, attendance behavior, and school heterogeneity. The central question is whether the upward trend in grades reflects changes in who is taking courses and how prepared they are, or whether it reflects changes in grading standards themselves. We then use an Oaxaca–Blinder decomposition to summarize the relative contribution of each.

4.1 Regression Framework

We estimate a linear probability model for receipt of an A grade in each SOL-linked course:

$$\Pr(A_{icgt}) = \alpha + \gamma \cdot t + f(\text{SOL}_{icg}) + X_i'\beta + \varphi_s + \varepsilon_{icgt} \tag{1}$$

where A_{icgt} is an indicator for student i receiving an A in course c, taken at grade level g, in cohort t; t enters as a linear cohort trend; $f(\text{SOL}_{icg})$ is a set of 50-point SOL score bin dummies that allow a nonlinear mapping between test performance and grades; X_i is a vector of student characteristics including grade level when the course was taken (8th–11th), demographic indicators, and the student's annual absence rate in the year the course was taken; and φ_s denotes school fixed effects, which enter only in Specification 4. The coefficient of interest, $\hat{\gamma}$, measures the annual change in P(A) after conditioning on measured ability. We focus on receipt of an A because grades are ordinal and most of the empirical change occurs at the A/B margin (Section 3). For AP courses, we replace the SOL bin dummies with dummies for AP exam score (1–5) plus an indicator for not taking the exam.

We estimate four sequential specifications. Specification 1 includes only the score bin dummies. Specification 2 adds demographic indicators (female, Black, Hispanic, Asian, economically disadvantaged). Specification 3 adds the student's absence rate in the year of the course. Specification 4 adds school fixed effects, identifying $\hat{\gamma}$ from within-school variation. The progressive structure serves a diagnostic purpose—showing whether the upward trend in grades survives controls for compositional shifts, attendance changes, and differential sorting across schools. We also estimate models replacing the linear trend with year fixed effects to capture the dynamics of inflation, including any discontinuous shifts associated with the COVID-19 pandemic. Standard errors are clustered at the school level throughout.

The bin specification of test scores is motivated by the strongly nonlinear mapping from test performance to grades. Figure 5, Panel A plots the SOL score bin coefficients from the pre-COVID Specification 3 regression, with one line per subject; Panel B plots the analogous AP exam score category coefficients. Two features stand out. First, the score–grade gradient is steep and approximately linear above the SOL pass threshold (400): the 450–499 bin is 17–26 percentage points above the omitted 400–449 bin, the 500–549 bin is 41–56 pp above, and the 550+ bin is 50–72 pp above. Below 400, all bin coefficients cluster between -0.02 and -0.09—a floor effect in which sub-proficient SOL performance rarely yields an A regardless of subject. Second, the gradient has essentially the same shape across all seven SOL subjects. The AP gradient (Panel B) is even more tightly bundled: the score-5 premium relative to score-3 is 27–43 pp across all six AP sub-

jects, reflecting the standardization of AP scoring.

4.2 Trend Estimates

Table 2 reports $\hat{\gamma}$ from the linear-trend specification for the pre-COVID period (graduation cohorts 2014–2019), a window in which SOL testing was near-universal and unaffected by pandemic-era policy changes. The results reveal substantial and precisely estimated grade inflation in six of the seven subjects. In the preferred specification (Specification 4, with school fixed effects), the annual $\hat{\gamma}$ coefficients range from 0.7 percentage points per year in Geometry to 2.2 percentage points per year in Algebra I. Over the five-year pre-COVID window, these translate to cumulative increases of 3–11 percentage points in P(A). Algebra I shows the largest effect: an annual increase of 2.2 pp that, against a 2014 baseline P(A) of 16.7% (Table C1), implies an 11 pp gain—a 66% increase in the share of students earning an A. Biology ($\hat{\gamma} = 0.0186$, 9.3 pp over five years) and Earth Science ($\hat{\gamma} = 0.0184$, 9.2 pp) are similarly large. English Reading shows moderate inflation ($\hat{\gamma} = 0.0093$, 4.7 pp), while Chemistry ($\hat{\gamma} = 0.0075$, 3.8 pp) and Geometry ($\hat{\gamma} = 0.0068$, 3.4 pp) show smaller but precisely estimated trends.[7]

A key finding across all subjects is the stability of $\hat{\gamma}$ as controls are added. Moving from Specification 1 (SOL bins only) to Specification 4 (adding demographics, absences, and school fixed effects) changes the Algebra I coefficient from 0.0250 to 0.0219—a modest 12% reduction. The pattern is similar across subjects: compositional changes in the student population, shifts in attendance behavior, and differential sorting across schools explain little of the observed grade inflation. The dominant source of rising grades is changes in how a given level of measured performance is translated into a course grade. The demographic and absence coefficients from Specifications 3 and 4 are reported in Table 4 and form the basis for the cross-sectional analysis in Section 4.3.

When we replace the linear trend with year fixed effects (Figure 6, reference year 2018), the

[7]We exclude Algebra II from further exposition of changes over time. The subject experienced an SOL test-form transition during the first part of the pre-COVID period, and post-2021 SOL participation in Algebra II fell sharply—to 14.4% of enrollees by the 2024 cohort (15.1% by 2025)—creating severe selection that complicates over-time comparison.

cohort-level pattern reveals three distinct phases. From 2014 to 2019, P(A) drifts steadily upward in most subjects, consistent with the linear-trend estimates. The COVID-era cohorts (2020–2022) show a discrete jump of roughly 5–10 pp above the pre-existing trend, reflecting the emergency grading policies documented in Section 2. By 2025, P(A) has partially reverted but remains well above the pre-COVID trend line. The post-COVID trajectory suggests that while the most extreme pandemic-era leniency has receded, a permanent upward shift in grading standards persists. Appendix Table B1 reports the full-period regressions analogous to Table 2.

The dynamics are similar for AP courses. Table 3 reports the pre-COVID trend estimates for six AP subjects, with demographic coefficients in Table 5. In the preferred specification, the annual $\hat{\gamma}$ coefficients range from 0.8 percentage points (AP English Language) to 1.6 percentage points (AP U.S. History and AP Calculus). These magnitudes are comparable to the SOL subjects despite the higher baseline P(A) in AP courses (37–48% versus 24–30% for SOL), consistent with grade inflation operating broadly across the curriculum. Figure 6, Panel B confirms that the year fixed effect pattern in AP courses mirrors the three-phase structure—pre-COVID drift, pandemic jump, partial reversion—documented for SOL subjects. The full-period AP regressions are in Appendix Table B2.

4.3 Cross-Sectional Grading Structure by Demographics

Tables 4 and 5 report the demographic and absence coefficients from Specifications 3 and 4 for the pre-COVID period: Panel A estimates without school fixed effects, Panel B with school fixed effects added. Beyond the over-time trend documented in Section 4.2, the demographic coefficients reveal systematic differences in the grade–test score relationship by gender, race, and economic status. Following Tyner and Gershenson (2020), we refer to these as patterns of "differential grade inflation"—differences in how the same measured performance maps into a grade.

The Hispanic gap is somewhat smaller in magnitude than the Black gap (−3 to −6 pp in SOL subjects, −5 to −8 pp in AP subjects). The Asian coefficient is positive in 4 of 7 SOL subjects without school fixed effects (Algebra I, Biology, Earth Science, and English Reading, 6.6–9.1 pp) but is near zero or negative in Algebra II, Geometry, and Chemistry; with school fixed effects

added it is positive in every subject, ranging from 3.2 pp (Algebra II) to 10.0 pp (Biology), and remains near zero in AP. Economically disadvantaged students face a penalty of 0.25–7.1 pp in SOL subjects and 1.25–7.9 pp in AP subjects, with the largest gaps in Earth Science (−7.1 pp) and English Reading (−4.6 pp); in contrast to the racial gap, the ED penalty widens slightly when school fixed effects are added.

The female premium is the largest and most consistent demographic pattern. In Table 4, Panel A, the female coefficient is positive in every SOL subject: 7.3 pp in Algebra I, 6.8 pp in Geometry, 7.7 pp in Algebra II, 14.3 pp in Biology, 10.4 pp in Chemistry, 12.4 pp in Earth Science, and 13.2 pp in English Reading. The premium in AP courses (Table 5) is similar in magnitude (8–12 pp). Importantly, the female coefficient is essentially unchanged when school fixed effects are added (Panel B versus Panel A), indicating that the female grading advantage is a within-school phenomenon rather than a reflection of female sorting across schools. This pattern is consistent with evidence that grades reward conscientiousness and behavioral engagement—traits on which female students tend to score higher—beyond what standardized tests capture (Borghans et al., 2016).

The coefficient on the indicator for Black students is negative in every subject: −6.2 pp in Algebra I, −7.1 pp in Algebra II, −5.8 pp in Geometry, −5.5 pp in Biology, −7.8 pp in Chemistry, −4.9 pp in Earth Science, and −7.5 pp in English Reading (Table 4, Panel A). When school fixed effects are added (Panel B), the Black coefficient compresses in SOL subjects—from −0.062 to −0.028 in Algebra I, from −0.071 to −0.036 in Algebra II, and from −0.075 to −0.050 in English Reading—implying that between-school sorting accounts for roughly 20–55 percent of the unconditional gap across subjects. In AP courses, the Black coefficient is largely unchanged by the addition of school fixed effects in some subjects (Table 5): the Calculus coefficient moves from −0.102 to −0.110, and AP U.S. History from −0.072 to −0.074. This is not uniform, however: the coefficient compresses by 18 percent in AP English Language (from −0.107 to −0.088) and by 29 percent in AP U.S. Government (from −0.124 to −0.088), indicating that between-school sorting plays a larger role in those two subjects.

We treat the racial, gender, and economic-status grading structures documented here as per-

sistent features of the grading process rather than transient artifacts. While we are not positioned to resolve the underlying mechanism, competing hypotheses include omitted skill measures correlated with demographics, differential measurement error in test scores across groups, and subjective components of grading that correlate with demographic characteristics.[8]

4.4 Oaxaca–Blinder Decomposition

To summarize these results and distinguish trend changes from compositional changes, we turn to an Oaxaca–Blinder decomposition. The decomposition separates the change in P(A) between two periods into a portion attributable to changes in student composition (the "explained" component) and a portion attributable to changes in the mapping from characteristics to grades (the "unexplained" component, interpretable as a change in grading standards).

Using the specification with demographic and attendance covariates (see Appendix Tables C1 and C3), we report decompositions for two period pairs in Table 6, with the comparison between the 2014 and 2019 high school graduation cohorts in Panel A capturing the pre-COVID portion of the trend, and the comparison between 2014 and 2024 in Panel B spanning the full period including the post-COVID observations.

For the pre-COVID comparison (Panel A), the unexplained component is the dominant driver of rising P(A) in nearly every subject. In Algebra I, 102% of the 11.7 pp increase in P(A) is unexplained: the composition of test-takers shifted slightly against higher grading, making the standards-based change marginally larger than the raw increase. An unexplained share above 100% is indicative of the compositional change moving the other direction, in this case toward lower grades. English Reading shows 107% unexplained. Across SOL subjects, changes in who takes the courses and their measured ability contribute little to the rise in P(A); changing grading standards do nearly all of the work. The AP results in the lower part of Panel A tell the same story: 33–136% unexplained across the six AP subjects. The full-period decomposition (Panel B) amplifies these

[8] Adding 8th grade math and reading SOL scores to Specification 3 modestly attenuates the demographic coefficients but leaves the qualitative pattern intact. In the pre-COVID SOL sample, the Black coefficient attenuates by roughly 10–15% (e.g., from −0.062 to −0.056 in Algebra I), the Hispanic coefficient is essentially unchanged, the economically disadvantaged coefficient attenuates by 15–30%, and the female premium attenuates by 1–2 pp.

patterns.

What the descriptive regression analysis highlights is the presence of a meaningful wedge between grades and test scores, with this difference growing over time. Moreover, there is also a persistent difference by demographic group as shown in Section 4.3.

5 The Content of High School Grade Inflation

Changes in grading standards rather than shifts in the composition of test-takers drive grade inflation. This section characterizes the factors of assessment and measurement that explain demographic and over-time differences in grading.

5.1 Framework and Estimation

To fix ideas, we model the grading process as depending on two latent student traits. Let X_{ic} denote student i's content knowledge in course c: mastery of the curriculum-specific material that end-of-course examinations are designed to measure, which varies across courses within a student. Let Z_i denote a stable behavioral trait that grades reward and standardized tests do not measure directly, comprising class participation, attentiveness, homework completion, and task persistence. Course-specific departures from this student-level mean, such as engagement with particular material or the relationship with a particular teacher, enter the residual rather than Z_i.

The latent grade—the continuous assessment underlying the discrete letter grade that appears on the transcript—depends on both traits:

$$G^*_{ict} = \alpha_s^{(t)} + \beta_1^{(t)} X_{ic} + \beta_2^{(t)} Z_i + u_{ict} \tag{2}$$

where $\alpha_s^{(t)}$ is the grading intercept of the school student i attends and u_{ict} captures idiosyncratic grading noise, unobserved teacher effects, and course-specific effort. The superscript (t) indexes the graduating cohort and allows the mapping from latent traits to grades to shift across cohorts—

the mechanism through which grade inflation operates.

School-specific thresholds map G^* into letters, $G_{ic} \in \{A, B, C, D, F\}$. Because the school intercept and the school's A threshold are not separately identified, we assign all temporal variation in grading generosity to the parameters of (2) rather than to moving thresholds; this is a normalization.

While content knowledge (X_{ic}) is not observed, we observe a course-level test score T_{ic}. The test score measures content knowledge with error and that error has a persistent and transitory component:

$$T_{ic} = X_{ic} + \nu_i + \eta_{ic} \tag{3}$$

ν_i is the student's persistent tendency to over- or under-perform on standardized assessment relative to her content knowledge; a student who consistently tests below what she knows has $\nu_i < 0$. η_{ic} is transitory, capturing factors like a bad day or a lucky guess, with mean zero within student. The distinction matters because the two enter the estimating equation in different places.

Substituting $X_{ic} = T_{ic} - \nu_i - \eta_{ic}$ into (2):

$$G^*_{ict} = \alpha_s^{(t)} + \beta_1^{(t)} T_{ic} + \underbrace{\left[\beta_2^{(t)} Z_i - \beta_1^{(t)} \nu_i\right]}_{\text{stable within student}} + \underbrace{\left[u_{ict} - \beta_1^{(t)} \eta_{ic}\right]}_{\text{varies across courses}}$$

Everything stable within student across courses collects in the first bracket: the behavioral trait and the persistent component of test measurement error. A student fixed effect cannot separate them, because both are constant within student across courses. Everything that varies within student across courses collects in the second bracket and enters the course-level residual.

For each graduation cohort we estimate the following linear probability model, where g denotes the grade level in which student i took course c:[9]

$$P(A)_{ic} = \sum_b \beta_b \cdot \mathbf{1}[T_{ic} \in \text{bin } b] + \theta_c + \lambda_g + \mu_i + \varepsilon_{ic} \tag{4}$$

[9]We work with the probability that the latent grade clears the A threshold because most of the empirical change occurs at the A/B margin (Section 3).

The test score is partitioned into 50-point bins with 400–449 as the reference category; θ_c are course fixed effects; λ_g are grade-level fixed effects; μ_i is the student fixed effect; and ε_{ic} is the course-level residual. Estimation uses a high-dimensional fixed-effects estimator with student, course, and grade level absorbed, and standard errors clustered at the graduation school. School fixed effects do not appear: students are observed in a single school, so a school effect is absorbed by μ_i, and we remove school-level grading norms after estimation instead (Section 5.2).

The student fixed effect absorbs $\beta_2^{(t)} Z_i - \beta_1^{(t)} \nu_i$, net of the school mean, and the two components have different interpretations. Under the first, a student with high μ_i earns grades above what her test scores predict because she demonstrates some knowledge the examination does not measure. Under the second, she earns them because the examination understates content knowledge she in fact has: a student with $\nu_i < 0$ knows more than her SOL score indicates, and if her grades reflect that knowledge she has a positive μ_i from the measurement term alone, even with average behavior. The two interpretations differ in what grades are doing, rewarding an unmeasured trait or recovering mismeasured content, and the estimation does not distinguish them. We use "grade skill" as a label for this composite, not as a claim about its content.

Rising $P(A)$ conditional on test performance can operate through any parameter of (2). A shift in the school grading intercept raises $P(A)$ uniformly within a school-cohort, and the cross-cohort path of the estimated school means traces where that has happened. A change in the weight on content knowledge appears in the score-bin coefficients, which Section 4 reports by period. The remaining possibilities operate through the student effect, and the brackets above show why they cannot be separated within it: the weight on the behavioral trait may rise, the dispersion of that trait across students may widen, or the dispersion of persistent test mismeasurement may grow. Each is constant within student, so no rearrangement of fixed effects or variance calculations distinguishes them. Doing so requires information from outside the SOL grade-and-test-score system.

Demographic differences in μ_i suggest multiple interpretations beyond unequal distributions of Z_i. If tests are noisier measures of content knowledge for some groups, as may be the case for low-income students, then the attenuated test control leaves more true ability in the fixed effect

for those groups, generating demographic gradients in μ_i that reflect measurement rather than behavior. Alternatively, teachers may evaluate identical performance differently by race or gender, so that the grading function itself has group-specific parameters. Our framework cannot separate differences in behavior, in test reliability, and in grading; in estimation, all three are represented in demographic-specific loadings on the fixed effect.

5.2 Decomposing Grade Variation Beyond Test Scores

The sample for each cohort comprises all students observed in at least two SOL-linked course–exam pairs, yielding approximately 64,000–90,000 students per cohort observed in 2–7 courses each. Students in earlier graduation cohorts (2014–2019) average 4.9–5.6 courses per student; cohorts from 2022 onward average only 2.6–3.6 courses, while the 2020–2021 cohorts remain higher, at 4.4–5.1. Two forces drive this asymmetry. First, Virginia reduced the number of required end-of-course SOL examinations beginning with the 2021 graduation cohort, and the resulting decline in test-taking was sharpest in Algebra II, Geometry, Chemistry, and Earth Science (Table A1). Second, the COVID-era cancellation of spring 2020 SOL administrations eliminated one exam cycle for the 2021 cohort.[10]

We decompose the estimated student fixed effect into school and within-school components:

$$\hat{\mu}_i = \bar{\mu}_{s(i),t} + \tilde{\mu}_i \tag{5}$$

where $\bar{\mu}_{s(i),t}$ is the mean fixed effect among students graduating from school s in cohort t, and $\tilde{\mu}_i$ is the within-school residual—grade skill purged of institutional grading norms. Table 7 shows that $\text{Var}(\bar{\mu}_{s,t})$ is small (0.009–0.014) and roughly stable across cohorts; the within-school component carries the large majority of the variance in the student effect and all of its growth.

Because students contribute only a small number of SOL-course observations to the fixed-effect estimate, $\tilde{\mu}_i$ contains estimation noise, and its variance overstates the variance of grade

[10]The composition of courses contributing to the fixed-effect estimate changes accordingly: post-2020 estimates rest primarily on Algebra I, Earth Science, and Biology rather than the full seven-course battery available for earlier cohorts. This raises the question of whether grade skill estimated from a narrower set of courses measures the same construct, a limitation we acknowledge throughout.

skill (Walters, 2024). Student i's fixed effect has sampling variance $s_i^2 = \hat{\sigma}_\varepsilon^2/n_i$, where n_i is the number of courses in the estimation sample and $\hat{\sigma}_\varepsilon^2$ is the residual variance with degrees of freedom net of all absorbed fixed effects. School demeaning removes part of each student's noise and adds the noise in the school mean, so the sampling variance of $\tilde{\mu}_i$ and the bias-corrected variance of grade skill are

$$\tilde{s}_i^2 = s_i^2 \left(1 - \frac{2}{N_s}\right) + \frac{1}{N_s^2} \sum_{j \in s} s_j^2, \qquad \text{Var}_{BC} = \text{Var}(\tilde{\mu}_i) - \overline{\tilde{s}_i^2} \tag{6}$$

where N_s is the number of students in i's school-cohort. Where grade skill enters a regression as an explanatory variable, we also use Empirical Bayes estimates, which shrink each student's effect toward the school-cohort mean in proportion to its noise:

$$r_i = \frac{\text{Var}_{BC}}{\text{Var}_{BC} + \tilde{s}_i^2}, \qquad \tilde{\mu}_i^{EB} = r_i \cdot \tilde{\mu}_i \tag{7}$$

Students observed in more courses are shrunk less. Mean reliability $\bar{r}_i$ ranges from 0.707 to 0.742 in pre-COVID cohorts. It falls to 0.585 in 2023 as courses per student fall, then recovers to 0.681 by 2025 as coverage partially returns.

The variance of grades unexplained by test performance has grown, and the decomposition asks where. If the growth is in the student effect, the widening wedge between grades and test scores is systematic: it attaches to students rather than to particular course-year draws, and an outside observer could in principle learn from it. If the growth is in the course-level residual, grades have become noisier from course to course in ways unrelated to any stable student attribute. What the comparison cannot establish is which stable component is expanding. Growing transitory test noise raises $\text{Var}(\varepsilon)$ and leaves $\text{Var}(\tilde{\mu})$ unchanged, because η_{ic} varies within student; growing dispersion in persistent test mismeasurement does the opposite, raising $\text{Var}(\tilde{\mu})$ and leaving $\text{Var}(\varepsilon)$ unchanged, which is the same signature as a rising weight on the behavioral trait.

Table 7 presents the full decomposition for the preferred specification (binned SOL, full period). The within-school grade-skill variance rises from 0.085 in 2014 to 0.131 in 2024—a 54 percent increase in raw terms. The bias-corrected variance, which strips estimation noise, rises less:

from 0.060 in 2014 to 0.085 in 2024, a 42 percent increase. The residual variance also rises, from 0.104 in 2014 to 0.130 in 2024, a 25 percent increase that is smaller than the growth in the student effect. Figure 7 displays the time series of standard deviations: the corrected grade-skill SD rises from 0.244 to 0.291, and the residual SD from 0.322 to 0.360.

Because students contribute only a small number of SOL-course observations to the estimate, the noise correction is quantitatively important. In 2024, the raw SD of within-school grade skill is 0.363; the corrected SD is 0.291. The implied estimation-noise variance is $0.131 - 0.085 = 0.046$ (an SD of roughly 0.21), which would contaminate cross-cohort comparisons if uncorrected. The correction matters most for recent cohorts, where fewer courses per student inflate the raw variance mechanically.

The expansion of grade-skill variance is gradual in the pre-COVID period—SD_{BC} rises from 0.244 to 0.263 between 2014 and 2019—and then accelerates post-2020, reaching 0.291 by 2024. This acceleration is consistent with pandemic-era changes in grading practice (expanded retake policies, increased homework weight, pass/fail options) that persisted beyond the emergency period. We note, however, that the course-composition shift toward fewer and earlier-grade courses in post-2020 cohorts may contribute: grade skill estimated from a narrower set of ninth-grade courses could exhibit more dispersion than grade skill estimated from the full battery, a limitation that the noise correction cannot fully address.

Over the pre-COVID window the absence penalty steepens in every SOL subject, from -0.23 to -0.43 in Algebra I, -0.30 to -0.45 in Geometry, and -0.62 to -0.88 in Chemistry, with smaller increases elsewhere (Appendix Table C1), and it steepens in all six AP subjects as well (Appendix Table C2). The steeper penalty is consistent with a grading function that weights attendance more heavily. Because absence rates also rose over the period, however, this pattern alone cannot distinguish a change in the weight on attendance from a change in its distribution.

A rising weight on a behavioral trait unequally distributed by gender and socioeconomic status carries a further implication: the demographic gradients in grade skill should widen over time. Table 9 shows that they do. The female advantage in within-school grade skill grew from $+0.097$ in 2014 to $+0.133$ in 2024, a 37 percent increase, and the penalty for economically disadvantaged

students widened from −0.097 to −0.141 over the same cohorts, a 45 percent increase. The Asian advantage also grew, from +0.094 to +0.125. By contrast, the Black coefficient shows no trend, ranging from −0.042 to −0.057 across cohorts, and the Hispanic coefficient is similarly stable through 2023 (−0.046 to −0.056) before widening to −0.062 and −0.067 in the 2024 and 2025 cohorts. Demographics explain only 8–11 percent of within-school grade-skill variance; the remaining 89–92 percent is within-group variation. Figure 8 traces these coefficients across cohorts.

The demographic gradients in grade skill can be compared with the single-course demographic grading gaps documented in Section 4.3, since both are measured in units of $P(A)$. There the within-school female premium ranges from 0.067 to 0.143 and the economic-disadvantage penalty from −0.026 to −0.080 across SOL subjects (Table 4, Panel B). The female gradient in grade skill falls within the single-course range, while the economic-disadvantage gradient (−0.097 in 2014, −0.141 in 2024) exceeds every single-course penalty.

5.3 What Grade Skill Predicts: Overall HSGPA

Grade skill is estimated exclusively from SOL-linked courses—seven subjects that constitute only a fraction of a student's full transcript. If the fixed effect captures a genuinely stable and transferable trait, it should predict grades in the many courses outside the SOL-tested set. We assess this by regressing cumulative HSGPA on within-school grade skill, controlling for course-specific SOL scores and demographics:

$$\text{HSGPA}_i = \alpha + \gamma\tilde{\mu}_i + \sum_c \beta_c \text{SOL}_{ic} + \mathbf{D}_i'\delta + \varepsilon_i \tag{8}$$

Table 8 reports the results. The coefficient $\hat{\gamma}$ on raw grade skill ranges from 0.878 to 1.025 across cohorts. Because $\tilde{\mu}_i$ contains estimation noise, this coefficient is attenuated toward zero. Replacing $\tilde{\mu}_i$ with the Empirical Bayes estimate $\tilde{\mu}_i^{EB}$ from (7), with the same controls and sample, corrects the attenuation and yields $\hat{\gamma}_{EB}$ between 1.269 and 1.572: a one-unit increase in the SOL-derived grade-skill measure corresponds to more than a one-unit increase in overall GPA conditional on test performance. The partial R^2—the incremental explanatory power of grade

skill beyond SOL scores and demographics—rises by roughly 56 percent over the study period, from 0.186 in 2014 to 0.291 in 2024 (Table 8). As grades become more compressed and inflated, the within-school behavioral component captured by our fixed effect becomes relatively more informative about overall academic credentials. Raw GPA carries less marginal information—it is increasingly a near-ceiling measure that fails to discriminate among students—while the component of grades orthogonal to test scores grows in both variance and predictive relevance.

5.4 What Grade Skill Predicts: AP Examinations

Advanced Placement courses are taken by a selected subset of students—those choosing a more rigorous curriculum—and AP examinations are not universally required. Nevertheless, the AP setting provides a distinctive validation opportunity: course grades are assigned by classroom teachers while examination scores are assigned independently by the College Board, yielding a within-student pairing of local grading and external assessment in the same subject. We estimate:

$$\mathbf{1}[G_{ij}^{AP} = A] = \sum_{s} \gamma_s \cdot \mathbf{1}[\text{AP Exam}_{ij} = s] + \phi\tilde{\mu}_i + \mathbf{D}_i'\delta + \varepsilon_{ij} \tag{9}$$

where G_{ij}^{AP} is the course grade in AP subject j, AP Exam dummies control for the externally assessed score (1–5), and ϕ captures the predictive content of SOL-derived grade skill for AP course grades conditional on demonstrated subject mastery. Estimation noise in $\tilde{\mu}_i$ attenuates $\hat{\phi}$, and because AP courses are not among the SOL-linked courses from which $\tilde{\mu}_i$ is estimated, that noise is unrelated to the AP course grade. We therefore report $\hat{\phi}_{EB}$, the coefficient on the Empirical Bayes estimate $\tilde{\mu}_i^{EB}$ from (7), as the main estimate; Figure 9 displays its time series for each subject. Across the three subjects with full-period estimates (Biology, English Literature, and U.S. Government), $\hat{\phi}_{EB}$ ranges from 0.637 to 0.874 (raw: 0.405 to 0.637), confirming that grade skill transfers substantially beyond the SOL grading environment (Appendix Table C5 reports the full results).

The evidence on change over time is weaker. From 2014 to 2024, $\hat{\phi}_{raw}$ falls by 0.051 to 0.067 in these subjects, and each decline is statistically significant ($t = 2.1$ to 3.3 for the 2014–2024 difference). $\hat{\phi}_{EB}$ falls by a similar 0.040 to 0.065, but its standard errors are larger, and none of

these declines is statistically significant at the 5 percent level ($t = 1.1$ to 1.9). Mean reliability also falls over the period (Table 7), so part of the raw decline reflects greater attenuation in later cohorts rather than a weaker link between grade skill and AP grades. Within the period, $\hat{\phi}_{EB}$ peaks between 2014 and 2018, and its largest peak-to-2024 decline is 0.118 (Figure 9). What decline there is coincides with rising mean $P(A)$ in these AP courses, from 0.382–0.491 in 2014 to 0.568–0.598 in 2024. This is consistent with ceiling compression: when most students receive As, the course grade discriminates less among students on any dimension, including the one our fixed effect captures. Our estimates do not establish that mechanism.

Both tests establish that the student component is stable and transfers across grading contexts, teachers, and subjects. Neither separates behavior from persistent test mismeasurement. Transferability to graded outcomes is necessary but not sufficient for the behavioral interpretation.

6 Discussion

Over the last decade, high school grades in Virginia have risen substantially with the increase concentrated at the top of the distribution. Mean overall HSGPA increased from 2.78 for the 2014 graduation cohort to 2.92 by 2021, and the share of students with a GPA of 3.5 or above grew from 23 to 31 percent. In SOL-tested subjects, the probability of earning an A conditional on test performance rose markedly (by 3 to 11 percentage points across subjects) over the pre-COVID window, and AP course grades followed the same pattern. Yet test scores showed no commensurate gains. Pandemic-era grading policies added further increases, with only partial reversion in the most recent years.

Neither changes in the demographic composition of students nor the distribution of students across schools accounts for the observed increases in grades. Oaxaca–Blinder decompositions attribute essentially all of the rise to the mapping from characteristics to grades rather than to the mix of students taking the courses. In several subjects the compositional change works against higher grades, so the change in grading standards exceeds the observed increase. Similar stu-

dents, with the same measured achievement, receive higher grades at the end of the period than at the beginning.

Grades and standardized tests both attempt to measure content knowledge and what a student has learned, but grades may also capture conduct and achievement beyond material tested, while exams may measure learning with error. These features complicate inference. A student who earns a higher grade than her test score predicts may have supplied engagement the examination ignores, or the examination may have understated what she knows. Neither is visible in one course, but both are properties of the student, not the occasion: a bad testing day does not follow her from course to course. Pooling a student's tested courses and estimating a student fixed effect therefore separates what is persistent in the grade–test wedge from what is transitory.

The expansion of grade variation beyond test scores concentrates in the persistent component, the student fixed effect. Net of estimation noise, its within-school variance grew 42 percent between the 2014 and 2024 graduation cohorts, while the variance of the course-level residual grew a smaller 25 percent, so the growth is not simply grades becoming noisier from course to course. This student-level effect from courses with SOL exams explains other academic outcomes, offering significant explanatory power in the prediction of overall academic performance and AP course grades conditional on externally scored AP examinations. While identification of the exact traits and skills associated with these effects is beyond what our data allow, these effects are consistent with the presence of a student-level "grade skill" which can be interpreted as a bundle of behavioral traits that impact success in academic settings. The behavioral interpretation is supported by the observation that absences (or non-attendance) are increasingly tied to lower grades over the pre-COVID period conditional on test performance. In addition, the demographic gradients in this student-level component are similar in magnitude to the course-level grading differences by gender, race, and economic circumstances documented above. Classroom- and student-level observation is needed to better understand how student behaviors, background and prior achievement impact grades.

One open question concerns how the policies that shape grading have changed over the last decade and how these policies impact the information-content of grades. School divisions in Vir-

ginia (like those in many other states) adopted standards-based and equity-oriented grading practices unevenly and at different times over our period, and reforms such as minimum grade floors, expanded retakes, and the removal of penalties for late work plausibly contribute to the changes in grades that we observe. In addition, the mandated accountability environment shifted as well as the Every Student Succeeds Act (2015) supplanted No Child Left Behind (2001). Wyckoff (2026) notes that the resulting relaxation in accountability tied to test outcomes is one candidate explanation for the decline in NAEP achievement which began before the pandemic. Whether and how changes in grading guidelines and accountability policy change the assignment of grades and the relationship between grades and other outcomes is an open question where we hope to add more evidence.

How changes in the assignment of grades affect postsecondary outcomes is a central unanswered question. Preliminary work linking the cohorts observed in this research to Virginia postsecondary records suggests that content preparation binds in quantitative gateway courses and matters considerably less elsewhere, so a change in what grades measure carries consequences that depend on where a student enrolls and what she takes. More broadly, future work to understand the information content of high school grades and test scores is important for helping students, along with colleges and employers, to identify strong matches in college choice, course-taking, and employment.

References

Arundel, K. 2020. "How educators are tweaking grading approaches in response to the pandemic." December, https://www.k12dive.com/news/how-educators-are-tweakinggrading-approaches-in-response-to-the-pandemic/591729/.

Betts, Julian R., and Jeff Grogger. 2003. "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings." *Economics of Education Review* 22 (4): 343–352.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries. 2016. "What Grades and Achievement Tests Measure." *Proceedings of the National Academy of Sciences* 113 (47): 13354–13359. 10.1073/pnas.1601135113.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2009. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.

Chan, William, Hao Li, and Wing Suen. 2007. "A Signaling Theory of Grade Inflation." *International Economic Review* 48 (3): 1065–1090.

Coffey, M. 2023. "Grading for equity isn't grounded in reality." May, https://fordhaminstitute.org/national/commentary/grading-equity-isnt-grounded-reality.

Denning, Jeffrey T., Rachel L. Nesbit, Nolan G. Pope, and Merrill Warnick. 2026. "Easy A's, Less Pay: The Long-Term Effects of Grade Inflation." Working Paper 34952, National Bureau of Economic Research. 10.3386/w34952.

Feldman, Joe. 2019. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. Thousand Oaks, CA: Corwin Press.

Figlio, David N., and Maurice E. Lucas. 2004. "Do High Grading Standards Affect Student Performance?" *Journal of Public Economics* 88 (9–10): 1815–1834.

Franco, Thomas, and Matthew Eberhardt. 2021. "Grading for Equity: A Primer." technical report, Virginia Department of Education.

Friedman, John, Bruce Sacerdote, Douglas O. Staiger, and Michele Tine. 2025. "Standardized Test Scores and Academic Performance at Ivy-Plus Colleges." *American Economic Association Papers and Proceedings* 115 676–681.

Gershenson, Seth. 2018. "Grade Inflation in High Schools (2005-2016)." technical report, Thomas B. Fordham Institute, Washington, DC.

Gershenson, Seth, Stephen B. Holt, and Adam Tyner. 2024. "Making the Grade: The Effect of Teacher Grading Standards on Student Outcomes." *Contemporary Economic Policy* 42 (2): 305–318. 10.1111/coep.12637.

Goldhaber, Dan, and Maia Goodman Young. 2023. "Course Grades as a Signal of Student Achievement: Evidence on Grade Inflation Before and After COVID-19." Policy Brief 35-1123, CALDER.

Guskey, Thomas R., and Jane M. Bailey. 2015. *Developing Standards-Based Report Cards.* Thousand Oaks, CA: Corwin Press.

Harbaugh, Rick, and Eric Rasmusen. 2018. "Coarse Grades: Informing the Public by Withholding Information." *American Economic Journal: Microeconomics* 10 (1): 210–235. 10.1257/mic.20160210.

Heckman, James J., and Tim Kautz. 2012. "Hard Evidence on Soft Skills." *Labour Economics* 19 (4): 451–464. 10.1016/j.labeco.2012.05.014.

Hess, F. M. 2024. "The trouble with 'equitable grading'." May, https://rickhess99.medium.com/the-trouble-with-equitable-grading-f936bd6fe27e.

Hoxby, Caroline M. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23 (4): 95–118.

Hurwitz, Michael, and Jangmin Lee. 2018. "Grade inflation and the role of standardized testing." In *Measuring success: Testing, grades, and the future of college admissions*, edited by Jack Buckley, Lynn Letukas, and Ben Wildavsky, Baltimore, MD: Johns Hopkins University Press.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy* 126 (5): 2072–2107. 10.1086/699018.

Kurlaender, Michal, and Kramer A. Dykeman. 2019. "Predicting College Success: How Do Different High School Assessments Measure Up?", working paper, PACE, https://edpolicyinca.org/publications/predicting-college-success-how-do-different-highschool-assessments-measure-2019.

McManus, Brian, Jessica Howell, and Michael Hurwitz. 2023. "Student Test Score Disclosure in the Presence of Test-Optional Admissions Policies." *American Economic Review: Insights* 5 (4): 464–480.

Ostrovsky, Michael, and Michael Schwarz. 2010. "Information Disclosure and Unraveling in Matching Markets." *American Economic Journal: Microeconomics* 2 (2): 34–63. 10.1257/mic.2.2.34.

Pattison, Elias, Eric Grodsky, and Chandra Muller. 2013. "Is the sky falling? Grade inflation and the signaling power of grades." *Educational Researcher* 42 (5): 259–265.

Romano, M. 2022. "Adopting equitable grading practices to improve student outcomes." June, https://www.instructure.com/resources/blog/adopting-equitable-grading-practices-improve-student-outcomes.

Sacerdote, Bruce, Douglas O. Staiger, and Michele Tine. 2025. "How Test Optional Policies in College Admissions Disproportionately Harm High Achieving Applicants from Disadvantaged Backgrounds." Working Paper 33389, National Bureau of Economic Research.

Sanchez, Edgar I. 2025. "The Implications of Grade Inflation: Evaluating the Prediction of First-Year College GPA Using High School GPA and ACT Scores." Research Report R2432, ACT, Inc.

Sanchez, Edgar I., and Raeal Moore. 2022. "Grade Inflation Continues to Grow in the Past Decade." technical report, ACT, Inc. https://www.act.org/content/dam/act/unsecured/do

cuments/2022/R2134-Grade-Inflation-Continues-to-Grow-in-the-Past-Decade-Final-Accessible.pdf.

Tyner, Adam, and Seth Gershenson. 2020. "Conceptualizing Grade Inflation." *Economics of Education Review* 78 102037. 10.1016/j.econedurev.2020.102037.

University of California Academic Senate. 2020. "Report of the UC Academic Council Standardized Testing Task Force." technical report, University of California Academic Senate, https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf.

U.S. Department of Education, National Center for Education Statistics. 2022. "High School Transcript Study." technical report, National Assessment of Educational Progress.

Virginia Department of Education. 2024. "Postsecondary Enrollment Report of Virginia Public School Graduates and Completers." https://p1pe.doe.virginia.gov/postsec_public/postsec.do?dowhat=LOAD_REPORT_C11, State Fiscal Stabilization Fund Indicator (C)(11), 2023 FGI cohort.

Virginia Department of Education. 2025. "The New Era of Student Excellence, Innovation, and Support." News release, December 10, 2025, December, https://www.doe.virginia.gov/Home/Components/News/News/510/227, Underlying data from VDOE Cohort Graduation and Dropout Reports.

Walters, Christopher. 2024. "Empirical Bayes Methods in Labor Economics." In *Handbook of Labor Economics*, edited by Christian Dustmann, and Thomas Lemieux, Volume 5. Chap. 3, 183–260, Elsevier. 10.1016/bs.heslab.2024.11.001.

Whipple, Guy Montrose. 1913. "Preface." In *The Eighteenth Yearbook of the National Society for the Study of Education, Part II*, v–vi, Bloomington, IL: Public School Publishing.

Wyckoff, Jim. 2026. "Puzzling over Declining Academic Achievement." *Journal of Policy Analysis and Management* 45 (2): e70052. 10.1002/pam.70052.

Tables

Table 1. Characteristics of Virginia High School Graduates by Period

	2014–2015	2018–2019	2024–2025
Female	0.488	0.485	0.486
Black	0.237	0.227	0.216
Hispanic	0.112	0.148	0.195
Asian	0.061	0.065	0.074
White	0.547	0.510	0.454
Econ. Disadvantaged	0.451	0.502	0.547
Absence Rate	0.114	0.118	0.138
N (students)	210,453	220,455	223,876

Notes: Demographic shares and absence rates are student-level means. Each student counted once per period.

Table 2. P(A) Trend Coefficient, SOL Subjects, 2014–2019

	Algebra I	Algebra II	Geometry	Biology	Chemistry	Earth Science	English
Cohort trend ($\hat{\gamma}$)	0.0250***	-0.0032*	0.0073***	0.0171***	0.0070***	0.0210***	0.0068***
	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0019)	(0.0022)	(0.0018)
+ Demographics	0.0230***	-0.0024	0.0076***	0.0172***	0.0074***	0.0209***	0.0079***
	(0.0017)	(0.0016)	(0.0017)	(0.0017)	(0.0019)	(0.0022)	(0.0018)
+ Absence rate	0.0225***	-0.0015	0.0077***	0.0169***	0.0079***	0.0201***	0.0084***
	(0.0017)	(0.0016)	(0.0017)	(0.0017)	(0.0019)	(0.0022)	(0.0018)
+ School FE	0.0219***	-0.0024	0.0068***	0.0186***	0.0075***	0.0184***	0.0093***
	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0019)	(0.0020)	(0.0018)
R^2 (spec 1)	0.228	0.260	0.283	0.224	0.259	0.161	0.110
R^2 (spec 2)	0.243	0.271	0.293	0.255	0.276	0.192	0.145
R^2 (spec 3)	0.246	0.275	0.295	0.260	0.281	0.197	0.158
R^2 (spec 4)	0.315	0.332	0.360	0.311	0.347	0.264	0.210
Mean P(A)	0.242	0.258	0.236	0.313	0.301	0.297	0.242
N	393,124	348,751	424,687	458,459	316,393	374,443	339,269

LPM. Outcome: P(A). Each cell reports $\hat{\gamma}$, the coefficient on a linear cohort trend. Each row adds controls cumulatively. SE clustered at school level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 3. P(A) Trend Coefficient, AP Subjects, 2014–2019

	AP Calc	AP Bio	AP ELang	AP ELit	AP Gov	AP Hist
Cohort trend ($\hat{\gamma}$)	0.0115***	0.0124***	0.0088**	0.0096*	0.0146***	0.0171***
	(0.0034)	(0.0037)	(0.0042)	(0.0053)	(0.0041)	(0.0035)
+ Demographics	0.0117***	0.0126***	0.0086**	0.0096*	0.0154***	0.0174***
	(0.0034)	(0.0037)	(0.0041)	(0.0052)	(0.0040)	(0.0035)
+ Absence rate	0.0113***	0.0124***	0.0085**	0.0094*	0.0148***	0.0181***
	(0.0033)	(0.0037)	(0.0043)	(0.0053)	(0.0040)	(0.0035)
+ School FE	0.0143***	0.0087**	0.0077*	0.0123**	0.0142***	0.0163***
	(0.0032)	(0.0036)	(0.0044)	(0.0054)	(0.0040)	(0.0037)
R^2 (spec 1)	0.132	0.129	0.127	0.087	0.135	0.174
R^2 (spec 2)	0.143	0.141	0.145	0.114	0.158	0.182
R^2 (spec 3)	0.155	0.157	0.158	0.128	0.168	0.191
R^2 (spec 4)	0.229	0.281	0.262	0.224	0.279	0.311
Mean P(A)	0.365	0.408	0.436	0.478	0.421	0.381
N	62,405	44,521	78,808	76,640	94,376	94,126

LPM. Outcome: P(A). Each cell reports $\hat{\gamma}$, the coefficient on a linear cohort trend. Each row adds controls cumulatively. SE clustered at school level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 4. Regression of A Grade on SOL Scores and Covariates, Pre-COVID 2014–2019

Panel A: Specification 3 (No School Fixed Effects)

	Alg I	Alg II	Geom	Bio	Chem	Earth Sci	Eng Read
Female	0.0734***	0.0772***	0.0684***	0.1425***	0.1039***	0.1241***	0.1319***
	(0.0022)	(0.0023)	(0.0023)	(0.0031)	(0.0027)	(0.0030)	(0.0034)
Black	−0.0620***	−0.0709***	−0.0575***	−0.0546***	−0.0783***	−0.0492***	−0.0753***
	(0.0065)	(0.0059)	(0.0058)	(0.0076)	(0.0070)	(0.0089)	(0.0075)
Hispanic	−0.0326***	−0.0580***	−0.0566***	−0.0567***	−0.0637***	−0.0256***	−0.0523***
	(0.0057)	(0.0055)	(0.0050)	(0.0069)	(0.0068)	(0.0073)	(0.0066)
Asian	0.0792***	−0.0179*	−0.0003	0.0655***	−0.0038	0.0909***	0.0788***
	(0.0087)	(0.0098)	(0.0092)	(0.0096)	(0.0132)	(0.0191)	(0.0210)
Econ Disadvantaged	−0.0255***	−0.0025	−0.0071*	−0.0406***	−0.0094*	−0.0711***	−0.0459***
	(0.0049)	(0.0045)	(0.0043)	(0.0052)	(0.0056)	(0.0062)	(0.0052)
Absence Rate	−0.4122***	−0.6774***	−0.4185***	−0.5965***	−0.7805***	−0.5072***	−0.8067***
	(0.0292)	(0.0334)	(0.0277)	(0.0313)	(0.0418)	(0.0401)	(0.0318)
R^2	0.246	0.275	0.295	0.260	0.281	0.197	0.158
N	393,124	348,751	424,687	458,459	316,393	374,443	339,269
Mean P(A)	0.249	0.267	0.244	0.317	0.305	0.298	0.258

Panel B: Specification 4 (Adding School Fixed Effects)

	Alg I	Alg II	Geom	Bio	Chem	Earth Sci	Eng Read
Female	0.0726***	0.0742***	0.0674***	0.1429***	0.1007***	0.1250***	0.1310***
	(0.0015)	(0.0015)	(0.0014)	(0.0018)	(0.0017)	(0.0018)	(0.0019)
Black	−0.0279***	−0.0357***	−0.0296***	−0.0344***	−0.0442***	−0.0395***	−0.0499***
	(0.0017)	(0.0020)	(0.0016)	(0.0019)	(0.0021)	(0.0020)	(0.0020)
Hispanic	−0.0049**	−0.0091***	−0.0087***	−0.0095***	−0.0164***	−0.0165***	−0.0282***
	(0.0022)	(0.0022)	(0.0021)	(0.0024)	(0.0025)	(0.0026)	(0.0025)
Asian	0.0854***	0.0318***	0.0403***	0.1002***	0.0544***	0.0785***	0.0870***
	(0.0032)	(0.0033)	(0.0031)	(0.0031)	(0.0031)	(0.0043)	(0.0041)
Econ Disadvantaged	−0.0339***	−0.0262***	−0.0303***	−0.0691***	−0.0366***	−0.0801***	−0.0608***
	(0.0016)	(0.0017)	(0.0014)	(0.0018)	(0.0018)	(0.0019)	(0.0019)
Absence Rate	−0.4841***	−0.6921***	−0.4835***	−0.7178***	−0.7936***	−0.6772***	−0.8997***
	(0.0139)	(0.0182)	(0.0134)	(0.0157)	(0.0195)	(0.0151)	(0.0168)
R^2	0.353	0.345	0.381	0.329	0.368	0.298	0.239
N	392,996	348,645	424,556	458,358	316,332	374,312	339,199

Notes: OLS regression of **1**(earned A) on SOL score bin dummies (omitted: 400–449), cohort trend, demographics, and absence rate. Panel A: no school fixed effects. Panel B adds school fixed effects (absorbed via reghdfe). SE clustered at school level in parentheses. Score bin coefficients not shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5. Regression of A Grade on AP Exam Scores and Covariates, Pre-COVID 2014–2019

Panel A: Specification 3 (No School Fixed Effects)

	Calc AB	AP Bio	AP Eng Lang	AP Eng Lit	AP US Govt	AP US Hist
Female	0.0835***	0.0918***	0.1033***	0.1197***	0.1163***	0.0774***
	(0.0066)	(0.0066)	(0.0060)	(0.0051)	(0.0059)	(0.0054)
Black	−0.1021***	−0.0835***	−0.1067***	−0.1233***	−0.1238***	−0.0723***
	(0.0129)	(0.0154)	(0.0132)	(0.0130)	(0.0118)	(0.0132)
Hispanic	−0.0535***	−0.0663***	−0.0617***	−0.0608***	−0.0817***	−0.0653***
	(0.0102)	(0.0136)	(0.0091)	(0.0131)	(0.0097)	(0.0107)
Asian	−0.0063	−0.0345	−0.0017	0.0209	0.0023	−0.0033
	(0.0108)	(0.0216)	(0.0197)	(0.0144)	(0.0172)	(0.0160)
Econ Disadvantaged	−0.0228***	−0.0192**	−0.0492***	−0.0793***	−0.0467***	−0.0125
	(0.0088)	(0.0093)	(0.0098)	(0.0096)	(0.0089)	(0.0089)
Absence Rate	−1.4762***	−1.5328***	−1.4224***	−1.3001***	−1.1077***	−1.2987***
	(0.1273)	(0.1073)	(0.0982)	(0.1254)	(0.1012)	(0.1064)
R^2	0.155	0.157	0.158	0.128	0.168	0.191
N	62,405	44,521	78,808	76,640	94,376	94,126
Mean P(A)	0.365	0.408	0.436	0.478	0.421	0.381

Panel B: Specification 4 (Adding School Fixed Effects)

	Calc AB	AP Bio	AP Eng Lang	AP Eng Lit	AP US Govt	AP US Hist
Female	0.0777***	0.1032***	0.1128***	0.1248***	0.1160***	0.0775***
	(0.0043)	(0.0048)	(0.0037)	(0.0038)	(0.0036)	(0.0034)
Black	−0.1099***	−0.0725***	−0.0877***	−0.1046***	−0.0881***	−0.0737***
	(0.0076)	(0.0072)	(0.0056)	(0.0059)	(0.0052)	(0.0049)
Hispanic	−0.0312***	−0.0259***	−0.0364***	−0.0504***	−0.0536***	−0.0351***
	(0.0076)	(0.0080)	(0.0060)	(0.0063)	(0.0053)	(0.0052)
Asian	0.0423***	0.0724***	0.0411***	0.0391***	0.0539***	0.0464***
	(0.0052)	(0.0062)	(0.0052)	(0.0061)	(0.0052)	(0.0049)
Econ Disadvantaged	−0.0418***	−0.0501***	−0.0649***	−0.0827***	−0.0675***	−0.0556***
	(0.0051)	(0.0056)	(0.0044)	(0.0048)	(0.0042)	(0.0038)
Absence Rate	−1.4476***	−1.3347***	−1.4395***	−1.5255***	−1.1835***	−1.2817***
	(0.0719)	(0.0650)	(0.0499)	(0.0460)	(0.0428)	(0.0473)
R^2	0.266	0.322	0.308	0.289	0.322	0.355
N	62,330	44,454	78,761	76,593	94,321	94,056

Notes: OLS regression of **1**(earned A) on AP exam score dummies (omitted: score 3), cohort trend, demographics, and absence rate. Panel A: no school fixed effects. Panel B adds school fixed effects (absorbed via reghdfe). SE clustered at school level in parentheses. Score category coefficients not shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6. Oaxaca-Blinder Decomposition of P(A) Changes

Subject	Early P(A)	Late P(A)	Raw Δ	Explained	Unexplained	% Unexpl.	N
Panel A: 2014 vs. 2019							
SOL-Linked Courses							
Algebra I	0.167	0.284	0.117	−0.002	0.119	102.0	115,324
Algebra II	0.245	0.314	0.070	−0.003	0.073	104.3	116,991
Geometry	0.187	0.276	0.089	0.028	0.061	68.3	133,336
Biology	0.296	0.341	0.045	−0.039	0.084	186.5	148,916
Chemistry	0.276	0.346	0.070	0.023	0.046	66.7	102,598
Earth Science	0.281	0.310	0.029	−0.056	0.086	291.9	118,777
English	0.244	0.278	0.034	−0.002	0.036	106.8	115,026
AP Courses							
AP Calculus	0.346	0.390	0.044	−0.016	0.059	135.7	20,213
AP Biology	0.362	0.437	0.075	0.019	0.056	74.4	15,031
AP Eng. Language	0.388	0.475	0.087	0.058	0.029	33.3	22,531
AP Eng. Literature	0.463	0.493	0.030	−0.007	0.036	122.1	25,326
AP US Government	0.382	0.473	0.091	0.006	0.086	93.6	31,080
AP US History	0.340	0.428	0.089	−0.007	0.096	107.8	30,488
Panel B: 2014 vs. 2024							
SOL-Linked Courses							
Algebra I	0.167	0.338	0.171	−0.007	0.178	103.9	87,463
Algebra II	0.245	0.509	0.265	−0.024	0.289	109.0	69,214
Geometry	0.187	0.379	0.192	−0.049	0.241	125.3	81,617
Biology	0.296	0.411	0.114	−0.085	0.200	174.3	148,128
Chemistry	0.276	0.310	0.034	−0.126	0.160	474.0	51,931
Earth Science	0.281	0.374	0.094	−0.110	0.204	217.5	79,376
English	0.244	0.354	0.110	0.052	0.058	52.8	115,214
AP Courses							
AP Calculus	0.346	0.434	0.088	−0.026	0.114	128.9	20,111
AP Biology	0.362	0.522	0.160	0.016	0.144	90.0	14,551
AP Eng. Language	0.388	0.581	0.193	0.035	0.158	81.8	21,615
AP Eng. Literature	0.463	0.568	0.105	0.003	0.102	97.3	24,561
AP US Government	0.382	0.530	0.148	0.025	0.124	83.5	30,601
AP US History	0.340	0.508	0.169	−0.050	0.219	129.7	29,133

Twofold decomposition with early-period coefficients as reference structure. Positive Raw Δ indicates P(A) rose. "Explained" is the change attributable to shifts in observable student characteristics; "Unexplained" is the change attributable to how characteristics are mapped to grades. Algebra II uses 2015 as early anchor (test-form transition). Covariates: SOL bin/AP score level dummies + demographics + absence rate. Source: `04_oaxaca_decomposition.do`.

Table 7. Variance Decomposition of "Grade Skill," Full Period, Binned SOL Scores

Cohort	N_{FE}	N_{obs}	$\bar{n}_i$	$\text{Var}(\hat{\mu})$	$\text{Var}(\bar{\mu}_s)$	$\text{Var}(\tilde{\mu})$	$\text{SD}(\tilde{\mu})$	Var_{BC}	SD_{BC}	$\bar{r}_i$
2014	72,423	332,710	4.92	0.098	0.012	0.085	0.292	0.060	0.244	0.707
2015	75,012	384,564	5.44	0.096	0.011	0.085	0.291	0.061	0.247	0.727
2016	80,155	421,385	5.57	0.096	0.010	0.085	0.292	0.062	0.248	0.731
2017	82,579	437,058	5.57	0.095	0.009	0.086	0.293	0.062	0.250	0.731
2018	87,928	462,681	5.59	0.097	0.009	0.088	0.296	0.063	0.252	0.730
2019	89,267	464,888	5.52	0.103	0.009	0.094	0.306	0.069	0.263	0.742
2020	88,502	429,985	5.09	0.108	0.010	0.098	0.312	0.071	0.266	0.731
2021	85,336	355,476	4.36	0.116	0.011	0.105	0.325	0.073	0.270	0.697
2022	83,575	281,621	3.50	0.130	0.010	0.120	0.347	0.076	0.276	0.641
2023	63,780	190,177	2.62	0.147	0.014	0.133	0.364	0.077	0.277	0.585
2024	84,183	257,565	3.11	0.143	0.012	0.131	0.363	0.085	0.291	0.651
2025	89,691	306,373	3.57	0.137	0.011	0.126	0.355	0.085	0.291	0.681

Notes: Model: $P(A)_{ic} = \mathbf{X}_i'\beta + \mu_i + \theta_c + \lambda_g + \varepsilon_{ic}$, estimated by reghdfe with student, course, and grade level absorbed FEs, clustered at graduation school. $\tilde{\mu}_i = \hat{\mu}_i - \bar{\mu}_{school}$. Each student's sampling variance is $s_i^2 = \hat{\sigma}_\varepsilon^2/n_i$, with $\hat{\sigma}_\varepsilon^2 = \sum \hat{\varepsilon}^2/(N_{obs} - N_{students} - N_{courses} - N_{grades} - N_{bins})$; after school demeaning, $\tilde{s}_i^2 = s_i^2(1 - 2/N_s) + N_s^{-2}\sum_{j \in s} s_j^2$, where N_s is the number of students in i's school-cohort. Bias-corrected variance $\text{Var}_{BC} = \text{Var}(\tilde{\mu}) - \text{mean}(\tilde{s}_i^2)$, the variance of grade skill net of estimation noise; $\text{SD}_{BC} = \sqrt{\text{Var}_{BC}}$. Reliability $r_i = \text{Var}_{BC}/(\text{Var}_{BC} + \tilde{s}_i^2)$, used for the Empirical Bayes estimates $\tilde{\mu}_i^{EB} = r_i\tilde{\mu}_i$. N_{FE} = students with $n_i \geq 2$ courses in the estimation sample.

Table 8. Validation: Grade Skill Predicting Cumulative HSGPA

Cohort	$\hat{\gamma}_{raw}$	SE	N	R^2_{full}	$R^2_{-\mu}$	Partial R^2	$\hat{\gamma}_{EB}$	SE
2014	0.946	0.007	72,422	0.610	0.521	0.186	1.348	0.011
2015	0.936	0.007	74,883	0.633	0.552	0.181	1.277	0.010
2016	0.940	0.007	80,017	0.643	0.559	0.190	1.269	0.010
2017	0.964	0.007	82,433	0.659	0.572	0.204	1.292	0.009
2018	0.960	0.006	87,923	0.665	0.574	0.215	1.292	0.009
2019	0.976	0.006	89,266	0.672	0.575	0.227	1.293	0.008
2020	0.970	0.006	88,499	0.662	0.561	0.230	1.297	0.008
2021	0.909	0.006	85,334	0.638	0.541	0.211	1.269	0.009
2022	0.918	0.006	83,573	0.594	0.469	0.236	1.416	0.009
2023	0.878	0.006	63,780	0.578	0.441	0.245	1.508	0.011
2024	1.013	0.005	84,183	0.606	0.444	0.291	1.572	0.009
2025	1.025	0.006	89,684	0.611	0.469	0.269	1.506	0.008

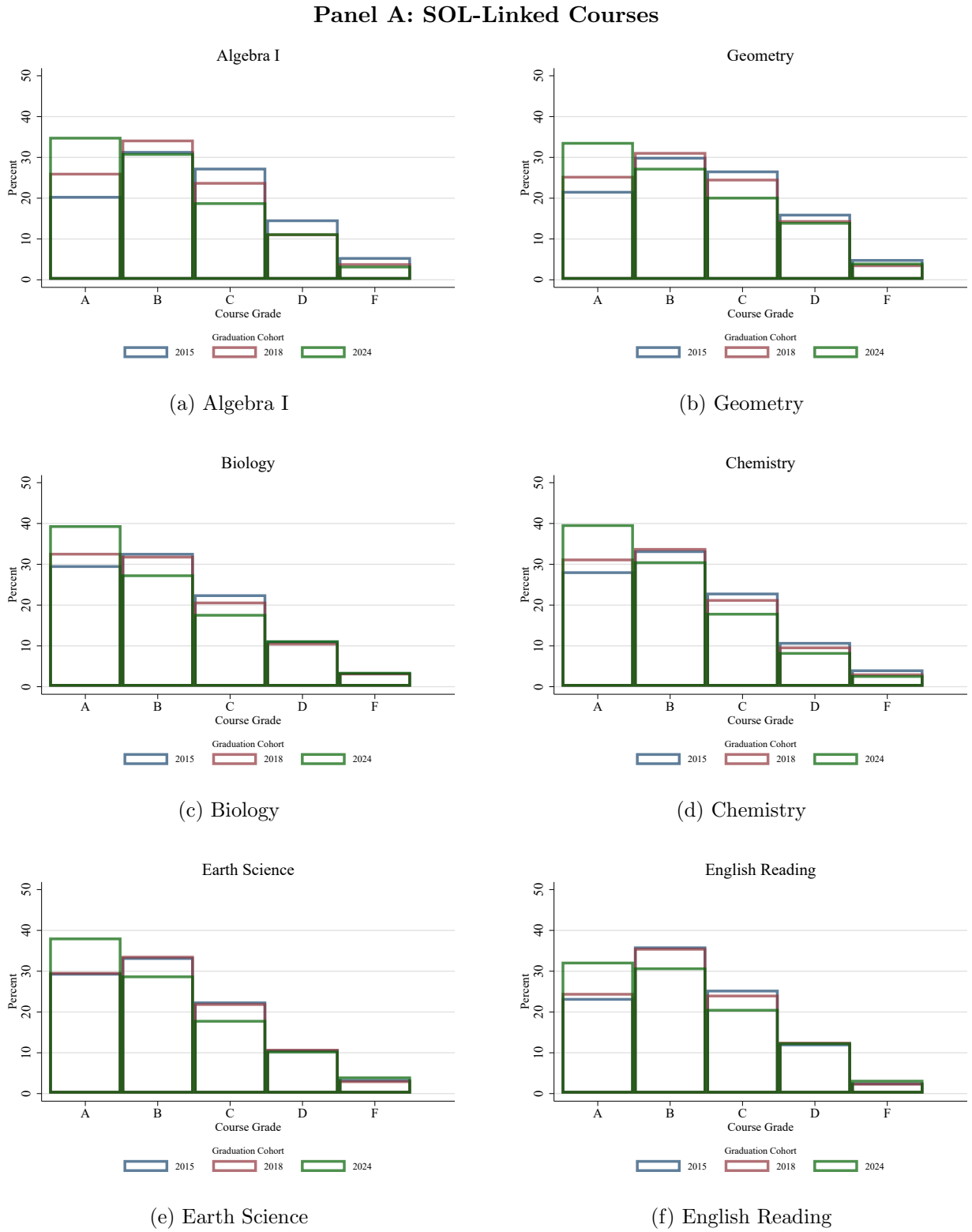
Notes: Dependent variable: cumulative HSGPA. Controls: 7 SOL scores (course-specific), 7 missing-SOL indicators, female, Black, Hispanic, Asian, economically disadvantaged. γ_{raw} = coefficient on school-demeaned student FE ($\tilde{\mu}_i$). Partial $R^2 = (R^2_{full} - R^2_{-\mu})/(1 - R^2_{-\mu})$, the incremental explanatory power of grade skill beyond SOL scores and demographics. γ_{EB} = coefficient from the same regression (same controls and sample) with the EB estimate $\tilde{\mu}_i^{EB} = r_i\tilde{\mu}_i$ in place of $\tilde{\mu}_i$; because $\tilde{\mu}_i$ contains estimation noise, γ_{raw} is attenuated toward zero, and using the EB estimate as the regressor corrects this attenuation (see Table 7 for r_i). Sample: students with ≥ 2 SOL-linked courses and non-missing HSGPA.

Table 9. Demographic Gradients in Within-School Grade Skill ($\tilde{\mu}_i$)

Cohort	Female	Black	Hispanic	Asian	Econ Disadv	R^2
2014	+0.097	−0.048	−0.046	+0.094	−0.097	0.080
2015	+0.098	−0.050	−0.049	+0.100	−0.102	0.087
2016	+0.097	−0.045	−0.051	+0.108	−0.106	0.089
2017	+0.106	−0.047	−0.050	+0.105	−0.113	0.100
2018	+0.113	−0.050	−0.051	+0.116	−0.111	0.104
2019	+0.118	−0.053	−0.056	+0.118	−0.117	0.109
2020	+0.126	−0.057	−0.054	+0.116	−0.117	0.112
2021	+0.124	−0.046	−0.047	+0.124	−0.125	0.105
2022	+0.132	−0.045	−0.049	+0.121	−0.133	0.101
2023	+0.134	−0.042	−0.048	+0.117	−0.139	0.094
2024	+0.133	−0.048	−0.062	+0.125	−0.141	0.101
2025	+0.115	−0.054	−0.067	+0.117	−0.143	0.101

Notes: OLS regression of $\tilde{\mu}_i$ on demographic indicators. All coefficients significant at $p < 0.001$. R^2 reflects the share of within-school grade-skill variance attributable to demographics. The female advantage grows from +0.10 in 2014 to +0.13 in 2024; over the same cohorts the economic disadvantage penalty grows from −0.10 to −0.14 and the Asian advantage from +0.09 to +0.13. The Black penalty shows no trend, ranging from −0.042 to −0.057; the Hispanic penalty ranges from −0.046 to −0.056 through 2023, then widens to −0.062 and −0.067 in 2024 and 2025.

Figures



(a) Algebra I

(b) Geometry

(c) Biology

(d) Chemistry

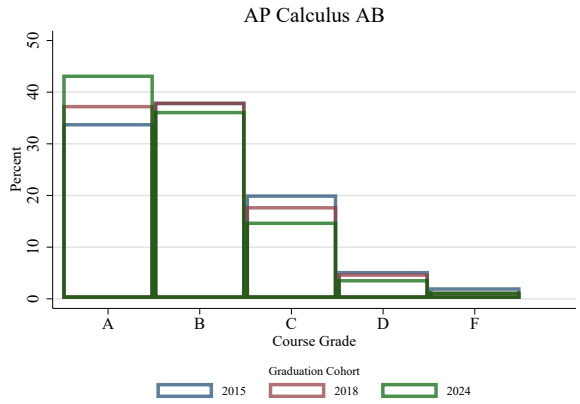
(e) Earth Science

(f) English Reading

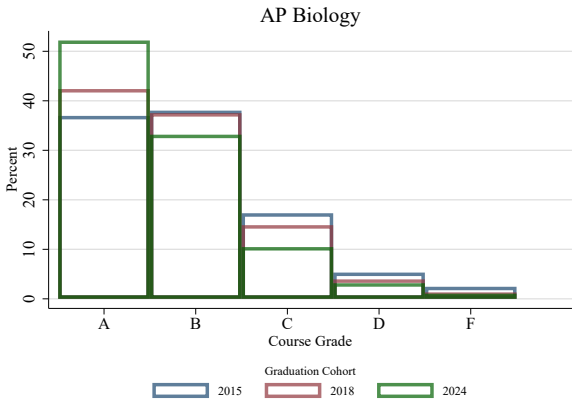
Figure 1. Distribution of course grades by graduation cohort.

Notes: Each panel compares the grade distribution across three graduation cohorts (2015, 2018, 2024). The dominant shift is from B to A grades, with D/F rates largely stable.
Source: A1_descriptive_trends.do.

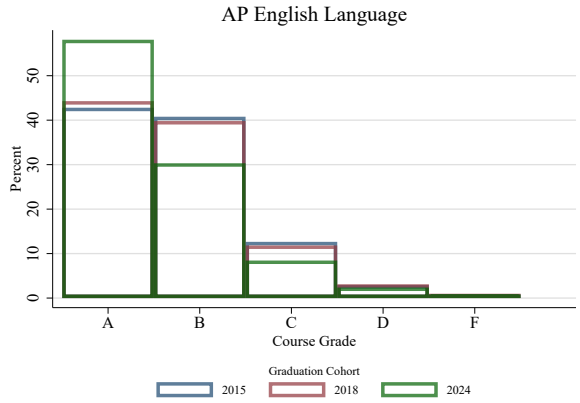
Panel B: AP Courses



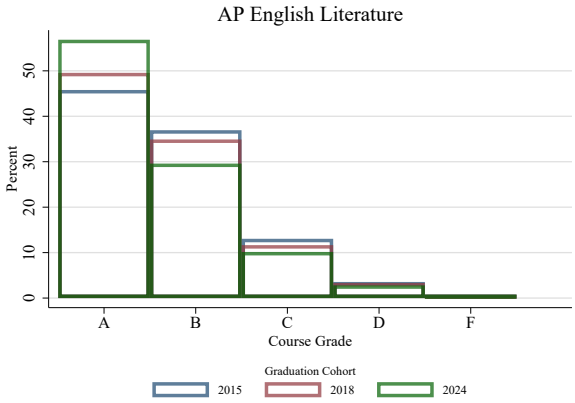
(a) AP Calculus AB



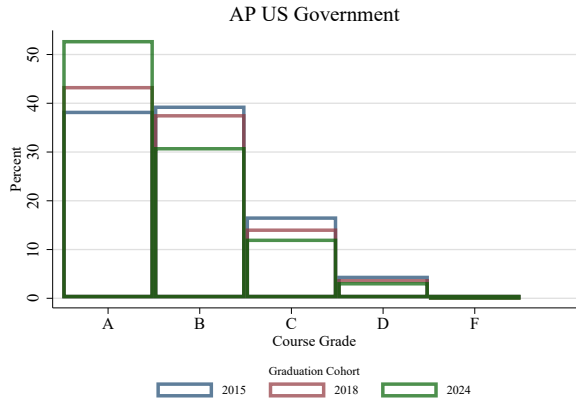
(b) AP Biology



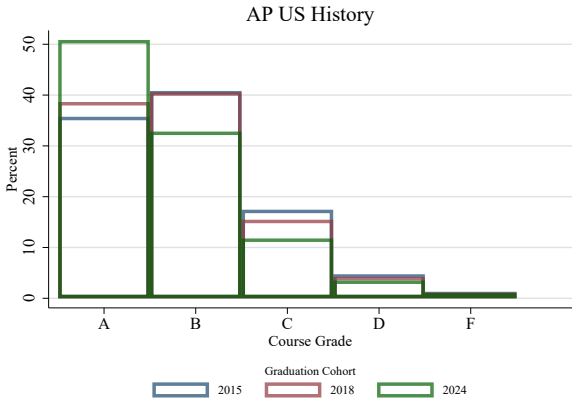
(c) AP English Language



(d) AP English Literature



(e) AP US Government



(f) AP US History

Figure 2. Distribution of AP course grades by graduation cohort.

Notes: Each panel compares the AP course grade distribution across three graduation cohorts (2015, 2018, 2024). As with SOL-linked courses, the dominant shift is from B to A grades.
Source: A1_descriptive_trends.do.

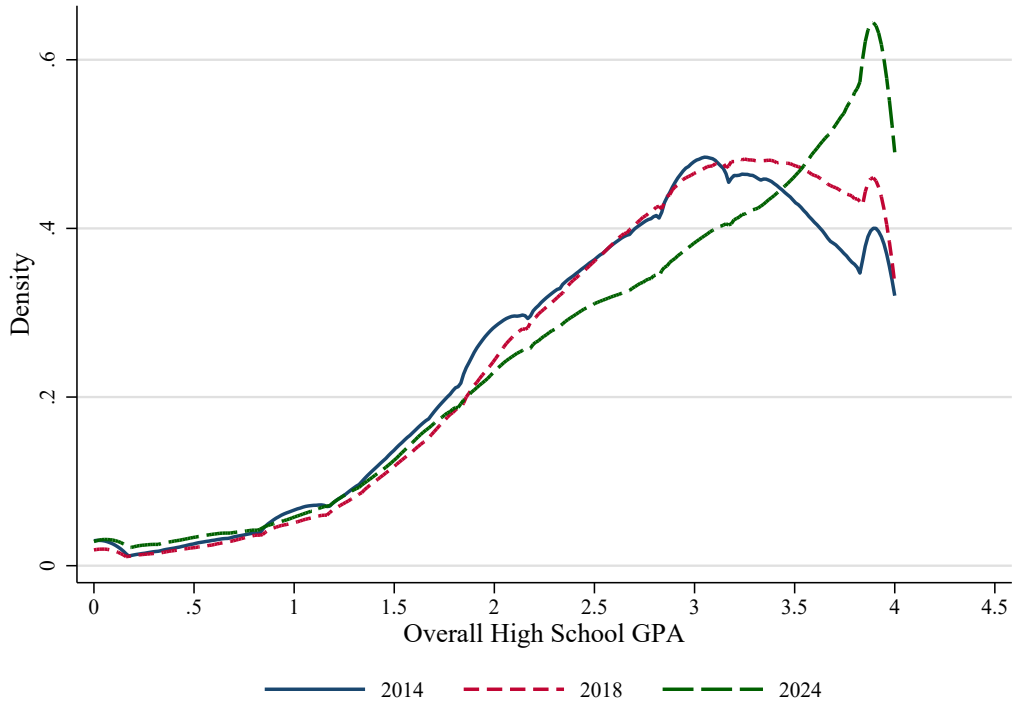
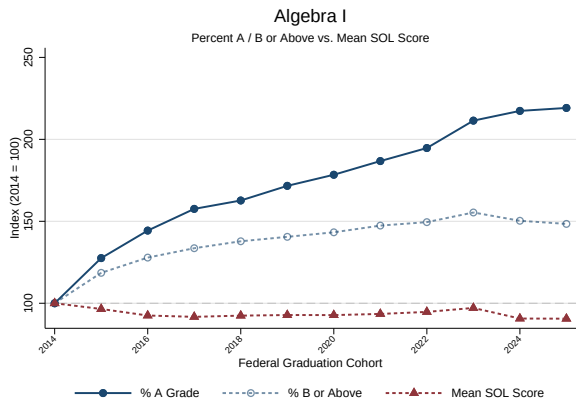
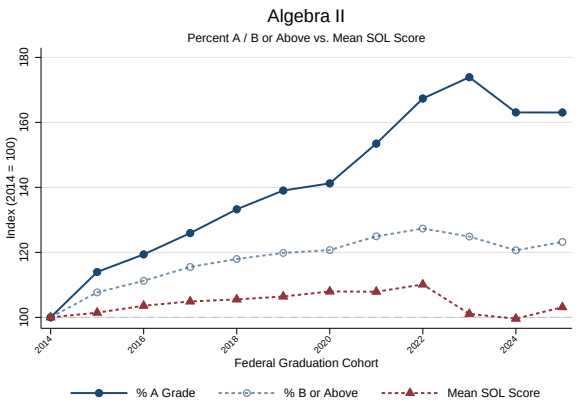


Figure 3. Kernel density of overall high school GPA, selected graduation cohorts.

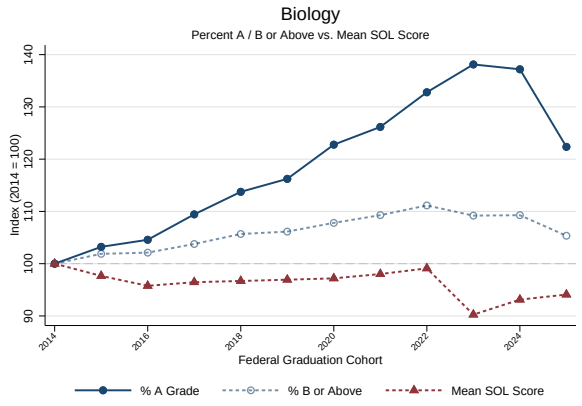
Notes: The rightward shift and increasing mass at the 4.0 ceiling illustrate both the mean increase and the top-end compression documented in Appendix Table D1.
Source: A1_descriptive_trends.do.



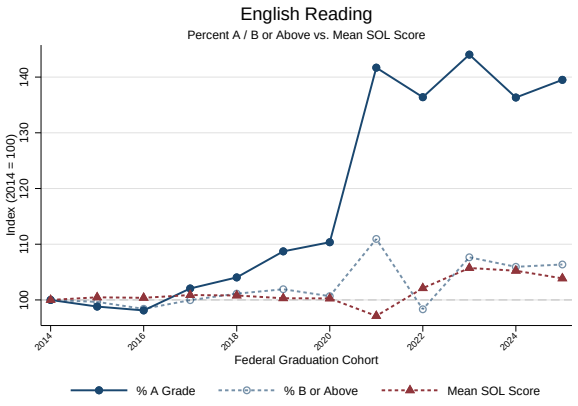
(a) Algebra I



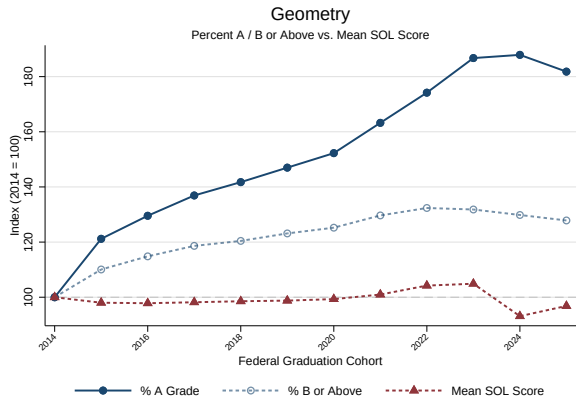
(b) Algebra II



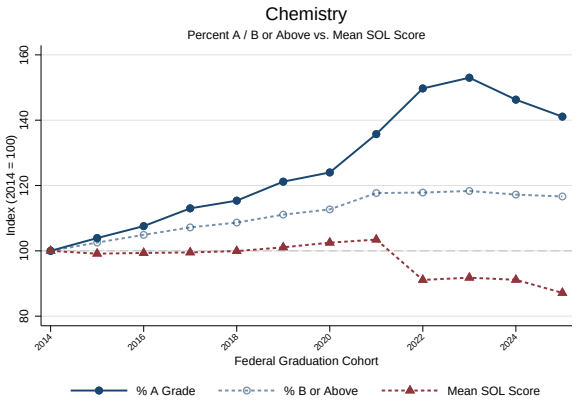
(c) Biology



(d) English Reading



(e) Geometry



(f) Chemistry

Figure 4. Course grades and SOL test scores over time, core SOL subjects.

Notes: Each panel plots the share of students earning an A (navy solid, circles) and the share earning B or above (navy dashed, open circles) alongside the mean SOL score (maroon dashed, triangles), all indexed to the first available graduation cohort = 100.
Source: grades_vs_scores_over_time.do.

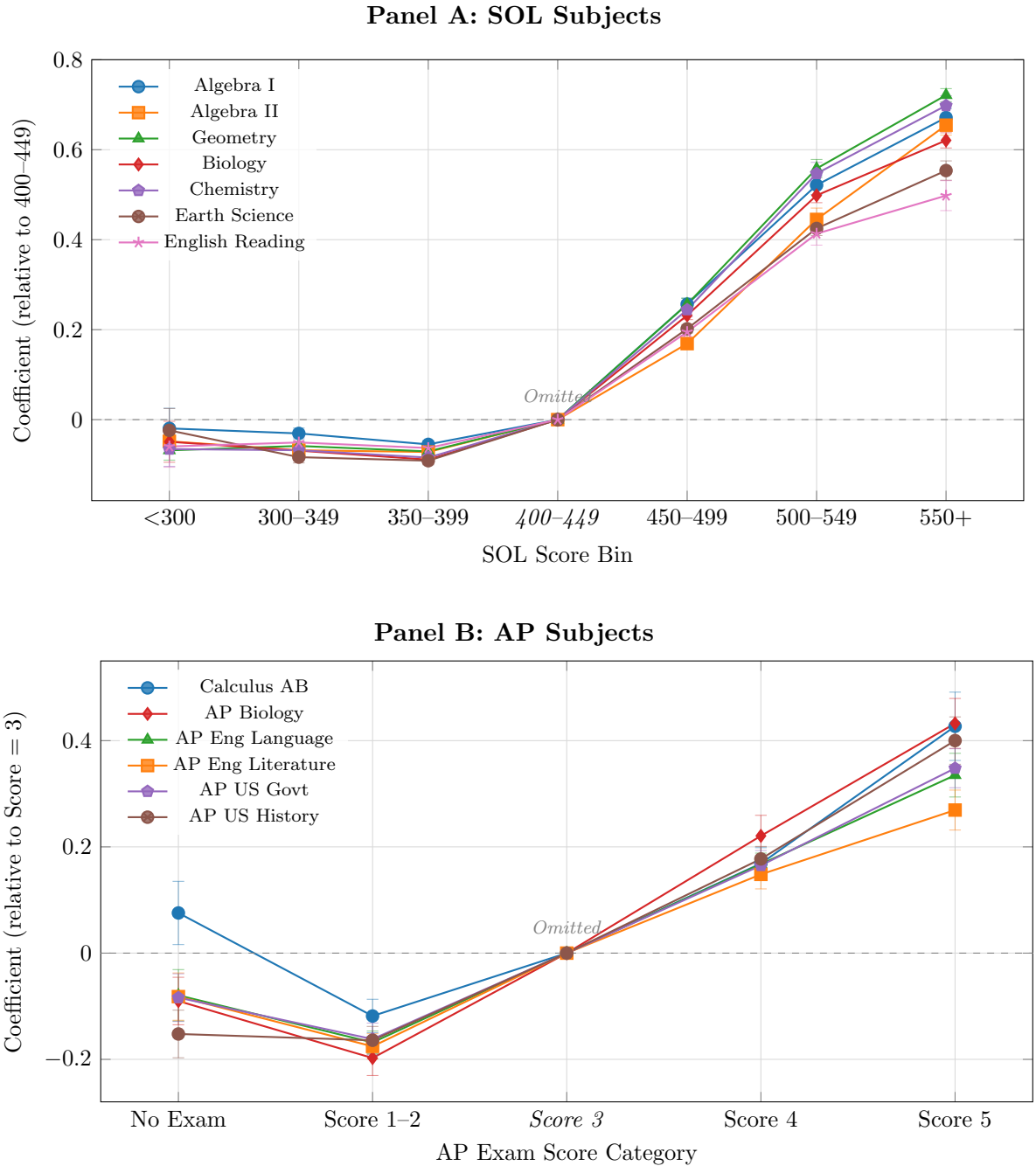
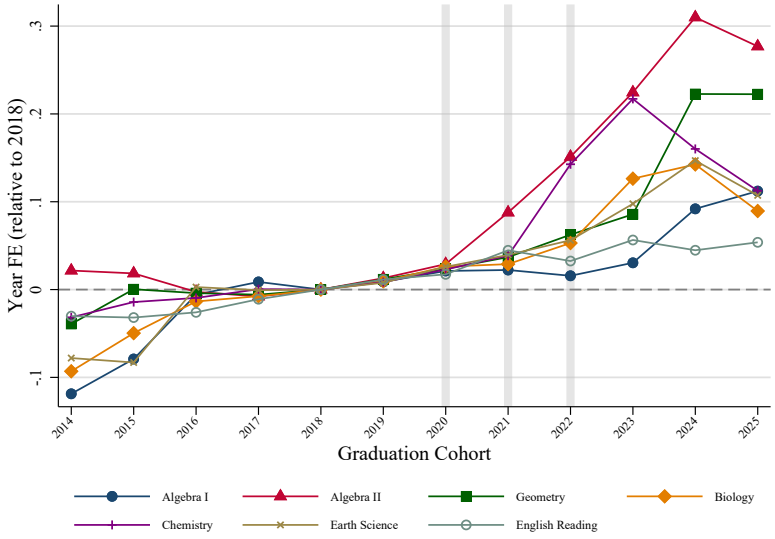


Figure 5. Score–grade gradient: test score bin coefficients from Specification 3, pre-COVID (2014–2019).

Notes: Panel A plots SOL score bin coefficients; Panel B plots AP exam score category coefficients. Each line is one subject. Omitted categories plotted at zero. Whiskers show 95% confidence intervals. The gradient is strikingly consistent across subjects within each domain.
Source: A2_workhorse_table.do.

Panel A: SOL Subjects



Panel B: AP Subjects

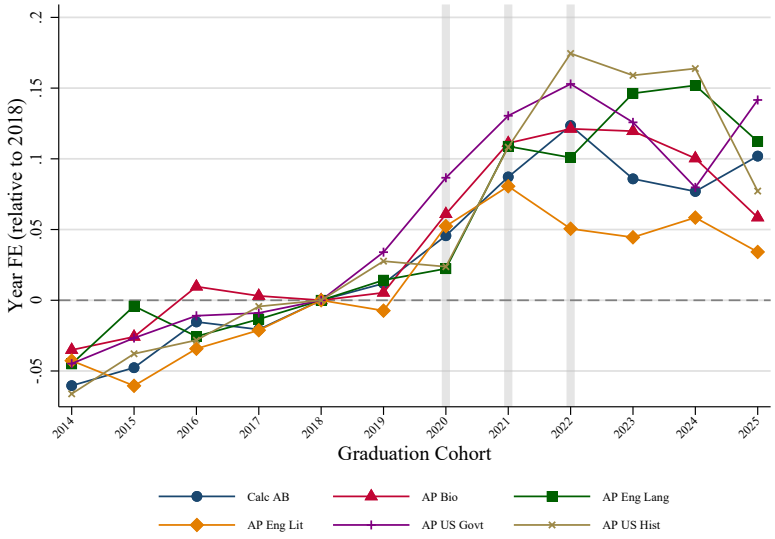


Figure 6. Year fixed effects on P(A), SOL and AP subjects.

Notes: Reference year: 2018. Shaded regions indicate COVID-affected cohorts (2020–2022). SE clustered at school level.
Source: A2_workhorse_table.do.

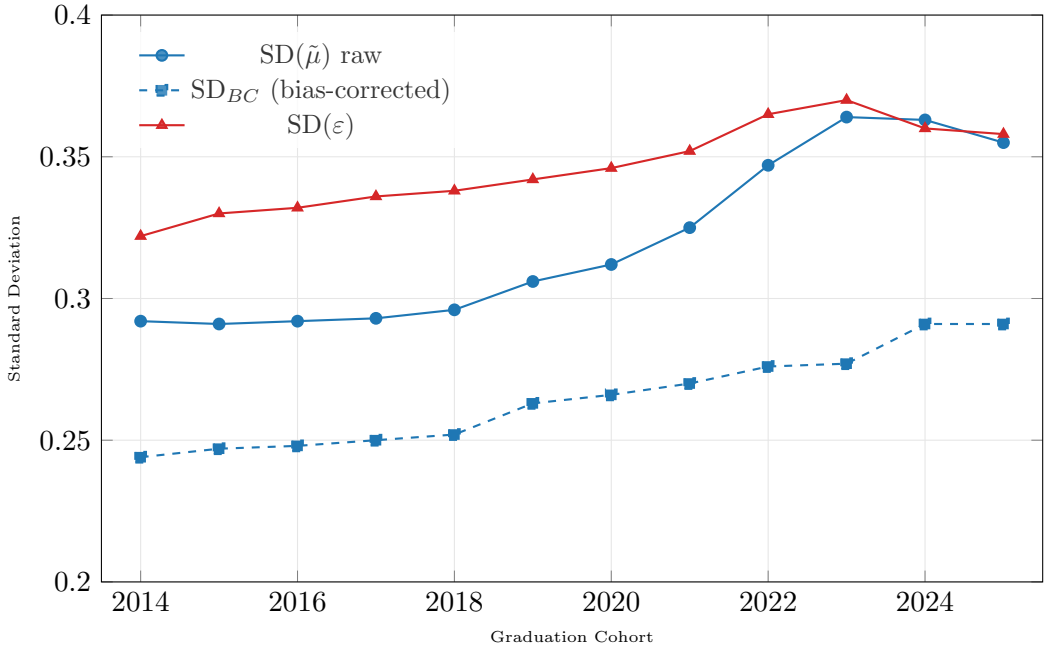


Figure 7. Standard deviation of grade skill and residual variance, by graduation cohort.

Notes: Blue solid line: raw SD of within-school grade skill ($\tilde{\mu}_i$). Blue dashed line: bias-corrected SD, $\text{SD}_{BC} = \sqrt{\text{Var}(\tilde{\mu}) - \text{mean}(\tilde{s}_i^2)}$, which subtracts each student's average estimation-noise variance (Table 7). Red line: SD of the idiosyncratic residual ($\hat{\sigma}_\varepsilon$, degrees of freedom net of student, course, and grade-level FEs). The gap between the raw and corrected blue lines represents estimation noise from finite courses per student. Preferred specification: binned SOL scores, full period.

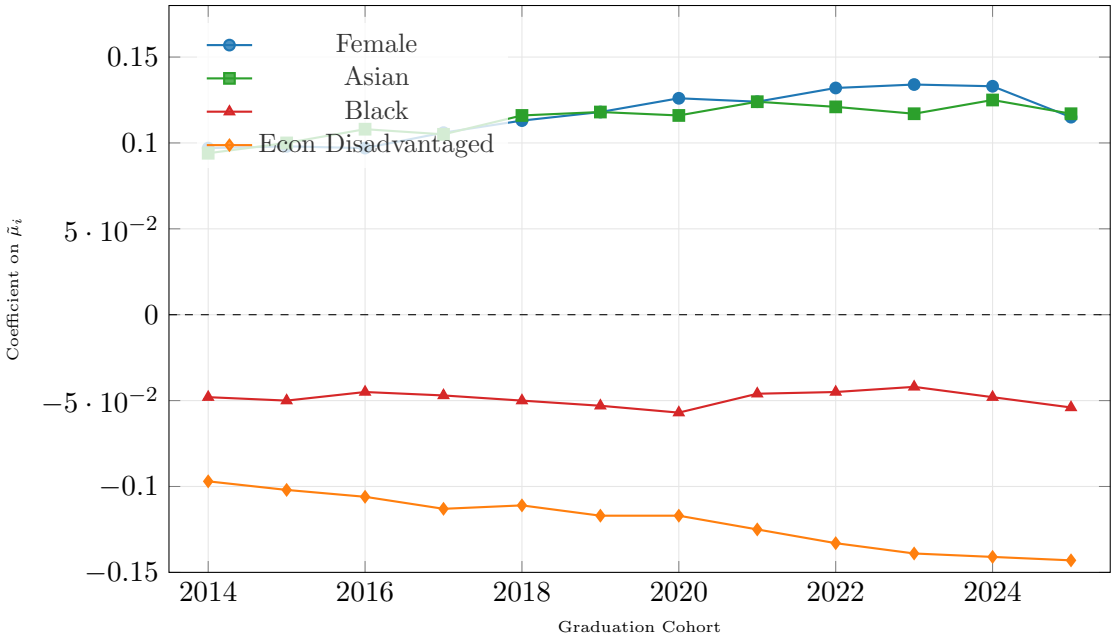


Figure 8. Demographic gradients in grade skill, by graduation cohort.

Notes: Coefficients from OLS regression of within-school grade skill ($\tilde{\mu}_i$) on demographic indicators, estimated separately by cohort. Positive values indicate higher grade skill; negative values indicate lower grade skill relative to White male students who are not economically disadvantaged. Hispanic coefficients (not shown for clarity; see Table 9) range from −0.046 to −0.056 through 2023, then widen to −0.062 and −0.067 in 2024 and 2025.

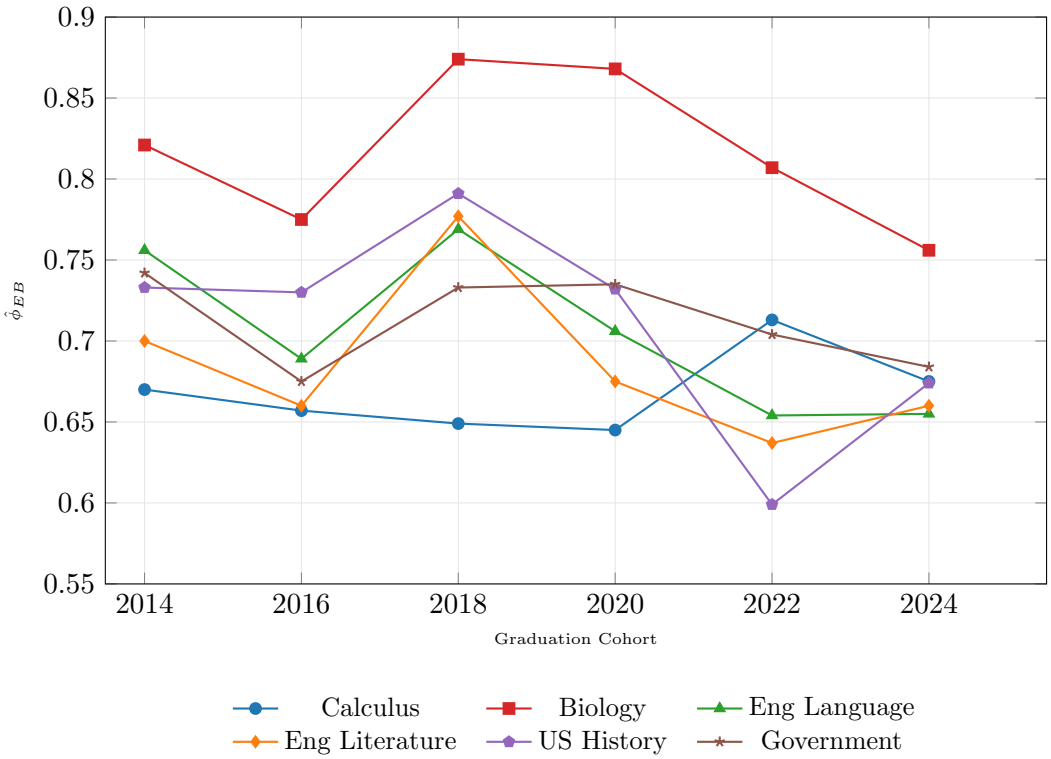


Figure 9. Grade skill coefficient ($\hat{\phi}_{EB}$) in AP course grade regressions, by subject and cohort.

Notes: Each point is $\hat{\phi}_{EB}$ from Equation 9, the coefficient on the Empirical Bayes estimate of SOL-derived grade skill ($\tilde{\mu}_i^{EB}$, Equation 7) in a regression of $\mathbf{1}$(AP course grade = A) on AP exam score dummies and demographics. Even-year cohorts shown. Appendix Table C5 reports $\hat{\phi}_{EB}$ and the raw coefficient $\hat{\phi}_{raw}$ with standard errors for Biology, English Literature, and U.S. Government.

A Test-Taking and Selection

Tables

Table A1. Test-Taking Rates by Graduation Cohort (%)

	SOL Subjects							AP Subjects						
Cohort	Alg I	Alg II	Geom	Bio	Chem	Earth	Eng	Calc	Bio	Eng L	Eng Lit	Gov	Hist	N
2014	91.2	68.9	87.0	91.0	63.1	75.1	86.5	7.5	5.7	15.4	11.2	14.1	15.4	83,497
2015	92.0	67.0	86.5	90.0	61.3	75.0	85.1	7.3	5.5	15.1	10.6	14.2	15.2	86,364
2016	90.1	66.7	86.2	88.6	60.6	73.5	83.4	7.5	5.5	15.4	10.0	13.5	14.6	91,698
2017	89.4	66.5	85.5	89.1	60.5	73.7	80.0	7.6	5.7	14.9	9.3	14.0	13.8	95,407
2018	89.8	66.7	85.0	88.4	60.2	74.4	77.6	7.6	6.1	14.4	8.4	13.2	13.7	101,802
2019	90.2	63.2	83.9	89.1	56.3	71.1	79.7	7.6	6.0	14.4	8.3	13.3	13.5	102,833
2020	89.9	49.9	78.9	87.2	34.7	67.0	82.6	7.0	5.7	14.1	7.1	12.0	13.5	103,083
2021	88.4	40.9	70.2	82.8	25.2	64.6	20.4	6.2	5.4	13.7	6.5	10.3	13.1	102,515
2022	83.2	14.1	43.9	49.5	0.9	54.5	68.4	6.5	5.9	12.6	6.8	11.1	11.8	103,945
2023	53.3	3.6	18.9	44.5	1.6	16.5	80.5	7.2	5.5	12.2	6.3	11.7	12.1	104,654
2024	55.1	14.4	28.6	85.0	3.8	26.8	79.7	6.8	5.0	12.7	6.1	10.6	11.5	105,376
2025	73.0	15.1	40.1	89.5	4.1	37.8	83.2	7.4	5.8	12.8	7.3	12.9	11.2	107,412

Percent of HS graduates with valid test score, by subject. SOL testing requirements relaxed post-2021 for Algebra II, Chemistry, Geometry, and Earth Science.

B Full-Period Workhorse Regressions

Tables

Table B1. P(A) Trend Coefficient, SOL Subjects, 2014–2025

	Algebra I	Algebra II	Geometry	Biology	Chemistry	Earth Science	English
Cohort trend ($\hat{\gamma}$)	0.0141***	0.0206***	0.0200***	0.0168***	0.0082***	0.0191***	0.0054***
	(0.0008)	(0.0011)	(0.0008)	(0.0008)	(0.0014)	(0.0012)	(0.0009)
+ Demographics	0.0139***	0.0206***	0.0197***	0.0168***	0.0085***	0.0189***	0.0074***
	(0.0008)	(0.0011)	(0.0008)	(0.0008)	(0.0014)	(0.0011)	(0.0008)
+ Absence rate	0.0140***	0.0205***	0.0194***	0.0169***	0.0088***	0.0183***	0.0086***
	(0.0008)	(0.0011)	(0.0008)	(0.0008)	(0.0014)	(0.0011)	(0.0009)
+ School FE	0.0143***	0.0201***	0.0192***	0.0172***	0.0109***	0.0171***	0.0087***
	(0.0008)	(0.0011)	(0.0009)	(0.0008)	(0.0013)	(0.0011)	(0.0008)
R^2 (spec 1)	0.248	0.258	0.288	0.225	0.263	0.172	0.118
R^2 (spec 2)	0.265	0.271	0.300	0.260	0.278	0.206	0.154
R^2 (spec 3)	0.270	0.276	0.304	0.268	0.283	0.212	0.171
R^2 (spec 4)	0.319	0.330	0.358	0.305	0.346	0.262	0.205
Mean P(A)	0.290	0.299	0.278	0.346	0.345	0.332	0.283
N	833,372	486,050	706,652	900,970	384,064	643,498	658,089

LPM. Outcome: P(A). Each cell reports $\hat{\gamma}$, the coefficient on a linear cohort trend. Each row adds controls cumulatively. SE clustered at school level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table B2. P(A) Trend Coefficient, AP Subjects, 2014–2025

	AP Calc	AP Bio	AP ELang	AP ELit	AP Gov	AP Hist
Cohort trend ($\hat{\gamma}$)	0.0144***	0.0156***	0.0182***	0.0102***	0.0178***	0.0188***
	(0.0017)	(0.0019)	(0.0021)	(0.0025)	(0.0022)	(0.0016)
+ Demographics	0.0146***	0.0162***	0.0187***	0.0116***	0.0186***	0.0194***
	(0.0017)	(0.0019)	(0.0021)	(0.0025)	(0.0022)	(0.0015)
+ Absence rate	0.0146***	0.0166***	0.0191***	0.0116***	0.0185***	0.0203***
	(0.0017)	(0.0020)	(0.0021)	(0.0026)	(0.0022)	(0.0016)
+ School FE	0.0164***	0.0152***	0.0189***	0.0130***	0.0188***	0.0209***
	(0.0018)	(0.0018)	(0.0022)	(0.0026)	(0.0021)	(0.0014)
R^2 (spec 1)	0.118	0.120	0.128	0.081	0.118	0.171
R^2 (spec 2)	0.128	0.134	0.145	0.112	0.143	0.184
R^2 (spec 3)	0.145	0.154	0.163	0.134	0.159	0.197
R^2 (spec 4)	0.208	0.262	0.246	0.202	0.240	0.290
Mean P(A)	0.408	0.468	0.501	0.517	0.475	0.439
N	126,087	90,634	175,791	136,124	178,991	190,572

LPM. Outcome: P(A). Each cell reports $\hat{\gamma}$, the coefficient on a linear cohort trend. Each row adds controls cumulatively. SE clustered at school level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

C Full Coefficient Tables

Tables

Table C1. Period-Specific Coefficients: 2014 vs. 2019 — Panel A: SOL Subjects

	Alg I		Alg II		Geom		Bio		Chem		Earth		Eng	
	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
SOL <300	-0.0166	-0.0263	-0.0393**	-0.0426***	-0.0714***	-0.0794***	-0.1244***	-0.0390	-0.0900**	-0.0351	-0.0385	-0.1473*	-0.0605	-0.0789*
	(0.0465)	(0.0386)	(0.0153)	(0.0143)	(0.0113)	(0.0154)	(0.0157)	(0.0432)	(0.0396)	(0.0354)	(0.0268)	(0.0799)	(0.0382)	(0.0444)
SOL 300–349	-0.0178	-0.0434***	-0.0750***	-0.0690***	-0.0493***	-0.0672***	-0.0211	-0.0787***	-0.0626***	-0.0838***	-0.0499***	-0.0898***	-0.0245*	-0.0577***
	(0.0156)	(0.0092)	(0.0079)	(0.0077)	(0.0081)	(0.0068)	(0.0138)	(0.0083)	(0.0096)	(0.0107)	(0.0131)	(0.0107)	(0.0132)	(0.0105)
SOL 350–399	-0.0122	-0.0732***	-0.0713***	-0.0712***	-0.0644***	-0.0823***	-0.0335***	-0.1075***	-0.0762***	-0.0960***	-0.0543***	-0.0975***	-0.0459***	-0.0775***
	(0.0103)	(0.0054)	(0.0060)	(0.0063)	(0.0056)	(0.0059)	(0.0064)	(0.0061)	(0.0071)	(0.0080)	(0.0112)	(0.0061)	(0.0071)	(0.0064)
SOL 450–499	0.1161***	0.3041***	0.1713***	0.1747***	0.1988***	0.2694***	0.1901***	0.2553***	0.2104***	0.2507***	0.1480***	0.2392***	0.1970***	0.1970***
	(0.0091)	(0.0090)	(0.0083)	(0.0068)	(0.0111)	(0.0082)	(0.0073)	(0.0064)	(0.0112)	(0.0101)	(0.0083)	(0.0073)	(0.0088)	(0.0068)
SOL 500–549	0.4087***	0.6249***	0.4706***	0.4347***	0.4256***	0.5941***	0.4359***	0.5188***	0.4988***	0.5164***	0.3730***	0.4847***	0.3971***	0.3957***
	(0.0192)	(0.0117)	(0.0183)	(0.0119)	(0.0231)	(0.0133)	(0.0120)	(0.0101)	(0.0176)	(0.0138)	(0.0134)	(0.0114)	(0.0173)	(0.0156)
SOL 550+	0.6448***	0.7531***	0.6677***	0.6545***	0.6390***	0.7353***	0.5949***	0.6411***	0.6736***	0.6803***	0.5672***	0.5996***	0.4444***	0.4368***
	(0.0264)	(0.0086)	(0.0183)	(0.0123)	(0.0212)	(0.0102)	(0.0128)	(0.0112)	(0.0159)	(0.0129)	(0.0176)	(0.0176)	(0.0489)	(0.0265)
Female	0.0486***	0.0877***	0.0743***	0.0884***	0.0514***	0.0794***	0.1338***	0.1559***	0.0912***	0.1118***	0.1165***	0.1341***	0.1208***	0.1404***
	(0.0040)	(0.0036)	(0.0040)	(0.0036)	(0.0036)	(0.0033)	(0.0048)	(0.0044)	(0.0043)	(0.0041)	(0.0045)	(0.0047)	(0.0054)	(0.0048)
Black	-0.0331*	-0.0611***	-0.0722***	-0.0813***	-0.0502***	-0.0613***	-0.0572***	-0.0456***	-0.0800***	-0.0817***	-0.0292*	-0.0600***	-0.0751***	-0.0775***
	(0.0170)	(0.0067)	(0.0077)	(0.0082)	(0.0088)	(0.0072)	(0.0119)	(0.0081)	(0.0103)	(0.0094)	(0.0170)	(0.0090)	(0.0118)	(0.0084)
Hispanic	-0.0152	-0.0300***	-0.0551***	-0.0721***	-0.0380***	-0.0564***	-0.0585***	-0.0401***	-0.0851***	-0.0620***	-0.0236**	-0.0286***	-0.0522***	-0.0542***
	(0.0099)	(0.0072)	(0.0078)	(0.0079)	(0.0080)	(0.0069)	(0.0100)	(0.0081)	(0.0096)	(0.0102)	(0.0104)	(0.0096)	(0.0094)	(0.0086)
Asian	0.0903***	0.0444***	-0.0219*	-0.0223**	0.0066	0.0021	0.0760***	0.0740***	-0.0164	0.0026	0.0994***	0.1061***	0.0821***	0.1142***
	(0.0182)	(0.0100)	(0.0127)	(0.0113)	(0.0121)	(0.0113)	(0.0147)	(0.0116)	(0.0161)	(0.0157)	(0.0381)	(0.0143)	(0.0241)	(0.0222)
Econ Disadv	-0.0081	-0.0192***	-0.0040	-0.0016	0.0044	-0.0110*	-0.0385***	-0.0447***	-0.0025	-0.0139**	-0.0654***	-0.0742***	-0.0380***	-0.0578***
	(0.0066)	(0.0059)	(0.0062)	(0.0066)	(0.0055)	(0.0060)	(0.0071)	(0.0064)	(0.0075)	(0.0071)	(0.0102)	(0.0071)	(0.0077)	(0.0068)
Absence Rate	-0.2313***	-0.4312***	-0.7195***	-0.7398***	-0.2996***	-0.4524***	-0.6282***	-0.7119***	-0.6243***	-0.8841***	-0.4209***	-0.5452***	-0.7553***	-0.8439***
	(0.0692)	(0.0336)	(0.0522)	(0.0478)	(0.0346)	(0.0406)	(0.0480)	(0.0362)	(0.0611)	(0.0625)	(0.0675)	(0.0433)	(0.0488)	(0.0401)
R^2	0.165	0.298	0.265	0.275	0.230	0.326	0.227	0.288	0.260	0.272	0.184	0.217	0.145	0.164
N	32,416	82,908	54,915	62,076	52,611	80,725	61,749	87,167	48,092	54,506	52,030	66,747	54,509	60,517
Mean P(A)	0.167	0.284	0.245	0.314	0.187	0.276	0.296	0.341	0.276	0.346	0.281	0.310	0.244	0.278

Separate OLS regressions for early and late cohorts. Covariates: SOL bin / AP score dummies + demographics + absence rate. SE clustered at school level. Algebra II uses 2015 as early anchor. * $p < .10$, ** $p < .05$, *** $p < .01$.

Preliminary Draft

Table C2. Period-Specific Coefficients: 2014 vs. 2019 — Panel B: AP Subjects

	Calc AB		AP Bio		AP Eng L		AP Eng Lit		AP Govt		AP Hist	
	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
No AP Exam	0.1091***	0.0771**	-0.1160***	-0.0927***	-0.1531***	-0.0938***	-0.0912**	-0.0475*	-0.0667**	-0.1331***	-0.1373***	-0.1243***
	(0.0353)	(0.0353)	(0.0294)	(0.0324)	(0.0367)	(0.0306)	(0.0394)	(0.0267)	(0.0305)	(0.0282)	(0.0296)	(0.0297)
AP Score 1–2	-0.0549**	-0.1353***	-0.2191***	-0.1962***	-0.2084***	-0.1644***	-0.1234***	-0.2051***	-0.1101***	-0.2033***	-0.1454***	-0.1418***
	(0.0240)	(0.0216)	(0.0227)	(0.0230)	(0.0234)	(0.0145)	(0.0237)	(0.0178)	(0.0277)	(0.0217)	(0.0224)	(0.0189)
AP Score 4	0.1675***	0.1901***	0.2586***	0.2330***	0.1845***	0.1727***	0.1417***	0.1609***	0.1663***	0.1503***	0.2054***	0.1814***
	(0.0249)	(0.0245)	(0.0268)	(0.0278)	(0.0275)	(0.0148)	(0.0261)	(0.0177)	(0.0209)	(0.0173)	(0.0185)	(0.0165)
AP Score 5	0.4164***	0.4375***	0.4288***	0.4352***	0.3743***	0.3372***	0.2791***	0.2618***	0.3685***	0.2811***	0.4880***	0.3828***
	(0.0323)	(0.0405)	(0.0389)	(0.0297)	(0.0328)	(0.0208)	(0.0342)	(0.0251)	(0.0276)	(0.0212)	(0.0380)	(0.0242)
Female	0.0723***	0.0910***	0.0889***	0.0922***	0.1086***	0.1081***	0.1008***	0.1299***	0.0954***	0.1280***	0.0693***	0.1022***
	(0.0112)	(0.0098)	(0.0121)	(0.0123)	(0.0179)	(0.0094)	(0.0104)	(0.0100)	(0.0082)	(0.0093)	(0.0091)	(0.0084)
Black	-0.1172***	-0.1045***	-0.0832***	-0.0888***	-0.0714**	-0.1081***	-0.1319***	-0.1374***	-0.1120***	-0.1323***	-0.0639***	-0.0851***
	(0.0266)	(0.0246)	(0.0220)	(0.0261)	(0.0341)	(0.0188)	(0.0238)	(0.0188)	(0.0189)	(0.0194)	(0.0195)	(0.0204)
Hispanic	-0.0703***	-0.0560***	-0.0698***	-0.0967***	-0.0150	-0.0616***	-0.0376	-0.1155***	-0.0787***	-0.1025***	-0.0665***	-0.0860***
	(0.0232)	(0.0199)	(0.0237)	(0.0206)	(0.0324)	(0.0150)	(0.0256)	(0.0167)	(0.0199)	(0.0146)	(0.0171)	(0.0154)
Asian	-0.0392**	-0.0175	-0.0277	-0.0413	0.0408	-0.0068	-0.0169	0.0027	0.0037	-0.0006	-0.0242	0.0197
	(0.0171)	(0.0179)	(0.0258)	(0.0274)	(0.0260)	(0.0219)	(0.0270)	(0.0170)	(0.0225)	(0.0244)	(0.0257)	(0.0221)
Econ Disadv	0.0026	-0.0404***	-0.0296**	-0.0256	-0.0228	-0.0544***	-0.0711***	-0.0797***	-0.0386***	-0.0502***	0.0096	-0.0312**
	(0.0162)	(0.0155)	(0.0124)	(0.0157)	(0.0193)	(0.0144)	(0.0164)	(0.0136)	(0.0150)	(0.0144)	(0.0143)	(0.0139)
Absence Rate	-1.4663***	-1.6085***	-1.4579***	-1.5388***	-1.0446***	-1.5218***	-1.0422***	-1.8224***	-0.9526***	-1.2884***	-1.0224***	-1.5137***
	(0.2380)	(0.1768)	(0.1852)	(0.1466)	(0.2083)	(0.1293)	(0.2560)	(0.1289)	(0.1928)	(0.1531)	(0.1617)	(0.1526)
R^2	0.139	0.159	0.185	0.162	0.179	0.164	0.106	0.152	0.144	0.173	0.214	0.184
N	8,758	11,455	6,556	8,475	5,548	16,983	12,332	12,994	13,862	17,218	13,938	16,550
Mean P(A)	0.346	0.390	0.362	0.437	0.388	0.475	0.463	0.493	0.382	0.473	0.340	0.428

Separate OLS regressions for early and late cohorts. Covariates: SOL bin / AP score dummies + demographics + absence rate. SE clustered at school level. Algebra II uses 2015 as early anchor. * $p < .10$, ** $p < .05$, *** $p < .01$.

Preliminary Draft

Table C3. Period-Specific Coefficients: 2014 vs. 2024 — Panel A: SOL Subjects

	Alg I		Alg II		Geom		Bio		Chem		Earth		Eng	
	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
SOL <300	-0.0166	-0.0535	-0.0393**	-0.2619***	-0.0714***	-0.2706***	-0.1244***	-0.1787***	-0.0900**	-0.3217***	-0.0385	-0.2992***	-0.0605	-0.0777***
	(0.0465)	(0.1079)	(0.0153)	(0.0458)	(0.0113)	(0.0281)	(0.0157)	(0.0314)	(0.0396)	(0.0236)	(0.0268)	(0.0319)	(0.0382)	(0.0226)
SOL 300–349	-0.0178	-0.1160***	-0.0750***	-0.2199***	-0.0493***	-0.2266***	-0.0211	-0.1901***	-0.0626***	-0.2355***	-0.0499***	-0.1723***	-0.0245*	-0.0701***
	(0.0156)	(0.0152)	(0.0079)	(0.0185)	(0.0081)	(0.0132)	(0.0138)	(0.0087)	(0.0096)	(0.0237)	(0.0131)	(0.0152)	(0.0132)	(0.0108)
SOL 350–399	-0.0122	-0.1267***	-0.0713***	-0.1734***	-0.0644***	-0.2055***	-0.0335***	-0.1665***	-0.0762***	-0.2248***	-0.0543***	-0.1331***	-0.0459***	-0.0773***
	(0.0103)	(0.0070)	(0.0060)	(0.0133)	(0.0056)	(0.0097)	(0.0064)	(0.0058)	(0.0071)	(0.0200)	(0.0112)	(0.0087)	(0.0071)	(0.0085)
SOL 450–499	0.1161***	0.2639***	0.1713***	0.2290***	0.1988***	0.2572***	0.1901***	0.2528***	0.2104***	0.2968***	0.1480***	0.2053***	0.1970***	0.1355***
	(0.0091)	(0.0107)	(0.0083)	(0.0127)	(0.0111)	(0.0132)	(0.0073)	(0.0056)	(0.0112)	(0.0343)	(0.0083)	(0.0087)	(0.0088)	(0.0057)
SOL 500–549	0.4087***	0.4980***	0.4706***	0.4001***	0.4256***	0.4316***	0.4359***	0.4096***	0.4988***	0.4561***	0.3730***	0.3533***	0.3971***	0.2980***
	(0.0192)	(0.0153)	(0.0183)	(0.0186)	(0.0231)	(0.0175)	(0.0120)	(0.0084)	(0.0176)	(0.0375)	(0.0134)	(0.0175)	(0.0173)	(0.0077)
SOL 550+	0.6448***	0.5968***	0.6677***	0.4340***	0.6390***	0.4956***	0.5949***	0.4836***	0.6736***	0.5380***	0.5672***	0.4586***	0.4444***	0.3961***
	(0.0264)	(0.0122)	(0.0183)	(0.0204)	(0.0212)	(0.0236)	(0.0128)	(0.0108)	(0.0159)	(0.0363)	(0.0176)	(0.0337)	(0.0489)	(0.0131)
Female	0.0486***	0.0889***	0.0743***	0.1038***	0.0514***	0.1134***	0.1338***	0.1490***	0.0912***	0.0603***	0.1165***	0.1341***	0.1208***	0.1349***
	(0.0040)	(0.0040)	(0.0040)	(0.0079)	(0.0036)	(0.0050)	(0.0048)	(0.0034)	(0.0043)	(0.0117)	(0.0045)	(0.0056)	(0.0054)	(0.0043)
Black	-0.0331*	-0.0471***	-0.0722***	-0.0995***	-0.0502***	-0.0378***	-0.0572***	-0.0311***	-0.0800***	-0.0517***	-0.0292*	-0.0448***	-0.0751***	-0.0687***
	(0.0170)	(0.0087)	(0.0077)	(0.0149)	(0.0088)	(0.0112)	(0.0119)	(0.0085)	(0.0103)	(0.0185)	(0.0170)	(0.0125)	(0.0118)	(0.0091)
Hispanic	-0.0152	-0.0112	-0.0551***	-0.0591***	-0.0380***	-0.0397***	-0.0585***	-0.0432***	-0.0851***	-0.0339*	-0.0236**	-0.0132	-0.0522***	-0.0687***
	(0.0099)	(0.0081)	(0.0078)	(0.0135)	(0.0080)	(0.0105)	(0.0100)	(0.0075)	(0.0096)	(0.0183)	(0.0104)	(0.0131)	(0.0094)	(0.0084)
Asian	0.0903***	0.0805***	-0.0219*	0.0396**	0.0066	0.0645***	0.0760***	0.1087***	-0.0164	-0.0077	0.0994***	0.1440***	0.0821***	0.0804***
	(0.0182)	(0.0134)	(0.0127)	(0.0158)	(0.0121)	(0.0132)	(0.0147)	(0.0095)	(0.0161)	(0.0180)	(0.0381)	(0.0174)	(0.0241)	(0.0138)
Econ Disadv	-0.0081	-0.0715***	-0.0040	-0.0617***	0.0044	-0.0552***	-0.0385***	-0.0810***	-0.0025	-0.0310**	-0.0654***	-0.1342***	-0.0380***	-0.0734***
	(0.0066)	(0.0075)	(0.0062)	(0.0129)	(0.0055)	(0.0088)	(0.0071)	(0.0065)	(0.0075)	(0.0155)	(0.0102)	(0.0108)	(0.0077)	(0.0071)
Absence Rate	-0.2313***	-0.6355***	-0.7195***	-1.1967***	-0.2996***	-0.9724***	-0.6282***	-0.8943***	-0.6243***	-0.6084***	-0.4209***	-0.7448***	-0.7553***	-0.9766***
	(0.0692)	(0.0386)	(0.0522)	(0.0985)	(0.0346)	(0.0601)	(0.0480)	(0.0382)	(0.0611)	(0.1013)	(0.0675)	(0.0517)	(0.0488)	(0.0369)
R^2	0.165	0.281	0.265	0.238	0.230	0.215	0.227	0.274	0.260	0.331	0.184	0.189	0.145	0.173
N	32,416	55,047	54,915	14,299	52,611	29,006	61,749	86,379	48,092	3,839	52,030	27,346	54,509	60,705
Mean P(A)	0.167	0.338	0.245	0.509	0.187	0.379	0.296	0.411	0.276	0.310	0.281	0.374	0.244	0.354

Separate OLS regressions for early and late cohorts. Covariates: SOL bin / AP score dummies + demographics + absence rate. SE clustered at school level. Algebra II uses 2015 as early anchor. * $p < .10$, ** $p < .05$, *** $p < .01$.

Preliminary Draft

Table C4. Period-Specific Coefficients: 2014 vs. 2024 — Panel B: AP Subjects

	Calc AB		AP Bio		AP Eng L		AP Eng Lit		AP Govt		AP Hist	
	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
No AP Exam	0.1091***	0.0265	-0.1160***	-0.0356	-0.1531***	-0.1265***	-0.0912**	0.0285	-0.0667**	0.0438*	-0.1373***	-0.1758***
	(0.0353)	(0.0308)	(0.0294)	(0.0300)	(0.0367)	(0.0272)	(0.0394)	(0.0226)	(0.0305)	(0.0240)	(0.0296)	(0.0307)
AP Score 1–2	-0.0549**	-0.1104***	-0.2191***	-0.1676***	-0.2084***	-0.1762***	-0.1234***	-0.1205***	-0.1101***	-0.1125***	-0.1454***	-0.1972***
	(0.0240)	(0.0284)	(0.0227)	(0.0282)	(0.0234)	(0.0184)	(0.0237)	(0.0240)	(0.0277)	(0.0219)	(0.0224)	(0.0217)
AP Score 4	0.1675***	0.1611***	0.2586***	0.1708***	0.1845***	0.1033***	0.1417***	0.1283***	0.1663***	0.1247***	0.2054***	0.1615***
	(0.0249)	(0.0250)	(0.0268)	(0.0266)	(0.0275)	(0.0144)	(0.0261)	(0.0233)	(0.0209)	(0.0207)	(0.0185)	(0.0161)
AP Score 5	0.4164***	0.3972***	0.4288***	0.3373***	0.3743***	0.2251***	0.2791***	0.2757***	0.3685***	0.3148***	0.4880***	0.2875***
	(0.0323)	(0.0400)	(0.0389)	(0.0315)	(0.0328)	(0.0174)	(0.0342)	(0.0241)	(0.0276)	(0.0243)	(0.0380)	(0.0262)
Female	0.0723***	0.0593***	0.0889***	0.0636***	0.1086***	0.0940***	0.1008***	0.1293***	0.0954***	0.1170***	0.0693***	0.0743***
	(0.0112)	(0.0108)	(0.0121)	(0.0134)	(0.0179)	(0.0096)	(0.0104)	(0.0102)	(0.0082)	(0.0090)	(0.0091)	(0.0083)
Black	-0.1172***	-0.0885***	-0.0832***	-0.0963***	-0.0714**	-0.0723***	-0.1319***	-0.1415***	-0.1120***	-0.1121***	-0.0639***	-0.1032***
	(0.0266)	(0.0251)	(0.0220)	(0.0291)	(0.0341)	(0.0208)	(0.0238)	(0.0231)	(0.0189)	(0.0218)	(0.0195)	(0.0194)
Hispanic	-0.0703***	-0.0854***	-0.0698***	-0.0954***	-0.0150	-0.0582***	-0.0376	-0.1124***	-0.0787***	-0.1137***	-0.0665***	-0.0820***
	(0.0232)	(0.0175)	(0.0237)	(0.0236)	(0.0324)	(0.0161)	(0.0256)	(0.0178)	(0.0199)	(0.0157)	(0.0171)	(0.0142)
Asian	-0.0392**	-0.0548***	-0.0277	-0.0495**	0.0408	-0.0306	-0.0169	-0.0095	0.0037	0.0366*	-0.0242	0.0334
	(0.0171)	(0.0167)	(0.0258)	(0.0236)	(0.0260)	(0.0250)	(0.0270)	(0.0193)	(0.0225)	(0.0213)	(0.0257)	(0.0228)
Econ Disadv	0.0026	-0.0181	-0.0296**	-0.0302	-0.0228	-0.0579***	-0.0711***	-0.0473***	-0.0386***	-0.0545***	0.0096	-0.0470***
	(0.0162)	(0.0139)	(0.0124)	(0.0222)	(0.0193)	(0.0132)	(0.0164)	(0.0139)	(0.0150)	(0.0154)	(0.0143)	(0.0138)
Absence Rate	-1.4663***	-1.6968***	-1.4579***	-1.8709***	-1.0446***	-1.8469***	-1.0422***	-2.0163***	-0.9526***	-1.5689***	-1.0224***	-1.4124***
	(0.2380)	(0.1547)	(0.1852)	(0.1480)	(0.2083)	(0.0936)	(0.2560)	(0.1065)	(0.1928)	(0.1336)	(0.1617)	(0.1163)
R^2	0.139	0.128	0.185	0.139	0.179	0.143	0.106	0.130	0.144	0.136	0.214	0.184
N	8,758	11,353	6,556	7,995	5,548	16,067	12,332	12,229	13,862	16,739	13,938	15,195
Mean P(A)	0.346	0.434	0.362	0.522	0.388	0.581	0.463	0.568	0.382	0.530	0.340	0.508

Separate OLS regressions for early and late cohorts. Covariates: SOL bin / AP score dummies + demographics + absence rate. SE clustered at school level. Algebra II uses 2015 as early anchor. * $p < .10$, ** $p < .05$, *** $p < .01$.

Preliminary Draft

Table C5. Validation: Grade Skill Predicting P(A) in AP Courses

Subject	Cohort	$\hat{\phi}_{raw}$	SE	$\hat{\phi}_{EB}$	SE	Mean P(A)	N	R^2
Biology	2014	0.532	0.020	0.821	0.030	0.382	4,274	0.317
	2016	0.563	0.021	0.775	0.029	0.431	4,623	0.279
	2018	0.632	0.020	0.874	0.027	0.449	5,502	0.308
	2020	0.637	0.020	0.868	0.027	0.521	5,526	0.287
	2022	0.518	0.020	0.807	0.030	0.590	5,612	0.199
	2024	0.465	0.022	0.756	0.034	0.568	4,451	0.236
Eng Literature	2014	0.456	0.016	0.700	0.025	0.491	7,787	0.187
	2016	0.472	0.016	0.660	0.022	0.507	8,033	0.201
	2018	0.555	0.016	0.777	0.023	0.526	7,598	0.287
	2020	0.494	0.017	0.675	0.024	0.585	6,712	0.225
	2022	0.405	0.017	0.637	0.027	0.624	6,375	0.185
	2024	0.405	0.018	0.660	0.029	0.598	5,297	0.212
US Government	2014	0.489	0.014	0.742	0.022	0.405	8,772	0.244
	2016	0.493	0.014	0.675	0.019	0.453	10,154	0.290
	2018	0.536	0.014	0.733	0.019	0.474	10,702	0.293
	2020	0.542	0.014	0.735	0.019	0.590	10,362	0.250
	2022	0.457	0.013	0.704	0.021	0.590	9,724	0.243
	2024	0.422	0.014	0.684	0.022	0.573	8,388	0.246

Notes: Dependent variable: $\mathbf{1}$(earned A in AP course). Controls: AP exam score dummies (ref. = 3), female, Black, Hispanic, Asian, economically disadvantaged. ϕ_{raw} = coefficient on school-demeaned student FE ($\tilde{\mu}_i$). ϕ_{EB} = coefficient from the same regression (same controls and sample) with the Empirical Bayes estimate $\tilde{\mu}_i^{EB} = r_i\tilde{\mu}_i$ in place of $\tilde{\mu}_i$, which corrects the attenuation of ϕ_{raw} from estimation noise; AP courses are not among the SOL-linked courses used to estimate $\tilde{\mu}_i$. R^2 is from the raw regression. Figure 9 displays the time series of $\hat{\phi}_{EB}$ for all six AP subjects; this table reports even-year cohorts for the three subjects with full-period estimates.

D GPA Distribution

Tables

Table D1. Distribution of Overall High School GPA by Graduation Cohort

Cohort	N	Mean	SD	P10	P25	P50	P75	P90	%≥3.0	%≥3.5	%≥3.75	%≥4.0
2014	92,615	2.78	0.84	1.64	2.24	2.90	3.43	3.81	47.1	22.8	12.4	4.4
2015	93,882	2.80	0.82	1.70	2.27	2.92	3.44	3.81	47.4	23.2	12.3	4.1
2016	98,230	2.82	0.81	1.72	2.30	2.93	3.46	3.80	48.2	23.6	12.4	3.7
2017	103,494	2.84	0.81	1.73	2.32	2.96	3.48	3.82	49.3	25.0	13.1	3.7
2018	106,899	2.85	0.81	1.75	2.33	3.00	3.50	3.83	50.1	25.5	13.6	3.6
2019	107,479	2.87	0.81	1.76	2.35	3.00	3.52	3.85	51.0	26.8	14.8	3.9
2020	107,596	2.90	0.82	1.77	2.38	3.04	3.56	3.87	52.6	28.5	16.0	4.3
2021	106,195	2.92	0.83	1.76	2.38	3.06	3.60	3.89	53.8	30.6	17.9	4.9
2022	108,049	2.92	0.85	1.74	2.36	3.08	3.63	3.90	54.2	31.9	18.8	5.1
2023	108,974	2.92	0.86	1.71	2.36	3.09	3.64	3.91	54.6	32.8	19.7	5.4
2024	108,874	2.89	0.90	1.64	2.31	3.06	3.64	3.91	53.4	32.1	19.3	5.5
2025	111,250	2.90	0.89	1.63	2.32	3.07	3.64	3.91	54.0	32.0	19.7	6.5

Overall high school GPA includes all courses (core and elective). Scale: 0–4 (unweighted). Sample: HS graduates, cohorts 2014–2025.

Table D2. Distribution of Overall High School GPA by Graduation Cohort and Economic Disadvantage Status

Cohort	N	Mean	SD	P25	P50	P75	%≥3.0	%≥3.5	%≥4.0
Econ Disadvantaged									
2014	41,198	2.45	0.86	1.90	2.50	3.08	30.2	11.6	2.4
2016	46,818	2.51	0.82	2.00	2.56	3.11	31.2	12.0	1.8
2018	52,695	2.54	0.83	2.00	2.60	3.15	32.8	13.2	1.6
2020	57,122	2.59	0.84	2.05	2.67	3.24	36.1	15.5	2.0
2022	58,182	2.60	0.87	2.00	2.67	3.29	37.3	17.6	2.3
2024	58,480	2.56	0.92	1.96	2.63	3.29	36.7	18.2	2.6
2025	61,905	2.60	0.90	2.00	2.68	3.32	38.6	18.9	3.2
Not Econ Disadvantaged									
2014	51,417	3.04	0.73	2.62	3.16	3.60	60.6	31.7	6.0
2016	51,412	3.11	0.68	2.71	3.23	3.64	63.6	34.2	5.5
2018	54,204	3.16	0.67	2.78	3.29	3.68	66.8	37.3	5.5
2020	50,474	3.24	0.65	2.88	3.38	3.75	71.3	43.2	7.0
2022	49,867	3.29	0.66	2.95	3.46	3.81	73.9	48.6	8.3
2024	50,394	3.26	0.70	2.91	3.46	3.81	72.7	48.2	8.9
2025	49,345	3.27	0.70	2.92	3.46	3.83	73.3	48.5	10.7

Overall HS GPA (0–4 scale, unweighted). Selected cohorts shown.