Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1456

# The Effects of Capped Piece-Rate Teacher Bonuses: Evidence from Advanced Placement

Md Twfiqur Rahman
Johns Hopkins University

I study a proficiency-based incentive program that rewards Advanced Placement (AP) teachers a piece-rate for each student scoring 3 or higher on the standardized exam. Using student-course-level administrative data and exploiting both withinand across-teacher variation, I find the program increased the probability that an AP enrollment results in a score of 3 or higher by 2.4 percentage points. I further find no evidence of educational triage, as gains were concentrated at the top of the AP score distribution (4s and 5s). Teacher responses also varied systematically, indicating that the effectiveness of performance pay depends jointly on incentive strength and the nature of the task.

VERSION: April 2026

Suggested citation: Rahman, Md Twfiqur. (2026). The Effects of Capped Piece-Rate Teacher Bonuses: Evidence from Advanced Placement. (EdWorkingPaper: 26-1456). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/4zqs-5835

# The Effects of Capped Piece-Rate Teacher Bonuses: Evidence from Advanced Placement*†

Md Twfiqur Rahman ‡
Johns Hopkins University

April 13, 2026

Access the latest version here.

**Abstract**

I study a proficiency-based incentive program that rewards Advanced Placement (AP) teachers a piece-rate for each student scoring 3 or higher on the standardized exam. Using student-course-level administrative data and exploiting both within- and across-teacher variation, I find the program increased the probability that an AP enrollment results in a score of 3 or higher by 2.4 percentage points. I further find no evidence of educational triage, as gains were concentrated at the top of the AP score distribution (4s and 5s). Teacher responses also varied systematically, indicating that the effectiveness of performance pay depends jointly on incentive strength and the nature of the task.

**JEL Codes:** D01, I20, J33, J38, J41

*The views expressed are those of the authors and do not necessarily reflect the positions or policies of the North Carolina Department of Public Instruction.

†I thank Daniel Kreisman, Jonathan Smith, Steven Hemelt, Vjollca Sadiraj, Tim Sass, Cade Lawson, and Brad Hershbein for helpful comments and suggestions. I am grateful to Ian Callen for his suggestions that helped improve the paper on several fronts. The paper also benefited from the suggestions and comments provided by the 2025 Southern Economic Association (SEA) meeting participants.

‡Email: mrahma44@jh.edu

# 1 Introduction

Teacher effectiveness is widely recognized as the most important school-based determinant of student achievement and long-term success (Chetty et al., 2014; Hanushek, 2011; Jackson, 2018). However, the traditional teacher compensation model in the U.S., based on credentials and experience, poorly correlates with teacher effectiveness and does not incentivize additional effort (Chingos & West, 2012; Hanushek, 2011). This disconnect has spurred significant interest in performance-based compensation, rewarding teachers based on their student achievement (Bleiberg et al., 2025; Imberman, 2015).

Empirical evidence on the efficacy of these programs in the U.S. remains mixed, with studies mostly finding lack of evidence to rule out a null effect (Bleiberg et al., 2025; Dee & Wyckoff, 2015; Fryer, 2013; Fryer Jr et al., 2022; Sojourner et al., 2014; Springer et al., 2016).[1] While it is debatable precisely why such muted impacts are observed, researchers commonly cite the complexity of bonus eligibility metrics and the free riding in group-based schemes among other reasons (Fryer, 2013; Goodman & Turner, 2013; Imberman & Lovenheim, 2015; Springer et al., 2016).

In this paper, I study a novel incentive program that rewards individual teachers based on a simple criterion. Beginning in the Academic Year (AY) 2015-16, North Carolina rewards Advanced Placement (AP) teachers $50 for each student of their class achieving a score of 3 or higher (out of 5) in the corresponding standardized AP exam, with total bonuses capped at $3,500 per teacher per academic year.[2] [3]

As theorized in the standard multitask framework detailed in Appendix A, this proficiency-based, piece-rate design induces educational triage: utility-maximizing teachers allocate greater instructional effort toward students just below the passing threshold, where marginal returns to effort are highest. In addition, the program's incentive strength

[1]In contrast, evidence from less developed countries largely supports the efficacy of incentives; for example, see Barrera-Osorio and Raju (2017), Behrman et al. (2015), Duflo et al. (2012), Loyalka et al. (2019), Mbiti et al. (2019), and Muralidharan and Sundararaman (2011).

[2]For example, an AP Biology teacher would receive $1,000 in bonuses if 20 students from their class obtain 3 or higher scores in the College Board administered standardized AP Biology exam. A teacher could earn a maximum bonus of $2,000 during the 2015-16 academic year. Starting in the 2016-17 academic year, the maximum annual bonus increased to $3,500. To put this into perspective, the maximum bonus is equivalent to ∼7% of the base salary for teachers with 13 years of experience in North Carolina.

[3]Throughout this paper, an AP exam score of 3 or higher is considered a passing score, and students receiving such scores are said to have passed the exam.

varies systematically across teachers based on baseline class achievement, summarized by the expected number of passers and non-passers from baseline instructional effort. For teachers with high baseline class achievement, the marginal incentive to exert additional effort declines due to diminishing marginal returns to effort and a closer proximity to the absolute bonus cap.

Testing these predictions requires isolating the bonus effect from potential confounders, such as teacher heterogeneity, changes in classroom composition, and contemporaneous shocks. To (attempt to) do so, I employ two complementary identification strategies using student-course level administrative data from academic year 2012-13 to 2018-19, spanning periods before and after the introduction of the bonus program in AY 2015-16. The first approach isolates the bonus effect using a within-teacher model, which compares the same teacher's student pass rate before and after the bonus conditional on teacher fixed effects. To control for non-random student-teacher sorting and changes in classroom composition, the model additionally includes a comprehensive set of student, grade, and course fixed effects. The second strategy is a Difference-in-Differences (DD) design built on the within-teacher model. I exploit the fact that the program's bonus cap ($3,500) created a credible control group of unincentivized teachers who had pre-bonus average number of passers reaching the maximum payout threshold ($n = 70$), allowing me to difference out contemporaneous shocks that are common to all AP teachers. To assess the sensitivity and robustness of the results, I further present estimates from models that use alternative thresholds ($n \in \{65, 75\}$) to define the control group.

I find robust evidence that the bonus program increased the probability of an AP enrollment resulting in a passed exam by 2.4 percentage points. I further find that this increase in pass rate is driven by broad-based improvement in achievement—significantly more high scores (4 and 5) and fewer low scores.[4]

As discussed above, the program's incentive strength varies across teachers depending on their baseline class characteristics, which I proxy by their respective pre-bonus average number of passers and non-passers. I find the program had significantly larger effect among teachers with lower (higher) number of baseline passers (non-passers). And

[4]In this paper, 'pass rate' refers to the fraction of the AP enrollments that result in a score of 3 or higher in the standardized AP exam. I distinguish this from 'exam pass rate' (pass rate among exam takers), which is defined by the ratio of the number of passed exams to the number of taken exams.

perhaps more importantly, I document evidence supporting a null effect on unincentivized teachers—those whose baseline number of passers predictably qualified them for maximum payout.

The effective strength of a financial reward may also depend on the teacher's marginal utility of income, suggesting the program's impact may vary systematically by locale due to significant spatial variation in teacher compensation. In North Carolina, K-12 teacher compensation is based on a uniform, statewide salary schedule, and local area salary supplements are substantially higher in city and suburban school districts than in rural ones.[5] A fixed-dollar piece-rate, therefore, represents a larger proportional increase in total compensation for teachers in lower-paying rural districts, providing them with a stronger marginal incentive. I find empirical evidence consistent with this notion: the bonus had its largest impact in rural schools, reducing the urban-rural AP credit achievement gap by one-third.[6]

This study advances the literature on incentive design and human capital accumulation in several key ways. First, this study provides new evidence on how incentive strength depends critically on bonus design. In particular, the theoretical predictions, largely supported by empirical evidence, suggest that an individual piece-rate scheme can generate marginal incentives that differ sharply from those created by group-based bonus systems, and may even reverse which teachers face the strongest incentives. Prior work on school-level, all-or-nothing bonuses finds that incentives are strongest for teachers whose effort has the greatest influence on the school's likelihood of meeting the performance threshold, generally those teaching a larger share of tested students (Goodman & Turner, 2013; Imberman & Lovenheim, 2015). By contrast, under the capped piece-rate scheme I study, marginal incentives weaken for teachers with a large number of baseline passers because diminishing returns and proximity to the bonus cap mechanically reduce

[5] This is evident in the data. Using school-district level average teacher salary supplements from North Carolina Public Schools Statistical Profiles, I find that in 2015 an average teacher salary supplement in city schools was approximately $4,800 compared to $2,700 in rural schools.

[6] Similar economic rationale also suggests a stronger response from novice teachers, who have less than 5 years of experience and face steep returns to experience (Papay & Kraft, 2015; Wiswall, 2013). I estimate a significant positive bonus effect for novice teachers, but find that the effect does not significantly diminish for early-career or veteran teachers. Although isolating the pure incentive effect from potential returns to experience is inherently difficult in this context, the fact that the bonus significantly motivates teachers across the entire experience spectrum alleviates concerns that the estimated average effect is primarily driven by novice teacher improvement.

the payoff to additional effort.

Second, in contrast to prior work, which has largely focused on core tested subjects such as Math, Reading, and Science, this paper's policy context provides a unique opportunity to examine how the effectiveness of a single incentive scheme varies across academic disciplines. I find the program's impact was concentrated in a specific subset of courses, including life sciences (e.g., Biology, Environmental Science) and certain social sciences (e.g., Microeconomics, World History). I find no statistically significant effect in many core humanities (e.g., English Literature, U.S. History) or other STEM subjects (e.g., Calculus, Statistics).[7] This heterogeneity highlights a non-negligible role of subject-specific production functions in mediating an incentive's effects.

Third, it provides new evidence on the behavioral mechanisms of teacher incentives. Contrary to the common concern that proficiency-based accountability induces educational triage (Ballou & Springer, 2017; Jacob, 2005; Neal & Schanzenbach, 2010), I find this proficiency-based incentive program led to broad-based score gains, significantly increasing the probability of students achieving the highest scores (4 and 5). This type of non-distortionary response, where incentives appear to boost broad-based performance, has been observed in other contexts, including pay-for-percentile tournaments (Loyalka et al., 2019) and group-based incentives that improved performance on both high- and low-stakes exams (Lavy, 2009; Muralidharan & Sundararaman, 2011). A plausible explanation that reconciles the divergence of these observed effects across programs may be the framing of the incentive: the threat of sanctions in accountability systems may induce a narrow, loss-avoiding focus on marginal students, whereas the prospect of a pure reward, as in this context, may encourage a broader improvement in overall instructional quality. The importance of this "punishment" versus "reward" framing on teacher behavior is highlighted in field experiments (Fryer Jr et al., 2022; Hossain & List, 2012).[8]

Fourth, to the best of my knowledge, the evidence that incentive is extinguished for capped-out teachers provides a new empirical result on how agents respond to an

[7] For contextual clarity, particularly when discussing heterogeneity, I use the term 'subject' to refer to a specific AP 'course' (e.g., AP Calculus AB or AP Biology). Throughout this paper, these terms are used interchangeably.

[8] The absence of evidence for triage may also arise from potential difficulty in identifying marginal students in the AP context. Because AP exams are externally designed and imperfectly aligned with classroom assessments, teachers may face substantial uncertainty in identifying students near the passing threshold, reducing the scope for targeted effort.

explicit contractual threshold. With caution, however, this result can be contrasted with the long-standing debate in the labor supply literature over behavioral responses to implicit, unobserved income targets (Camerer et al., 1997; Farber, 2005). A key challenge in that literature is that the income target (among taxi drivers) is unobserved and must be inferred by the researcher. While the result—that capped out teachers did not respond to the bonus—does not resolve that specific debate, it lends partial support toward threshold-based effort reductions when the target is an explicit and observable feature of an incentive contract.

Finally, this study provides a crucial refinement to the foundational work on Advanced Placement Incentive Programs (APIP). Examining the effects of Texas APIP, Jackson (2010) documented large positive effects of locally funded incentives on AP student achievement. However, the Texas program provided large, multifaceted financial rewards to both teachers and students, making it impossible to disentangle the effect of the teacher incentives from the direct student incentives.[9] This study sidesteps this challenge by analyzing a program that rewards teachers exclusively. This unique program design allows me to isolate the effect of a teacher-only incentive on student AP pass rate.

This paper is organized as follows. Section 2 delineates the policy background. Section 3 provides an intuitive discussion of the theoretical predictions of the teacher effort model formalized in Appendix A. Section 4 explains the data and provides summary statistics. Section 5 discusses the empirical strategies and the results. Section 6 discusses limitations related to external validity and the policy's cost-effectiveness, and Section 7 concludes.

# 2 Policy Background

The North Carolina Advanced Placement teacher bonus program was established by the General Assembly as a pilot program beginning in the 2015-16 academic year. A year later the program was made permanent, and under the provisions of the relevant Session Laws (SL 2016-94 and SL 2017-57), has been rewarding AP teachers $50 bonus for each

[9]Under the Texas APIP program, not only did students receive exam fee subsidies and bonuses of $100 to $500 for each passing score, but also teachers earned similar piece-rate bonuses and, in some cases, substantial salary supplements, ranging from $2,000 to $10,000 (Jackson, 2018). This bundle of high-powered incentives for both teachers and students differs fundamentally from the teacher-only piece-rate bonus I analyze.

student in their class achieving a score of 3 or higher on the corresponding AP exam. The maximum payout to a given teacher per academic year is capped at \$3,500.[10]

While the program also rewards teachers of other advanced coursework programs, such as International Baccalaureate (IB) and Cambridge AICE, this paper focuses exclusively on the AP program for two reasons. First, the AP program is by far the most prevalent advanced coursework option in the state, with significantly higher enrollment and a broader presence across all school districts. For instance, there were 149,583 enrollments in AP courses compared to 20,140 and 500 enrollments in IB and Cambridge AICE courses in AY 2016-17, respectively (North Carolina Department of Public Instruction, 2019). Second, while the administrative data I use in this study contain linkable records for AP course enrollment and exam performance, they do not include equivalent data for IB or AICE exams during the study period.

Importantly, North Carolina introduced the AP teacher bonus program a year after the state implemented its universal AP exam fee waivers policy, funding all AP exams taken by students enrolled in the corresponding AP courses in NC public schools beginning in AY 2014-15. Additionally, the state formed a partnership with College Board—the organization that designs AP curriculum and administers the standardized tests—to provide administrative and professional assistance primarily to a set of low-performing, target school districts each academic year beginning in AY 2013-14 (North Carolina Department of Public Instruction, 2014, 2015). While these overlapping policies constitute major identification threats, in Section 5, I provide additional institutional details and conduct extensive robustness and sensitivity analyses that circumvent concerns of contamination.

# 3 Conceptual Framework

This section outlines the theoretical predictions and associated mechanisms regarding how the incentive design under consideration affects teacher effort. Appendix A details the formal model. Intuitively, because the program pays per passing student while imposing a maximum payout limit, the marginal incentive to exert additional effort depends

[10] In the initial year of implementation (AY 2015-16), the bonus cap was set at \$2,000.

on a teacher's baseline number of passers, which determines both the expected reward and the proximity to the cap. Returns to additional effort also depends on the available pool of marginal students (the 'bubble' window) who can be pushed across the passing threshold with heightened instruction. Taken together, these forces shape the optimal allocation of teacher effort and generate clear predictions regarding teacher responses to the bonus.

1. *The cap effect (Proposition 1).*– Because effort is costly, teachers whose expected baseline number of passers meets or exceeds the bonus cap do not face positive marginal incentives to exert additional effort.

2. *Educational triage (Proposition 2).*– The piece-rate structure encourages teachers to allocate higher targeted effort toward students near the passing threshold. Since rewards depend solely on the number of students crossing a single threshold (i.e., a score of 3 or higher), a utility-maximizing teacher will not devote this costly effort on students who are expected to pass without additional tutoring or to those unlikely to pass even with substantial additional tutoring.

3. *Baseline number of passers and incentive strength (Proposition 4).*– The marginal incentive to exert additional effort decreases with a teacher's baseline number of passers. Under standard economic assumptions, teachers with higher baseline face weaker incentives due to diminishing returns to additional effort and higher likelihood that the bonus cap binds. From another perspective, a high baseline number of passers may reflect a higher ability composition of students within a teacher's class, shifting the achievement distribution upward and reducing the potential density of students expected to fall just below the passing threshold.

4. *Baseline number of non-passers and incentive strength (Proposition 6).*– A teacher's incentive strength depends not only on their baseline number of passing students but also on their baseline number of non-passing students. Under a standard monotonicity assumption, a larger baseline pool of non-passers creates a more target-rich environment, increasing the absolute number of marginal students who can be pushed across the passing threshold through a combination of heightened general instructional rigor and targeted effort.

These predictions partially guide the empirical analysis that follows. In particular, this study examines whether teacher bonuses increase overall pass rates, where gains are concentrated, and whether responses vary with baseline number of passing and non-passing students.

# 4 Data

### A. *Sample Construction*

This study utilizes student-course level administrative data from the North Carolina Education Research Data Center (NCERDC), covering the academic years 2012-13 through 2018-19. The analytic dataset is constructed by first extracting AP course enrollments from high school transcripts and merging them with official AP exam records from the College Board.[11] I then link the AP course enrollments to their respective course teachers using the course membership files. Unfortunately, approximately 10% of the student-courses could not be matched to the class teachers due to missing teacher identifiers in the course membership records. According to NCERDC documentation, these are not subject to systematic omission.[12]

To isolate the effects of the teacher bonus program, I apply a series of necessary sample restrictions. Starting with the full sample of AP course enrollments, I first limit the sample to observations matched to teachers, creating the "Matched Sample". Because the goal is to understand whether the bonus program increases teacher effort and consequently student achievement, I then restrict the sample to classes taught by teachers teaching both before and after the introduction of the bonus program, giving us the "Continuing Teacher Sample." Finally, to rigorously control for unobserved, time-invariant student-

[11] I follow Fazlul et al. (2021) and restrict the analytic sample to the second semester (i.e., Spring term) or year-long enrollments. This restriction is appropriate for two reasons. First, most AP courses span two semesters (Fall and Spring). Second, in North Carolina public schools, the few one semester-long AP courses (e.g., Microeconomics, U.S. Government and Politics, Psychology) are generally offered in the Spring to avoid a timing mismatch between course instruction and the exam. This institutional feature is evident in the data: AP Psychology, a semester-long course, is the second most popular AP course in the sample.

[12] The publicly available NCERDC report titled "Technical Report 1: Linking Teachers in the ABC Data to Teachers in the School Activity Report" provides further technical details about the matching process. Also importantly, approximately 98.4% of the student-courses are matched uniquely to only one teacher. This share increases to ∼99% for the main analysis sample, the construction of which I describe next.

level confounders using student fixed effects, I confine the analytic sample to students who are observed in multiple AP courses. This yields the primary "Analysis Sample" of 383,560 student-course observations, which is used to estimate the main coefficients on interest. In section 6, I provide evidence that these sample restrictions, if anything, may lead to understatement of the true effect.

### B. *Summary Statistics*

Table 1 reports the summary statistics from student-course level data from 2013 to 2019 for the Full and Analysis samples separately.[13] Comparing column 3 to column 6 of the table, we observe that pre-to-post-bonus changes in average student characteristics and academic variables are comparable between the samples. However, the average pre-period 8th-grade Math and Reading z-scores are about 0.1 Standard Deviation higher in the Analysis Sample (Column 1 vs. 4), implying that the main estimates come from a sample of students with relatively higher ability than the general AP population. Focusing on the Analysis sample, we observe a noticeable change in student composition in AP courses following the introduction of the bonus program, with the share of Asians and Hispanics increasing. The exam participation also increased by 6 percentage points, which is suggestive of potential spillover of the universal AP exam fee waivers policy introduced a year before the implementation of the bonus program (Column 6). Evidence presented later will demonstrate that this paper's main findings (on probability of an enrollment leading to a passed exam) are not contaminated by fee waivers. Finally, the raw mean difference suggests a 2.8 percentage point increase in AP pass rates (Column 6).

Figure 1 plots the percentage point change in the share of AP enrollments resulting in passed AP exams, using 2015 as the baseline reference.[14] For teachers who continued to teach AP courses starting in the pre-bonus period, the trend is flat throughout the pre-period. A steady rise is observed following the introduction of the program. However, the pre-trend for the Analysis Sample has been steadily decreasing in the pre-period and rising thereafter. Importantly, despite variation in pre-trend, pass rate across the samples followed similar, increasing trajectory in the post-bonus period.

[13] To clarify, I refer to Academic Year 2014-2015 using the ending year '2015', and other academic years referred to following this same convention.

[14] The year coefficients are estimated by regressing the year coefficients on the course-level binary indicator of exam passing.

## C. *Teacher-level Measures*

The main analysis of this paper centers on the 2,712 teachers observed teaching AP courses in both the pre- and post-bonus periods. On average, a teacher in this sample has about 14 years of teaching experience, an approximately 50% probability of holding an advanced degree, and a pre-bonus student load of approximately 32 students per academic year.[15]

The difference-in-differences (DD) design employed in this paper exploits the bonus cap to identify a credible control group of unincentivized teachers—i.e., teachers whose baseline number of passers predictably qualifies them for the maximum bonus. To avoid endogeneity, I calculate a proxy measure of each teacher's baseline number of passers by averaging the number of passers from the broader "Continuing Teacher Sample" over the pre-bonus years.[16] I identify 57 teachers who, with an average of 70 or more pre-bonus passers, serve as the primary (unincentivized) control group under the DD design [see Figure B.1]. However, a teacher's actual number of passers naturally varies over time, introducing a potential for misclassifying them into groups being compared. This potentially makes the DD estimates conservative, mainly because some unincentivized teachers may drop below the 70-passer threshold (facing strictly positive marginal incentives), while some incentivized teachers may surpass the threshold (facing no marginal incentive). Both types of misclassifications contaminate the comparison and would attenuate the estimated effects toward zero. Notwithstanding, I further examine the sensitivity of these estimates via alternatively defined control groups based on 65 and 75 average number of pre-bonus passers.

Figure 2a provides non-parametric evidence supporting the DD design. The binned scatter plot illustrates the relationship between a teacher's pre-bonus number of passers and their respective pre-to-post change in pass rates.[17] The pattern reveals two distinct

[15]Figure B.2 plots the frequency distribution of teachers' pre-bonus student load.

[16]I use the broader "Continuing Teacher Sample" to construct these baseline characteristics because the non-singleton students in the Analysis sample alone do not represent a teacher's complete pre-bonus workload or effectiveness. This procedure yields a more stable and representative proxy for a teacher's typical pre-period environment. However, I also note that, throughout the paper, the pre-to-post change in pass rate for a given teacher is calculated using the Analysis sample, providing comparable descriptive evidence for the main causal estimates.

[17]This plot is based on teacher-level aggregate data and generated by locally regressing teachers' pre-bonus number of passers on their pre-to-post change in pass rate within data-driven bins [See Cattaneo et al. (2024) for econometric details].

regimes. For teachers with a low baseline, a strong, downward-sloping relationship is observed, with the largest gains concentrated among teachers with least number of pre-bonus passers. Crucially, for teachers with a high baseline (more than 40 pre-bonus passers), the relationship flattens and becomes statistically indistinguishable from zero. This provides a strong descriptive evidence that the marginal incentives were indeed extinguished for this high-performing group, at least in part, due to the bonus cap. Moreover, Figure 2b highlights that a teacher's the baseline number of non-passers is positively associated with the pre-to-post change in pass rates. Taken together, they are indicative of varying teacher responses to the bonuses given a teacher's baseline class characteristics.

Finally, while the number of teachers in the control group is modest, they are highly active, accounting for over 17,000 student-course observations in the Analysis sample. On the other hand, the unincentivized group is on-average slightly more experienced than the incentivized group (14.1 vs. 13.4 years). However, Figure 3 exhibits that this small difference in means is not indicative of a fundamental disparity; the overall experience distribution for both groups are roughly similar and exhibit substantial overlap. This comparability alleviates potential concerns that the results may be driven by differential returns to experience across the groups being compared, which I formally investigate later.

# 5 The Impact of Teacher Bonuses

The empirical analysis of this paper proceeds in four steps, moving from the average effect of the bonus program to the deeper mechanisms that, in part, explain it. First, I answer the fundamental question of whether the bonus program effectively increased pass rate among AP enrollees by documenting robust significant estimates across both the within-teacher and difference-in-differences models. Second, I ask how this improvement was achieved by examining the program's impact across the full AP score distribution. Third, I test the theoretical predictions of the incentive design by examining heterogeneity in effects by teachers' baseline class characteristics. Finally, I document heterogeneity by AP subjects, student demographics, teacher experience, and school urbanicity.

## 5.1 The Average Effect

To examine the effect of the teacher bonus program on the probability of an AP enrollee passing the exam, I rely on the Analysis sample and estimate the following within-teacher model.

$$\text{Pass}_{icjt} = \tau \text{Post}_t + \theta_j + \gamma_c + \delta_g + \alpha_i + \beta \mathbf{X}_{icjt} + \varepsilon_{icjt} \quad (1)$$

Where, $Pass_{icjt}$ is the binary indicator that student $i$ in AP course $c$ taught by teacher $j$ at year $t$ achieves a passing score in the standardized AP test. $\text{Post}_t$ [$=\mathbb{I}(\text{Year} \geq 2016)$] is the policy variable, taking a value of 1 in years when the teacher bonus program is active and 0 otherwise. $\tau$ captures the effect of the bonus program on the probability of an AP course resulting in a passing exam score.

This model exploits outcome variation within teachers over time by leveraging teacher fixed effects ($\theta_j$) to absorb all stable differences in quality, effectiveness, and teaching style across teachers. To ensure that the within teacher comparison over time is not confounded by the variation of mix of students taught, I include grade ($\delta_g$), course ($\gamma_c$), and student ($\alpha_i$) fixed effects. The vector $\mathbf{X}_{icjt}$ contains two time-varying covariates: the district-level indicator for participation in the North Carolina Advanced Placement Partnership (NCAPP) to account for the effects of targeted professional assistance (if any) and the subject-year level annual national pass rate among exam takers to absorb year-to-year changes in exam difficulty.[18] $\varepsilon_{icjt}$ is the idiosyncratic error term.

A critical assumption underlying the validity of the within-teacher model is that, the pass rate among students assigned to a given teacher would not have changed over time in absence of the bonus program. I probe the validity of this assumption by estimating event study coefficients and find that the pre-bonus evolution of pass rate was stable

[18] I note again that "pass rate among exam takers" is fundamentally different from "pass rate." While the numerator is same for both measures, the denominator of the former is the number of exam takers and the denominator of the latter is the number of AP course enrollees. The data on national pass rate among exam takers are sourced from the College Board: https://apstudents.collegeboard.org/about-ap-scores/score-distributions [Retrieved: 06.10.2025]

leading up to the introduction of the bonus policy (see Figure 4).[19]

Table 2 reports estimates from the within-teacher model (Eq. 1). Estimates from the most rigorous specification, including student fixed effects in addition to teacher, subject, and grade fixed effects, show that the probability of an AP enrollment resulting in a passing exam score increased by 2.6 percentage points following the introduction of the bonus program. Comparing columns 4 and 5, I find that including student-level controls—sex, race, socioeconomic status, and 8th-grade Math and Reading z-scores—in lieu of individual fixed effects attenuates the estimated effect. This pattern is consistent with negative omitted variable bias in the specification without student fixed effects.

Such results, however, could be driven by unobserved, statewide time shocks common to all AP teachers in North Carolina. To surmount this challenge, I exploit the fact that the bonus program was capped at \$3,500, implying that teachers faced zero marginal incentives to exert additional effort to generate passes beyond 70 (\$3,500 / \$50). This allows me to the sample of teachers into treatment and control groups based on their respective average pre-bonus number of passers, $\text{NPassPre}_{j,pre}$.[20] I do this by defining a binary variable, $\text{Incentivized}_j$, that takes a value of 1 if $\text{NPassPre}_{j,pre} < 70$ and 0 otherwise. I then compare the within-teacher changes in pass rate among incentivized teachers to those unincentivized over time to estimate the effect of the bonus program.

$$\text{Pass}_{icjt} = \lambda(\text{Post}_t \times \text{Incentivized}_j) + \theta_j + \gamma_c + \alpha_i + \rho_t + \beta \mathbf{X}_{icjt} + \varepsilon_{icjt} \tag{3}$$

While preserving the notational conventions, I hereby transform the within-teacher model (Eq. 1) into a difference-in-differences model through two changes. First, the addition

[19] The dynamic specification of Equation 1 is as follows.

$$\text{Pass}_{icjt} = \tau_t \sum_{k \neq 2015} \mathbb{I}(k = t) + \theta_j + \gamma_c + \delta_g + \beta \mathbf{X}_{icjt} + \varepsilon_{icjt} \tag{2}$$

Where, $\tau_t$ captures the effect of the bonuses at year $t$ relative to the year 2015. Note that, I exclude student fixed effects from this model since the unbalanced structure of the student-course panel makes the year coefficients uninterpretable. This is primarily because, in this specification, each year-coefficient is estimated using a different, non-random sample of students, and due to near-perfect collinearity between within-student high school grade progression and year, the effect of student maturation cannot be separately identified in the presence of student fixed effects.

[20] To accurately proxy a teacher's potential total bonus amount in a given academic year, I calculate annual number of passes for a given teacher from the "Continuing Teacher Sample," which includes all singleton and non-singleton students taught by an AP teacher. I then average each teacher's respective annual number of passes over 2013 to 2015 to calculate $\text{NPassed}_{j,pre}$.

of an interaction between $Post_t$ and the Incentivized$_j$ dummies enable the comparison of outcome growth between the incentivized and unincentivized teachers over time. The coefficient of interest, $\lambda$, captures the effect of the bonus program on incentivized teachers. Second, I add year fixed effects, $\rho_t$ to absorb year-specific shocks common to all AP teachers. Because high school grade and year are collinear within students, I exclude grade fixed effects from the specification including student fixed effects.

The primary identifying assumption of the difference-in-differences model is that the pass rate across the incentivized and unincentivized teachers would have evolved similarly over time in the absence of bonus incentives. I test the validity of this assumption by estimating a dynamic DD specification.[21] Figure 5 plots the estimated coefficients for each year, with coefficients estimated relative to the year immediately preceding the bonus, 2015. It shows that the evolution of pass rate remained relatively stable over the pre-bonus period. The positive post-bonus year-coefficients provide suggestive evidence that the bonuses increased relative pass rates among incentivized teachers.

Another key assumption is that the change in the number of students assigned to a teacher is not endogenously determined by their response to the bonus. To test, I examine the correlation between the pre-to-post change in a teacher's student load and the corresponding change in their pass rate in Figure B.3. The estimated near-zero correlation (correlation coefficient = 0.019) provides strong evidence that the type of dynamic, performance-based sorting that would threaten the DD design is not a significant concern in the data.

Table 3 reports the estimates from the difference-in-differences model (Eq. 3). The estimates from the most rigorous specification in column 4 show that the bonus program

[21] I estimate a modified version of Equation 3, where I replace the single Post$_t$ variable with a series of year-specific dummies.

$$\text{Pass}_{icjt} \quad = \quad \sum_{k \neq 2015} \lambda_t \cdot \text{Incentivized}_{j,pre} \cdot \mathbb{I}(k=t) + \theta_j + \gamma_c + \rho_t + \alpha_i + \varepsilon_{icjt} \tag{4}$$

The event study coefficients $\lambda_t$ measure the difference in outcomes between incentivized and unincentivized teachers in each year, relative to 2015. Note that student fixed effects can be included in this dynamic specification under two assumptions: (i) the effect of student maturation is common across all students regardless of teachers they are assigned to, and (ii) the unbalanced nature of the nonrandom sample of AP students used to estimate the year-coefficient are roughly similar across the comparison groups over time. The similarity between the coefficient estimates with or without the student fixed effects (Figure 5 vs. Figure B.4) provide strong evidence supporting the validity of the DD design.

significantly increased pass rate among AP enrollees by 2.4 percentage points ($p$=0.052). This estimate is less precise but similar to that from the baseline within-teacher model in Table 2 and suggestive of negligible bias from common time trends that affect all AP teachers. I further test the robustness of this result by altering the threshold that defines the unincentivized control group. The estimated effect increases a little when the threshold is lowered to 65 pre-bonus passers and increases dramatically when using a stricter threshold of 75, with the parallel trends assumption well-supported in all cases [see Table B.3 and Figure B.5]. The robustness of the main effect across these specifications provides strong support of a consistent positive effect of the bonus program.[22]

*Additional robustness and sensitivity checks.-* A potential concern with the main difference-in-differences specification is that it abstracts from the policy's initial bonus cap of $2,000 in AY 2015–16 (implying an effective threshold of 40 passers), which was increased to $3,500 (70 passers) in subsequent years. To align treatment with this policy shift, I estimate specifications that allow incentive exposure to vary across groups and over time: teachers with fewer than 40 baseline passers are treated starting in 2016, while those with 40-69 baseline passers are treated starting in 2017, with teachers at or above 70 passers serving as the (never treated) capped-out group. Appendix Table B.1 presents the results. These estimates reinforce the main findings and are informative about heterogeneity in responses across baseline teacher capacity. The positive response is driven primarily by teachers with fewer than 40 baseline passers, while the estimated effect for teachers with 40–69 passers is not statistically different from zero (column 4). This pattern is consistent with declining marginal incentives as teachers approach the cap, which I examine in Section 5.3 using a continuous difference-in-differences framework.

A second potential concern is that the results may be confounded by the North Carolina Advanced Placement Partnership (NCAPP), a collaboration between the state and the College Board initiated in the 2013–14 academic year, through which the Col-

[22] I also acknowledge that AP teachers may also teach IB and Cambridge AICE courses. This has important implications for the analysis, since under the bonus program, teachers of all advanced courses are rewarded if their students achieve a passing score in the respective standardized exam. However, I argue that this is not a significant driver of the main results for two key reasons. First, it is relatively uncommon for a school to offer both AP and IB/AICE programs concurrently. Thus, the number of teachers who are the instructor of record for courses across these different programs are likely very small. Second, the teacher fixed effects are well-suited to handle this. If a teacher consistently teaches a mix of AP and IB courses, this is part of their stable, unobserved characteristic, and the fixed effect will absorb it.

lege Board provided administrative and professional support to a targeted set of low-performing school districts (North Carolina Department of Public Instruction, 2014, 2015).[23] In Appendix Table B.2, I provide evidence that the main estimates are qualitatively similar when I exclude all observations from the NCAPP target school districts (columns 2 and 4).[24]

A third, and critical, threat to the identification is the statewide AP exam fee waiver policy implemented in 2015, the year immediately preceding the introduction of the teacher bonus. This concern is unlikely to explain the results, as shown by several empirical checks. To begin with, the event study plots (Figures 4 and 5) show no statistically significant jump in pass rates in 2015, the first year all students received AP exam fee waivers when the teacher bonus program was inactive. In section 5.2, I further document that the estimated effect of the bonus program on exam participation is insignificant and very small in magnitude (0.006, SE=0.012), ruling out substantial confounding from the fee waivers.[25] Furthermore, Rahman and Lawson (2025), exploiting within-student variation induced by the fee waivers, find that the waivers increased exam-taking but not passing propensity among students taking multiple AP courses both before and after the waivers are implemented. Adopting this strategy as a falsification test, I document that for students observed taking AP courses both before and after the introduction of the bonus program, pass rates rose by a statistically significant 2.1 percentage points,

[23]Although a statewide partnership, NCAPP's assistance was primarily focused on a specific set of low-performing school districts—initially 23 and covering a total of 34 by the end of the study period (North Carolina Department of Public Instruction, 2015, 2019). Data from North Carolina Department of Public Instruction (NC-DPI) reports indicate that the vast majority (over 90%) of teachers participating in College Board AP training workshops during this period were employed by the NCAPP target districts (North Carolina Department of Public Instruction, 2014, 2015, 2016, 2017).

[24]Also, the main specifications include a 'target' dummy variable to separately control for the effects of NCAPP. Estimates in Columns 1 and 3 of Appendix Table B.2 show that exclusion of the dummy results in roughly identical estimates.

[25]This argument rests on the economic intuition that the effect of exam fee waivers on pass rates must operate primarily through exam participation. To further assess robustness, I rigorously test the sensitivity of the pass rate estimates by restricting the analysis to the post-waiver period (2015–2019) to isolate the marginal effect of the bonus; see Appendix Table B.4. When I define incentivized teachers using alternative cutoffs of 65 or 75 average pre-bonus number of passers (to account for potential measurement error or behavioral responses near the threshold), the estimated effects on pass rates remain similar to the main estimates. The estimate using the policy-determined cutoff of 70 is positive but less precise, with the 95% confidence interval fully covering the 95% confidence interval of the baseline estimate. Crucially, regardless of the threshold used to classify incentivized teachers, I find no statistically significant effect on exam participation (Panel B), directly falsifying the hypothesis that exam subsidies drive the results.

whereas exam participation did not (see Appendix Table B.5).[26]

Finally, both the NCAPP and the AP exam fee waiver policy were implemented as part of North Carolina's "Broaden Successful Participation to Advanced Courses" initiative legislated in 2013. This initiative also encourages schools to provide AP preparatory resources to middle school students. However, because benefits of these efforts were realized before students entered high school and enrolled in AP courses, the resulting differences in students' pre-AP academic inputs are absorbed by the student fixed effects.

### 5.2 Effects on AP Score Distribution

The program awards bonuses based on a strict absolute threshold: a score of 3 or higher. This creates an incentive for teachers to engage in educational triage—that is, to allocate targeted effort to students just below the passing threshold (Gillborn & Youdell, 1999; Neal & Schanzenbach, 2010). I test this theoretical prediction by examining the program's effect on the distribution of AP exam scores using the difference-in-differences framework (Eq. 1). I rely on two complementary outcome measures for each possible exam score, $k \in \{1, 2, 3, 4, 5\}$. First, I use a non-cumulative measure, $\mathbb{I}(\text{Score} = k)$, to investigate whether the score distribution is reshaped by the incentives. Second, I use a cumulative measure, $\mathbb{I}(\text{Score} \geq k)$, to estimate the program's effect on probability of clearing a certain threshold, such as a passing score.

Table 4 provides evidence of a shift in the entire score distribution, consistent with a broad-based improvement rather than a narrow spike at the passing threshold. Focusing on the non-cumulative measures in Panel A, I find a negative but insignificant effect on the probability of a course leading to the lowest score. Such negative impact is observed throughout the bottom score bands (1 to 3), whereas positive significant gains are observed for scores at the top of the distribution—2 and 1.9 percentage point increase in the probability of achieving a score of 4 and 5, respectively. These results provide strong evidence contrasting educational triage hypothesis, which predicts largest score gain at the passing threshold.

[26] I acknowledge that fee waivers may be a necessary condition for the observed gains in pass rates following the introduction of the bonus program. While the bonuses drove the marginal improvement in performance, financial barriers could have precluded exam participation among many AP enrollees who were likely to pass (Fazlul et al., 2021; Rahman & Lawson, 2025).

Panel B uncovers further details. I find a statistically insignificant effect on the probability of achieving 1 or higher score. Because 1 is the minimum score an enrollee taking the exam can get, this estimate indicates that changes in exam participation is not a key mechanism for this sample. Crucially, the effect peaks beyond the passing threshold at score 4, increasing the probability of achieving a score $\geq 4$ by approximately 4 percentage points. This reinforces the implication that the significant average effect discussed earlier is likely driven by improvement in general instructional quality.[27]

It is plausible that teachers might have allocated higher targeted effort toward students who do not enroll in multiple AP courses, whom the main Analysis sample omits. However, results from the difference-in-differences model estimated by excluding student fixed effects and on the broader "Continuing Teacher Sample"—which includes singleton students—reveal a similar pattern observed for the Analysis sample: the bonus effect is concentrated at the top of the AP score distribution, with significant gains for scores of 4 and 5 [see Table B.7]. This reinforces the suggestion that the bonuses primarily induced a broad improvement in instructional quality.

### 5.3 Incentive Strength

As argued in section 3 (and formally deduced in Appendix A), the incentive strength teachers face under this bonus program varies along their respective baseline number of passers and non-passers. To probe the empirical validity of such prediction, I estimate continuous difference-in-differences models, where the treatment intensity is determined by a teacher's baseline number of passers and non-passers. Specifically, I replace the binary indicator for teachers with baseline number of passers below 70 ($\text{Incentivized}_j$) in model 3 with continuous measures of baseline class characteristics that predictably determine incentive strength. I then estimate the coefficients of the interactions between the post indicator with pre-bonus average number of passers ($\text{NPass}_{j,pre}$) and non-passers ($\text{NFail}_{j,pre}$) separately. Theoretically, we would expect higher number of passers (non-passers) to have negative (positive) effects on pass rates.

[27]In Table B.6, I provide results from the within-teacher model (Eq. 1). The estimates tell a relatively weaker story, but still provides strong evidence of significant positive effects on the high scores (4 and 5).

The estimates in Table 5 provide strong empirical support for the theoretical predictions. Column 3 demonstrates that the marginal effect of the policy is highly responsive to baseline classroom composition. Specifically, one more student in the baseline non-passer pool increases pass rate by 0.09 percentage points. In contrast, an additional baseline passer reduces the effect by 0.08 percentage points.

Such interpretation of the continuous DD coefficients, however, is valid under a strong linearity assumption, which often fails to hold in the data (Callaway et al., 2024).[28] Consequently, I estimate an extended specification, which includes up to third-order polynomial of the baseline proxies. Appendix Table B.8 presents the results. Calculating the marginal effect from the cubic specification (column 3) at the sample mean (15.49 pre-bonus average number of passers) reveals a substantially steeper penalty of ∼0.21 percentage points from an additional baseline passer. This suggests that the parsimonious difference-in-differences model provides a conservative lower-bound estimate of the incentive effect for the average teacher, as it averages across the steep initial penalty and the flatter regions further up the distribution.[29]

### 5.4 Heterogeneity

To examine heterogeneity, I augment the DD model with an interaction between the heterogeneity dimension ($Z_{icjt}$) and the main difference-in-differences interaction term ($\text{Post}_t \times \text{Incentivized}_{j,pre}$), yielding a general model of the form:

$$\begin{aligned} \text{Pass}_{icjt} = \lambda \left(\text{Post}_t \times \text{Incentivized}_j\right) &+ \sum_{k \neq 0} \mu_k \left(\text{Post}_t \times \text{Incentivized}_j \times \mathbb{I}(\text{Z} = k)\right) \\ &+ \zeta_k + \theta_j + \gamma_c + \alpha_i + \rho_t + \beta \mathbf{X}_{icjt} + \varepsilon_{icjt} \end{aligned} \tag{5}$$

[28] This potential nonlinearity can be observed in the nonparametric bin-scatter plots base on teacher level data (see Appendix Figure B.1).

[29] While the bonus program directly rewards the absolute number of passing students, a teacher's incentive strength can alternatively be parameterized by their baseline pass rate and total student load. Appendix Table B.9 reveals a similar story: Column 3 estimates indicate that a 10-percentage-point increase in a teacher's baseline pass rate reduces the treatment effect by approximately 1.7 percentage points, whereas total student load has no statistically significant impact. Importantly, the binscatter plot (Figure B.6a) provides strong visual evidence supporting the linearity assumption of the continuous DiD model for the pass rate.

Where, $Z$ represents the dimension of heterogeneity, and $k$ indexes the different groups within. The coefficient $\lambda$ captures the effect of the omitted reference category, and $\mu_k$ captures the differential effect for category $k$ relative to reference category. The total effect for a given category, $k$, is given by $\lambda + \mu_k$.[30]

I examine heterogeneity along four key dimensions. First, I study subject-level heterogeneity, where Z represents the vector of AP subject indicators. Second, I examine heterogeneity by locale, where Z represents the vector of school urbanicity indicators.[31] Third, I analyze heterogeneity across student demographics by investigating how the effects of the program vary by gender, socioeconomic status, and underrepresented minority (URM) status, where URM students are those who are neither White nor Asian. Finally, I examine heterogeneity by teacher experience, where Z represents a categorical variable classifying teachers based on their initial pre-period experience as Novice (less than 5 years), Early-Career (5 to 9 years), or Veteran (10 or more years).[32]

### A. *Heterogeneity by Subjects*

Figure 6 plots the estimated effect for different AP subjects along with 95% confidence interval. The first number within the parentheses accompanying each subject title reports the pre-bonus percentage of student-courses and the second number reports the point estimate for the given subject. The plot reveals significant heterogeneity in the program's effect across AP subjects.

The results for STEM subjects are mixed. The bonus had large positive effects on the life sciences: Biology and Environmental Science. In contrast, for most physical science courses, including mathematics, Chemistry, and Calculus AB/BC, the estimated effects are not statistically distinguishable from zero.

Among history and social sciences, pass rates significantly improved in US Government and Politics, Microeconomics, and World History courses. However, for other popular subjects in this group, such as US History and Psychology, I estimated precise null effects.

[30] The standard errors are estimated using the delta method.

[31] The urbanicity indicator is a categorical variable constructed from the detailed ULOCAL codes in the National Center for Education Statistics (NCES) Common Core of Data (CCD), collapsed into three categories: City, Suburban, and Town/Rural.

[32] Note that, the category fixed effects, $\zeta_k$, are collinear with other fixed effects and are absorbed, except for when capturing permanent differences across school locales.

Within the Arts, the effects are positive but noisy.[33] A few language courses, such as Spanish Language and Spanish Literature show a statistically significant increase in pass rates, whereas the effect for Art History and English Language courses are positive but imprecisely estimated.

Taken together, there appears to be no single characteristic, such as the subject's overall popularity or its broad disciplinary category, that separates the courses for which the bonus program was effective. However, this complex pattern hints at the underlying importance of the subject-specific educational production function. It is speculatively plausible, that incentives were most effective in subjects where the educational production function is straightforward, meaning that there exists subject-specific core concepts with widespread application. In Microeconomics, for example, a teacher can focus extra effort on a portable concept like supply and demand, confident that mastery of this single, codifiable model will directly equip students to answer a wide range of questions across different contexts. Similarly, in Environmental Science, the vast but well-defined curriculum allows a teacher to target specific, testable content, such as the nitrogen cycle or key environmental laws, to reliably produce score gains. This contrasts sharply with a cumulative subject like Calculus, where a teacher's effort can be rendered ineffective if a student lacks the prerequisite algebra skills from prior years. In such cases, the roadmap from effort to success is less certain.

### B. *Heterogeneity by Student Demographics*

Turning to heterogeneity by student demographics, I find substantial baseline disparity in AP performance. In the pre-bonus period, AP enrollments by female students are on-average 5 percentage points less likely to result in a passed exam compared to those by males. This gap is nearly four-fold larger (≈23 pp) for students from low-socioeconomic (SES) or underrepresented minority (URM) backgrounds compared to their respective demographic complement group.

Figure 7a presents the difference-in-differences estimates stratified by the subgroups. I find that the gains are not statistically different across demographic groups. While

[33] I caution against strong causal interpretation of most of these estimates, however. As indicated by asterisks (*), a large subset of AP subjects lacked a direct comparison group of unincentivized teachers in the sample, meaning their estimated effects are not identified from a within-subject comparison.

point estimates suggest relatively larger gains for female and low-SES students, the confidence intervals largely overlap.[34] Although weak, the differential gain among female AP students is consistent with prior literature suggesting they may better utilize additional schooling inputs due to differences in non-cognitive skills (Angrist et al., 2009; Jacob, 2002).

### C. *Heterogeneity by Locale*

A key strength of the data is the within-teacher variation in teaching environments, which allows examination of heterogeneity across school locale. Specifically, I find that 421 teachers in the Analysis sample (approximately 15%) moved between City, Suburban, and Rural locales during the study period. And, among these movers, 249 made a permanent move to a different locale, with moves occurring in all directions. This mobility enables separate identification of the average effects of teaching in specific locales from the teacher fixed effects.[35]

Figure 7b exhibits that the bonus effects were significantly higher for AP enrollees at rural schools, increasing the probability of passing the linked AP exam by nearly 4 percentage points. The program had imprecise but sizable positive impact within the city and suburban schools as well. The within-teacher model yields conservative estimates, while still estimating a statistically significant 2.7 percentage point increase in rural schools (see Appendix Figure B.8). These estimates suggest that performance pay generates larger gains in under-resourced areas, reducing the persistent urban-rural AP achievement gaps by approximately one-third.[36]

### D. *Heterogeneity by Teacher Experience*

[34] More specifically, the interaction terms are not statistically significant (see Panel A of Appendix Table B.10). However, estimates from the within-teacher model (reported in Panel A of Appendix Table B.10) reveal evidence of heterogeneity by gender. The estimates suggest that the bonus increased pass rate among female enrollments by an additional 2.8 percentage points ($p < 0.01$) relative to male enrollments.

[35] Note that 25 out of the 57 teachers in the unincentivized group were observed to teach AP courses in rural schools at least one year in the pre-bonus period, alleviating concerns regarding the identification of the difference-in-difference parameters.

[36] In the analysis sample, the pre-bonus AP exam pass rate among rural school enrollments were 6.95 percentage points (14%) lower than those in city and suburban schools, implying that the estimated effects close roughly one-third of the baseline urban-rural gap.

While Figure 3 indicates substantial overlap between the experience distribution of incentivized and unincentivized teachers, the difference in average experience raises a potential concern regarding whether the main estimates conflate the policy's impact with differential returns to teaching experience. In particular, if the bonus effects were concentrated entirely among novice teachers—who are modestly overrepresented within the incentivized group—the causal interpretation would be weakened.

To probe this directly, I estimate the DD model allowing the bonus effect to vary across three experience categories: novice (<5 years), Early-Career (5-9 years), and Veteran (10 or more years).[37] Figure 7c shows that, while the estimated bonus effect is larger among novice teachers, the estimated effects for groups with 5 or more years of teaching experience are close to the average effect and statistically significant at 10% level.[38] And because the incentive appears effective for both novice and more experienced teachers, this alleviates the concern that the average effect is primarily driven by the over-representation of less experienced teachers in the incentivized group.

# 6 Discussion

A key feature of the empirical design is the use of pre-bonus averages to proxy for the teachers' baseline class characteristics and expected incentive exposure. Because actual bonus payments are not observed, I infer incentives from the known policy formula, under which bonuses are a deterministic function of the number of students achieving a passing score, subject to a cap. Accordingly, a teacher's average number of passers in the pre-bonus period provides a natural proxy for baseline productivity and expected bonus exposure, while ensuring that the classification of incentive status is not mechanically affected by post-policy outcomes. This, however, introduces measurement error, as these

[37] This categorization is well-supported by the prior literature, suggesting teachers' learning curve is steepest in the first 5 years Papay and Kraft (2015) and Wiswall (2013).

[38] Table B.11 presents the background raw estimates for Figure 7c, where veteran teachers serve as the reference group. I estimate a statistically significant differential effect for novice teachers, implying the bonus had been most effective for this group. While this could also be partially driven by higher returns to experience for the early-career teachers, the crucial finding is that the bonus yields positive total effects across all experience levels. Moreover, these results remain robust to a stricter definition of treatment group. Defining incentivized teachers using the 65 average pre-waiver passers threshold expands the control group to 78 teachers (28,310 enrollments); however, the results remain stable (see Appendix Figure B.9 and Figure B.10).

averages may imperfectly capture a teacher's true baseline, particularly in the post-bonus years. Such noise can lead to misclassification of teachers' incentivized status—some teachers classified as unincentivized may in fact face positive incentives, while some classified as incentivized may be effectively capped-out. This misclassification attenuates the difference-in-differences estimates toward zero. Reassuringly, the consistency of the core findings across multiple baseline proxies (e.g., pass rate, number of passers, number of non-passers) and the robustness across various specifications instill confidence.

Additionally, extensive robustness and sensitivity checks in Section 5.1 establish that the increase in pass rates among AP enrollees is not driven by AP exam fee waivers. However, it is important to emphasize that the success of the teacher bonus program is achieved under a universal no-fee testing regime.[39]

This paper explicitly isolates the intensive margin of teacher response: how financial incentives alter the effort and strategy of incumbent AP teachers. In this paper, I do not estimate the policy's effects on the extensive margin of labor supply, such as recruitment or retention. However, economic theory predicts that performance pay is likely to improve workforce quality through positive selection (Lazear, 2000). Though I cannot directly verify, I provide evidence supporting this notion: the estimates from the Matched Sample (Table B.12, Panel B), which includes all AP teachers, are consistently larger than the estimates from samples restricted to continuing teachers (Panels C and D). Consequently, it is plausible that the policy's total effect is likely larger than the estimates suggest.

Finally, back-of-the-envelope calculations suggest that the bonus program may be highly cost-effective. Under a set of simplifying assumptions that allow comparison of the policy's benefits and costs, the expected increase in future tax revenues may fully offset the program's upfront cost. Comparing an enrollee's expected lifetime after-tax benefit to the expected cost of the bonuses yields a benefit–cost ratio of 4.92.[40]

[39] A further unique feature of this policy context is that AP credits carry high signaling value for students during the college admissions process. As a result, students may have strong intrinsic motivation to achieve a passing score, which may in turn reinforce the effectiveness of teacher responses to the financial incentives.

[40] Bartik et al. (2016) estimate that a college degree increases the net present value (NPV) of lifetime earnings by \$846,300 for non-economically disadvantaged (non-EDS) students and \$314,800 for EDS students (in 2012 dollars). Weighting these figures by the pre-policy EDS share in the Analysis sample (14%) yields an average lifetime earnings increase of \$771,890. Assuming a combined marginal tax rate of 24% (applying a flat 20% federal income tax and the 4% North Carolina state income tax), the after-tax benefit to the student is 76% of this total. Because passing an AP exam increases the probability

In the framework of Hendren and Sprung-Keyser (2020), policy efficiency is summarized by the Marginal Value of Public Funds (MVPF), which accounts for fiscal externalities. Under these assumptions, the implied MVPF is infinite, as projected tax revenues exceed program costs. While these calculations rely on strong assumptions, they provide suggestive evidence that the program represents a highly efficient investment in human capital.[41]

# 7 Concluding Remarks

This paper examines a novel piece-rate incentive program that awards Advanced Placement (AP) teachers \$50 for each student in their class achieving a score of 3 or higher on the standardized AP exam—typically eligible for college credit. Exploiting both within- and across-teacher variation in pass rate, I document the policy increased the probability that an AP enrollee passes the corresponding AP exam by 2.4 percentage points. I further show that this effect was driven by a broad-based improvement in achievement.

This average effect, however, masks significant heterogeneity. I find that the bonus was less (more) effective for teachers with a high (low) baseline number of passers, a finding that contrasts with the literature on group-based bonuses (Goodman & Turner, 2013; Imberman & Lovenheim, 2015) and highlights the uniqueness of the marginal incentives of a capped, piece-rate design. I also find that the effectiveness of the bonus varied across school locale, suggesting that the bonuses reduced the urban-rural AP achievement gap by one-third. Taken together, the findings demonstrate that the impact of performance

of college completion by an estimated 1 percentage point (Smith et al., 2017), the expected increase in the NPV of after-tax lifetime earnings for passing one additional exam is \$5,866.36. The per enrollment expected benefit of the bonuses can be achieved by scaling this benefit by the estimated effect of 2.4 percentage point, yielding \$140.79. Given post-bonus pass rate of 57.20 percent, the per enrollment program cost is \$28.60. The benefit-cost ratio is therefore 4.92.

[41] To calculate the MVPF, we must account for the government's 24% tax recovery on the gross earnings increase. The expected tax recovery per enrollment is \$44.46 (=\$771,890×0.01× 0.24× 0.024). Because this expected tax recovery (\$44.46) exceeds the expected upfront cost (\$28.60), the net cost to the government is negative, resulting in an infinite MVPF. Importantly, this estimate may understate the true benefits, as it conservatively assumes AP credits generate no direct tuition or time savings for students once in college and do not reduce public higher-education subsidies. Conversely, it may overstate the MVPF since the administrative overhead of implementing the bonus program and the heterogeneity of AP subjects' respective impacts on college completion are not accounted for.

pay depends critically on contract design, the distribution of incentive strength across agents, and the underlying production technology of the task being rewarded.

## References


Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136–163 (cited on page 23).

Ballou, D., & Springer, M. G. (2017). Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains. *Education Finance and Policy*, *12*(1), 77–106 (cited on page 5).

Barrera-Osorio, F., & Raju, D. (2017). Teacher performance pay: Experimental evidence from pakistan. *Journal of Public Economics*, *148*, 75–91 (cited on page 2).

Bartik, T. J., Hershbein, B., & Lachowska, M. (2016). The merits of universal scholarships: Benefit-cost evidence from the kalamazoo promise. *Journal of Benefit-Cost Analysis*, *7*(3), 400–433 (cited on page 25).

Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning learning incentives of students and teachers: Results from a social experiment in mexican high schools. *Journal of Political Economy*, *123*(2), 325–364 (cited on page 2).

Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M. A., & Springer, M. G. (2025). Taking teacher evaluation to scale: The effect of state reforms on achievement and attainment. *Journal of Political Economy Microeconomics*, *3*(3), 568–610 (cited on page 2).

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). *Difference-in-differences with a continuous treatment* (tech. rep.). National Bureau of Economic Research. (Cited on page 20).

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, *112*(2), 407–441 (cited on page 6).

Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2024). On binscatter. *American Economic Review*, *114*(5), 1488–1514 (cited on pages 11, 33, A16).

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679 (cited on page 2).

Chingos, M. M., & West, M. R. (2012). Do more effective teachers earn more outside the classroom? *Education Finance and Policy*, *7*(1), 8–43 (cited on page 2).

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from impact. *Journal of Policy Analysis and Management*, *34*(2), 267–297 (cited on page 2).

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American economic review*, *102*(4), 1241–1278 (cited on page 2).

Farber, H. S. (2005). Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of political Economy*, *113*(1), 46–82 (cited on page 6).

Fazlul, I., Jones, T., & Smith, J. (2021). College credit on the table? advanced placement course and exam taking. *Economics of Education Review*, *84*, 102155 (cited on pages 9, 18).

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from new york city public schools. *Journal of Labor Economics*, *31*(2), 373–407 (cited on page 2).

Fryer Jr, R. G., Levitt, S. D., List, J., & Sadoff, S. (2022). Enhancing the efficacy of teacher incentives through framing: A field experiment. *American Economic Journal: Economic Policy*, *14*(4), 269–299 (cited on pages 2, 5).

Gillborn, D., & Youdell, D. (1999). *Rationing education: Policy, practice, reform, and equity*. McGraw-Hill Education (UK). (Cited on page 18).

Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the new york city bonus program. *Journal of Labor Economics*, *31*(2), 409–420 (cited on pages 2, 4, 26).

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466–479 (cited on page 2).

Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly journal of economics*, *135*(3), 1209–1318 (cited on page 26).

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, *58*(12), 2151–2167 (cited on page 5).

Imberman, S. A. (2015). How effective are financial incentives for teachers? *IZA World of Labor* (cited on page 2).

Imberman, S. A., & Lovenheim, M. F. (2015). Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics*, *97*(2), 364–386 (cited on pages 2, 4, 26).

Jackson, C. K. (2010). A little now for a lot later: A look at a texas advanced placement incentive program. *Journal of Human Resources*, *45*(3), 591–639 (cited on page 6).

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107 (cited on pages 2, 6).

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education review*, *21*(6), 589–598 (cited on page 23).

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of public Economics*, *89*(5-6), 761–796 (cited on page 5).

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979–2011 (cited on page 5).

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361 (cited on page 25).

Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2019). Pay by design: Teacher performance pay design and the distribution of student achievement. *Journal of Labor Economics*, *37*(3), 621–662 (cited on pages 2, 5).

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly journal of economics*, *134*(3), 1627–1673 (cited on page 2).

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of political Economy*, *119*(1), 39–77 (cited on pages 2, 5).

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, *92*(2), 263–283 (cited on pages 5, 18).

North Carolina Department of Public Instruction. (2014). Report to the north carolina general assembly: Broaden successful participation in advanced courses (date: November 15, 2014). Retrieved January 5, 2025, from https://webservices.ncleg.gov/ViewDocSiteFile/16688 (cited on pages 7, 17).

North Carolina Department of Public Instruction. (2015). Report to the north carolina general assembly: Broaden successful participation in advanced courses (date: November 15, 2015). Retrieved January 5, 2025, from https://webservices.ncleg.gov/ViewDocSiteFile/16770 (cited on pages 7, 17).

North Carolina Department of Public Instruction. (2016). Report to the north carolina general assembly: Broaden successful participation in advanced courses (date: December 15, 2016). Retrieved January 5, 2025, from https://www.dpi.nc.gov/documents/fbs/finance/reporting/2015-16-ap-legislative-report/download (cited on page 17).

North Carolina Department of Public Instruction. (2017). Report to the north carolina general assembly: Broaden successful participation in advanced courses (date: December 15, 2017). Retrieved January 5, 2025, from https://www.dpi.nc.gov/documents/fbs/finance/reporting/2016-17-ap-legislative-report/download (cited on page 17).

North Carolina Department of Public Instruction. (2019). Report to the north carolina general assembly: Broaden successful participation in advanced courses (date:december 15, 2019). Retrieved January 5, 2025, from https://webservices.ncleg.gov/ViewDocSiteFile/17108 (cited on pages 7, 17).

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119 (cited on pages 4, 24).

Rahman, M. T., & Lawson, M. C. (2025). *Removing barriers to college credits: Where and for whom ap exam fee waivers work* (EdWorkingPaper No. 25-1345). EdWorkingPaper, Annenberg Institute at Brown University. https://doi.org/10.26300/df8y-4q69 (cited on pages 17, 18).

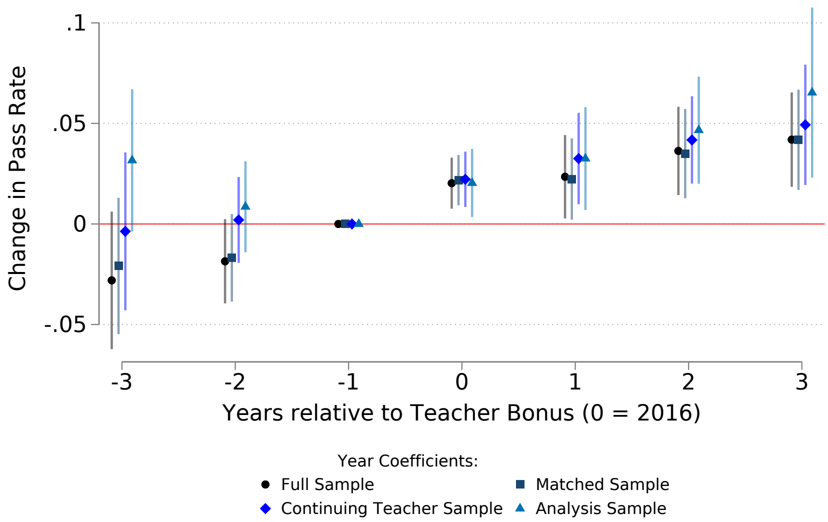
Smith, J., Hurwitz, M., & Avery, C. (2017). Giving college credit where it is due: Advanced placement exam scores and college outcomes. *Journal of Labor Economics*, *35*(1), 67–147 (cited on page 26).

Sojourner, A. J., Mykerezi, E., & West, K. L. (2014). Teacher pay reform and productivity: Panel data evidence from adoptions of q-comp in minnesota. *Journal of Human Resources*, *49*(4), 945–981 (cited on page 2).

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective teacher retention bonuses: Evidence from tennessee. *Educational Evaluation and Policy Analysis*, *38*(2), 199–221 (cited on page 2).

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, *100*, 61–78 (cited on pages 4, 24).

Figure 1—trends in ap exam pass rate




Notes.- The year-coefficients are estimated using Linear Probability model based on student-course level data, including all second-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The AY 2014-15 (2015) coefficient is normalized to zero.

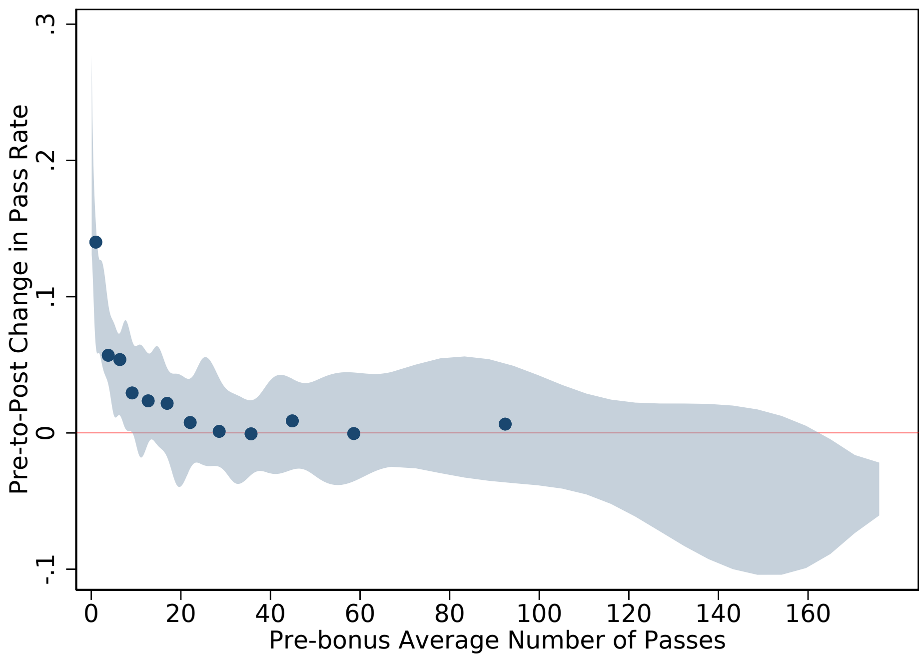
Table 1—summary statistics

| | **Full Sample** | | | **Analysis Sample** | | |
|---|---|---|---|---|---|---|
| | Pre 2013-2015 | Post 2016-2019 | Difference (se) | Pre 2013-2015 | Post 2016-2019 | Difference (se) |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Demographic Characteristics** | | | | | | |
| Female | 0.563 | 0.568 | 0.006(0.001) | 0.558 | 0.566 | 0.008(0.002) |
| Low-SES | 0.171 | 0.146 | -0.025(0.001) | 0.139 | 0.118 | -0.021(0.001) |
| Race | | | | | | |
| White | 0.689 | 0.666 | -0.023(0.001) | 0.706 | 0.691 | -0.015(0.002) |
| Black | 0.123 | 0.114 | -0.009(0.001) | 0.101 | 0.093 | -0.008(0.001) |
| Asian | 0.079 | 0.096 | 0.017(0.001) | 0.095 | 0.109 | 0.014(0.001) |
| Hispanic | 0.068 | 0.084 | 0.016(0.001) | 0.059 | 0.070 | 0.010(0.001) |
| Other | 0.033 | 0.034 | 0.002(0.000) | 0.033 | 0.032 | -0.001(0.001) |
| **B. Academic Variables** | | | | | | |
| Course Grade Point | 3.040 | 3.216 | 0.176(0.002) | 3.139 | 3.300 | 0.161(0.003) |
| Exam Participation | 0.831 | 0.937 | 0.106(0.001) | 0.893 | 0.953 | 0.060(0.001) |
| Exam Pass | 0.462 | 0.507 | 0.045(0.001) | 0.544 | 0.572 | 0.028(0.002) |
| Exam Score | 2.309 | 2.574 | 0.265(0.003) | 2.620 | 2.781 | 0.161(0.005) |
| 8th Grade Math Score | 1.042 | 1.070 | 0.028(0.002) | 1.160 | 1.197 | 0.037(0.002) |
| 8th Grade Reading Score | 0.915 | 0.940 | 0.025(0.002) | 0.999 | 1.026 | 0.027(0.002) |
| Observations | 307,580 | 467,036 | | 158,916 | 224,644 | |
| Number of Teachers | 5,173 | 5,269 | | 2,712 | 2,712 | |

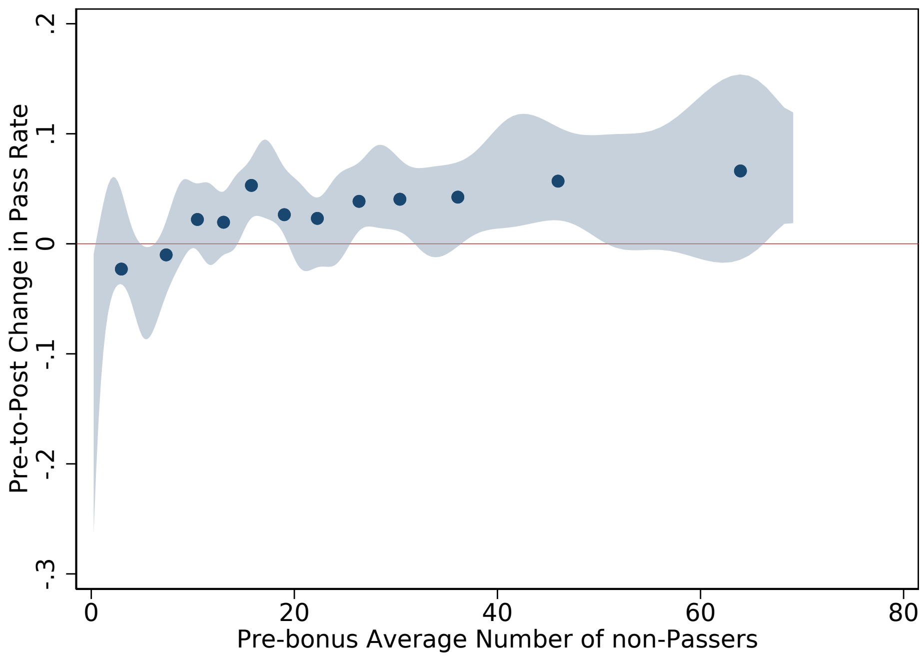
Notes.– This table reports the summary statistics from student-course level data including all Spring-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The ‘Analysis Sample’ is restricted to students who have taken two or more AP courses with continuing AP teachers—i.e., those who were observed teaching in both pre- and post-bonus periods. The 8th Grade Math and Reading scores are standardized within year×subject cells.

Figure 2—relationship between a teacher's pre-bonus average class characteristics and pre-to-post-bonus change in pass rate

(a) (Baseline) Pre-bonus Average Number of Passers

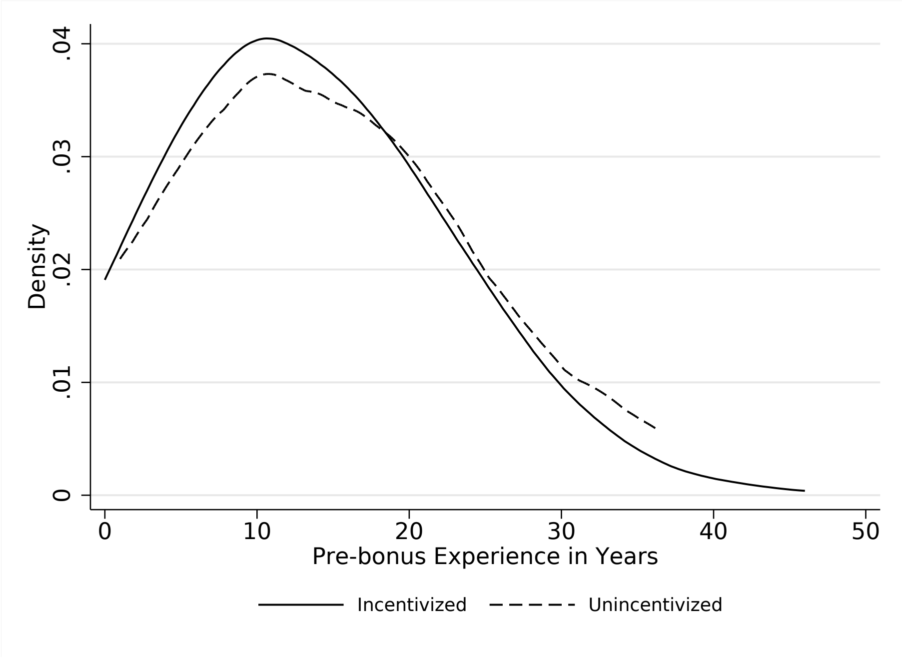



(b) (Baseline) Pre-bonus Average Number of Non-passers



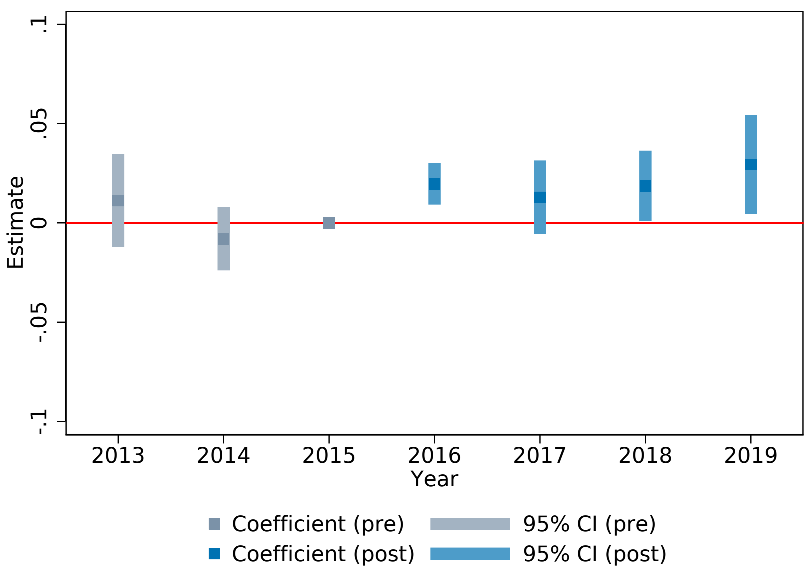

Notes.- The binned scatter plots illustrate the relationship between teachers' pre-bonus average number of passers (non-passers) and their pre-to-post change in pass rate. The plot uses teacher-level aggregate data from the sample of 2,712 continuing AP teachers, where the pre-bonus average number of passers (non-passers) are computed from the "Continuing Teacher Sample" and the pre-to-post waiver change in pass rates is computed from the "Analysis Sample" to maintain consistency with the main estimates. The number of bins are selected using the data-driven procedure of Cattaneo et al. (2024).

Figure 3—distribution of teacher experience
(incentivized versus unincentivized teachers)




Notes.- This figure plots the kernel density estimates of initial (pre-bonus) teaching experience among the incentivized and unincentivized teachers separately. The sample is restricted to the 2,712 continuing teachers observed teaching AP courses in both the pre- and post-bonus periods. Densities are estimated using an Epanechnikov kernel with a 5-year bandwidth.

Figure 4—event coefficient estimates from the dynamic within-teacher specification



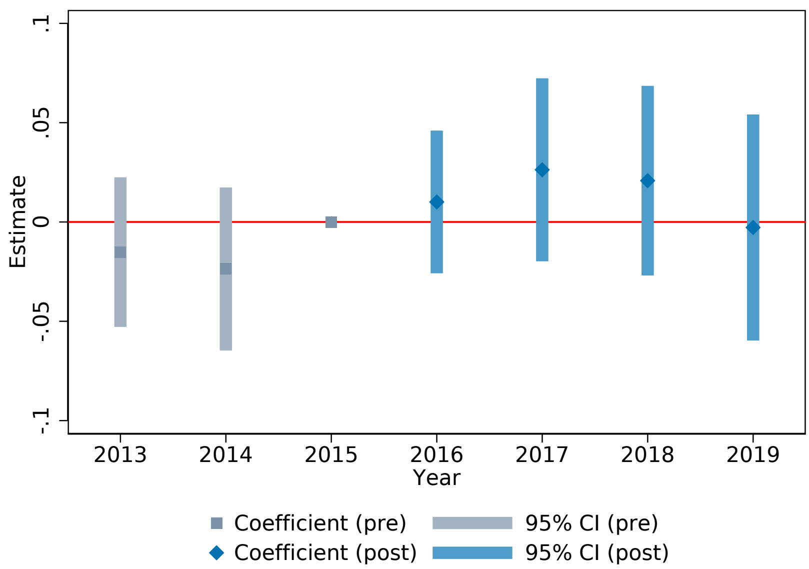

Notes.- This figure plots the estimated year coefficients from the dynamic within teacher model (Eq. 2), including teacher, high school grade, course, and student fixed effects. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods.

Table 2—Within-teacher model estimates

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Post | 0.020*** | 0.021*** | 0.019*** | 0.017*** | 0.026*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) |
| Pre-bonus Mean | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 |
| adj. $R^2$ | 0.20 | 0.20 | 0.21 | 0.33 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 | 383,560 |
| Teacher FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | ✓ | ✓ |
| Student FE | | | | | ✓ |
| Student Covariates | | | | ✓ | |

Notes.– This table reports regression estimates from the within-teacher model (Eq. 1). Observations are at the student-course (enrollment) level. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. Student covariates include sex, race, disability status, socioeconomic status, and 8th grade Math and Reading z-scores. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure 5—event coefficient estimates from dynamic difference-in-differences specification




Notes.– The coefficients are estimated using the dynamic difference-in-difference specification (Eq. 4), including teacher, course, and student fixed effects. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. Teachers with pre-bonus average number of passers below 70 are categorized as 'incentivized'.

Table 3—difference-in-differences estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Post×Incentivized$_j$ | 0.023** | 0.028** | 0.030*** | 0.024* |
| | (0.011) | (0.011) | (0.011) | (0.012) |
| Pre-bonus Mean | 0.52 | 0.52 | 0.52 | 0.52 |
| adj. $R^2$ | 0.20 | 0.21 | 0.21 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Teacher FE | ✓ | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Student FE | | | | ✓ |

Notes.– This table reports regression estimates from the difference-in-differences model (Eq. 3). Observations are at the student-course level. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. Teachers with pre-bonus average number of passers below 70 are categorized as 'incentivized'. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 4—effects on ap score distribution

| | **AP Exam Score:** | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| **A. Non-cumulative**: Outcome = $\mathbb{I}$(Score$= k$) | | | | | |
| Post×Incentivized$_j$ | -0.007 | -0.011 | -0.015* | 0.020** | 0.019** |
| | (0.005) | (0.007) | (0.008) | (0.008) | (0.009) |
| Pre-bonus Mean | 0.16 | 0.21 | 0.23 | 0.18 | 0.11 |
| adj. $R^2$ | 0.45 | 0.22 | 0.13 | 0.15 | 0.37 |
| **B. Cumulative**: Outcome = $\mathbb{I}$(Score$\geq k$) | | | | | |
| Post×Incentivized$_j$ | 0.006 | 0.013 | 0.024* | 0.039*** | 0.019** |
| | (0.012) | (0.013) | (0.012) | (0.010) | (0.009) |
| Pre-bonus Mean | 0.89 | 0.73 | 0.52 | 0.29 | 0.11 |
| adj. $R^2$ | 0.48 | 0.49 | 0.52 | 0.48 | 0.37 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 | 383,560 |

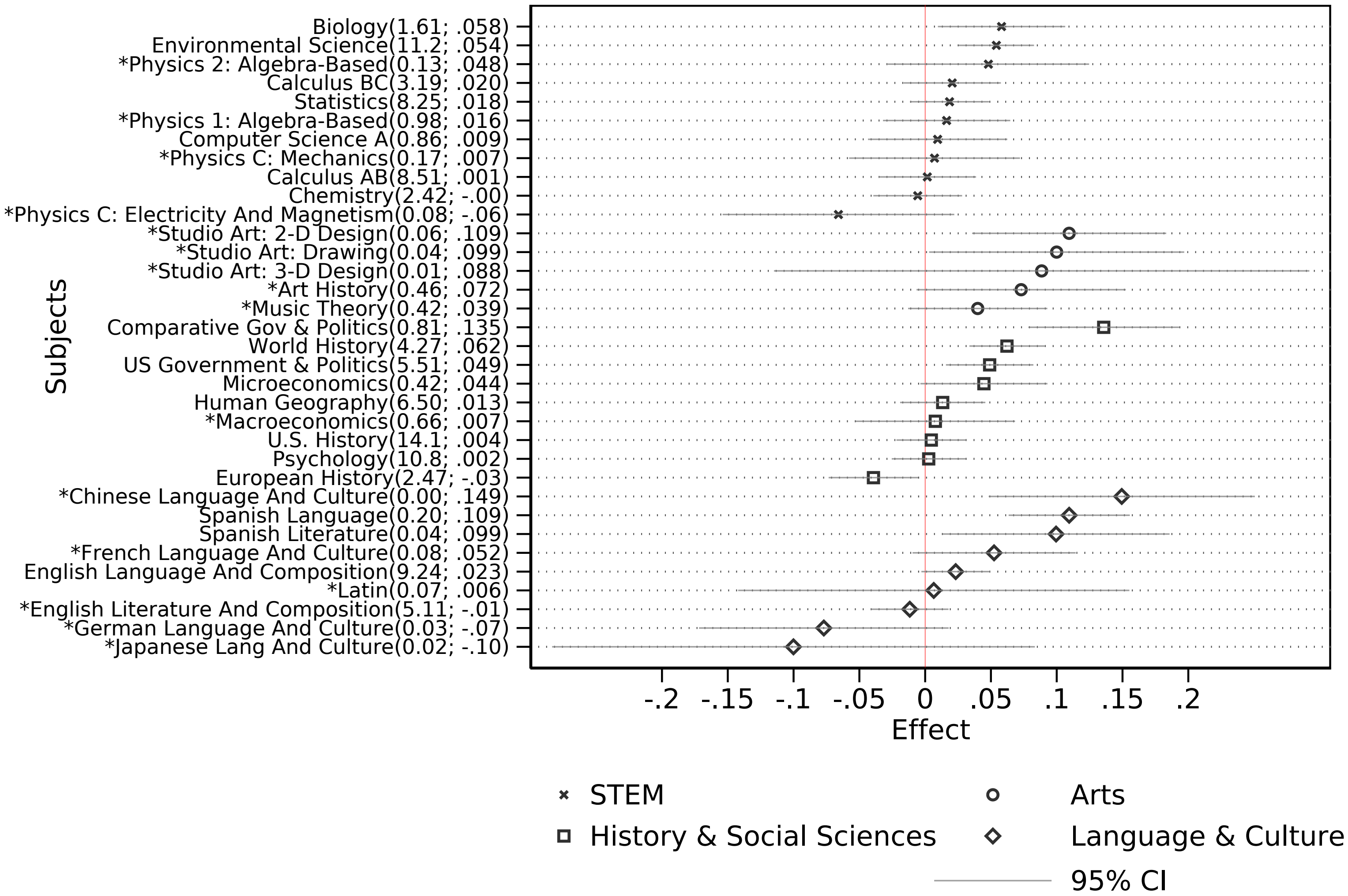
Notes.– This table reports regression estimates from the difference-in-differences specification (Eq. 3), including teacher, course, year, and student fixed effects. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. Teachers with pre-bonus average number of passers below 70 are categorized as 'incentivized'. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 5—heterogeneity by teachers' baseline number of passers and non-passers

| | (1) | (2) | (3) |
|---|---|---|---|
| Post$\times$(NPass$_{j,pre}$/100) | -0.063*** | | -0.081*** |
| | (0.011) | | (0.012) |
| Post$\times$(NFail$_{j,pre}$/100) | | 0.059*** | 0.090*** |
| | | (0.020) | (0.021) |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 |

Notes.– This table reports regression estimates from the continuous difference-in-differences specification (Eq. 3), including teacher, course, year, and student fixed effects. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

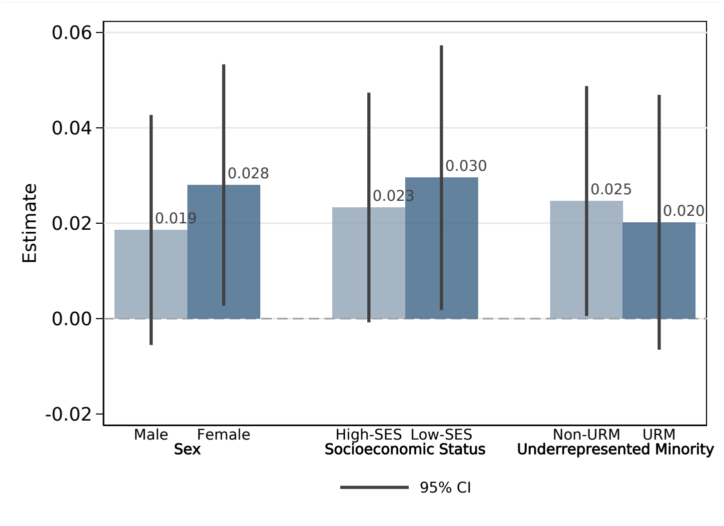
Figure 6—heterogeneity across AP subjects



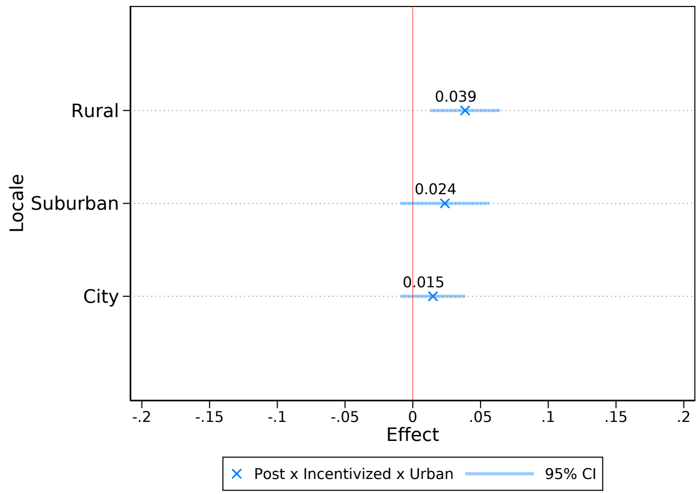

Notes.– This figure plots the regression estimates from the heterogeneous difference-in-differences specification (Eq. 5), with teacher, course, year, and student fixed effects. The first number in the parenthesis accompanying each subject title reports the pre-bonus share of student-courses (%) and the second number reports the point estimate for the given subject. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. Teachers with pre-bonus average number of passers below 70 are categorized as 'incentivized'. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level.

Figure 7—Heterogeneous Effects

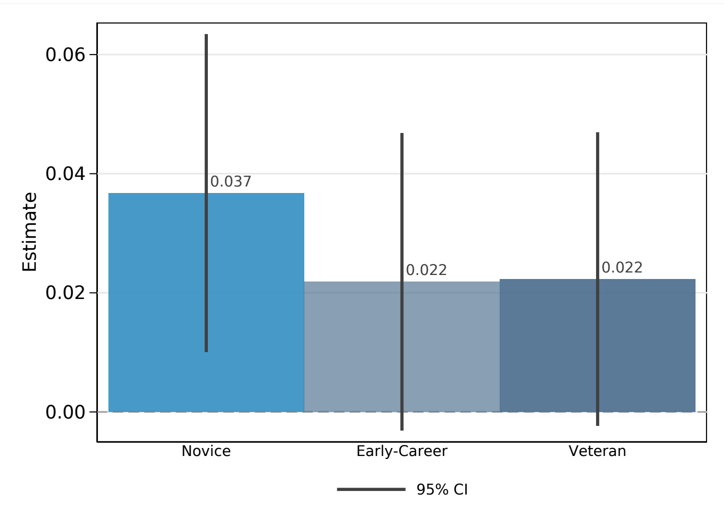
(a) by student demographics




(b) by locale




(c) by teacher experience




Notes.– This figure plots the regression estimates from the heterogeneous difference-in-differences specification (Eq. 5), including teacher, course, year, and student fixed effects. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. Teachers with pre-bonus average number of passers below 70 are categorized as 'incentivized'. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. Based on their initial pre-bonus experience, in Panel C, teachers are assigned to Novice (less than 5 years), Early-Career (5 to 9 years), or Veteran (10 or more years) categories.

# ONLINE APPENDIX

## A Theoretical Framework

This section develops a simple model to illustrate how optimal effort allocation across teachers change in response to the bonus design under study. Grounded in the human capital framework of Becker (1964), the model incorporates the multi-task perspective of Holmstrom and Milgrom (1991) and Barlevy and Neal (2012) by framing a teacher's utility maximization decision as an allocation of their total effort between two types of instructional activities: those that apply to the entire class (General Instruction) and those that can be targeted to specific students (Targeted Instruction). For tractability, the model assumes that teachers are homogeneous in their effectiveness and personal cost of effort. However, the model predictions are valid even if teachers vary in their effectiveness and costs but respond rationally to incentives. Additionally, this assumption is justified by empirical strategies employed in this paper, which leverage teacher fixed effects to absorb stable, cross-teacher differences.

A teacher is assigned to $S$ students, with their respective baseline ability, $a_i$, drawn from a distribution with probability density function, $\gamma(a)$.[42] Student $i$'s underlying academic achievement, the latent test score $z_i$, is a function of their own ability ($a_i$), the teacher's General instructional effort ($e_G$), and a random shock ($\epsilon_i$):

$$z_i = a_i + g(e_G) + \epsilon_i$$

Here, $g(e_G)$ captures learning from foundational, whole-class instruction and is assumed to be increasing and concave in general instructional effort, with $g'(\cdot) > 0$ and $g''(\cdot) < 0$. A teacher can also exert targeted instructional effort ($e_T$). This represents activities like individualized tutoring or small-group drills. While any effective targeted effort also builds a student's latent knowledge ($z_i$), a rational teacher responding to the bonus will direct this scarce resource toward students near the passing threshold—often referred to as "bubble students"—where it has the highest marginal product in generating a passing

[42] Without loss of generality, the baseline ability, $a_i$, in this case, may also capture a student's prior academic preparation and motivation.

score.[43] To cleanly model this strategic effort allocation, the model makes a deliberate simplifying abstraction: it treats the teacher's choice of targeted effort, $e_T$, as if it is a separate technology that specifically increases the pass probability for the bubble student group. This allows the model to frame a teacher's decision as a clear trade-off between a broad 'knowledge-building' strategy and a focused 'pass-harvesting' strategy.

A student is considered to have passed the exam if their latent achievement, $z_i$, is equal or greater than the passing threshold, $P$. To determine the expected number of passers for a given teacher, I group students into three categories based on their expected achievement, $a' = a + g(e_G)$, after receiving general instruction. The 'sure passers' are those with expected score equal or greater than the passing threshold $[a' \geq P]$, giving them a passing probability of 1. Students whose expected score falls just below the threshold $[a' \in [P - \Delta, P)]$ are the 'bubble students,' whose success is probabilistic and depends on the teacher's targeted effort.[44] The return to targeted effort on bubble students passing the test is captured by $f(e_T)$, with $f'(\cdot) > 0$ and $f''(\cdot) < 0$. Finally, the 'sure non-passers' are those with expected score below the bubble window $[a' < P - \Delta]$, and their passing probability is zero. This allows us to calculate the expected number of passers, $\mathbb{E}[\Pi]$, by integrating the passing probability for each group across the distribution of baseline abilities:

$$\mathbb{E}[\Pi] = S\left(f(e_T)\int_{P-\Delta-g(e_G)}^{P-g(e_G)} \gamma(a)da + \int_{P-g(e_G)}^{\infty} \gamma(a)da\right)$$

Here, the first term within the parentheses captures the number of passers generated from targeted instruction, and the second term captures the number of passers resulting from general instruction.

A teacher's objective is to maximize their expected utility, given by

$$\max_{e_G, e_T} \mathbb{E}[U] = W_0 + \min\{b \cdot \mathbb{E}[\Pi], \bar{B}\} - C(e_G, e_T; S) \tag{A1}$$

where, $W_0$ denotes the baseline salary, $b$ the piece-rate bonus, and $\bar{B}$ the bonus cap. A teacher's reward is the minimum of the piece-rate multiplied by the number of passes

[43] Effort dilution of this form is commonly referred to as 'educational triage,' a term coined by Gillborn and Youdell (1999).

[44] Assume $\Delta$ is strictly positive and very small.

and the bonus cap. The cost of effort, $C(\cdot)$, is increasing in both effort types and student load, $S$, reflecting the diseconomies of scale (e.g., grading). The economies of scale (e.g., lecture preparation) is captured on benefit side, as the potential bonus for any strategy scales with $S$.

$$FOC(e_G): \quad b \cdot S \cdot g'(e_G)\left[\gamma(P - g(e_G))(1 - f(e_T)) + f(e_T)\gamma(P - \Delta - g(e_G))\right] = \frac{\partial C}{\partial e_G}$$

$$FOC(e_T): \quad b \cdot S \cdot f'(e_T) \int_{P-\Delta-g(e_G)}^{P-g(e_G)} \gamma(a)da = \frac{\partial C}{\partial e_T}$$

Analyzing the first order conditions of the teacher's maximization problem—which state that optimal effort is exerted until marginal benefit equals marginal cost—yield a set of clear predictions.

**Proposition 1** *The bonus cap extinguishes marginal financial incentives for teachers of high-achieving classes.*

A teacher whose expected number of passers meet or exceed the cap has an utility function: $\mathbb{E}[U] = W_0 + \bar{B} - C(e_G, e_T; S)$. It follows that their objective switches from maximizing passes to minimizing cost to achieve the maximum reward, given by the derivative with respect to effort, $\partial\mathbb{E}[U]/\partial e = -\partial C/\partial e$.

**Proposition 2** *The bonus scheme incentivizes teachers to allocate targeted effort disproportionately toward the "bubble" students.*[45]

A rational teacher allocates costly effort to the margin where the marginal returns to effort is highest. Both FOCs highlight that effort is sensitive to the number of bubble students. $FOC(e_T)$ shows that the marginal benefit of the targeted effort is directly proportional to the share of bubble students, derived from the integral. On the other hand, $FOC(e_G)$ implies that the marginal benefit from general instruction is realized by increasing the number of net additional passers, collectively defined by the terms within brackets, through two mechanisms. First, additional $e_G$ increases the number of passers

[45]It can also be shown that the bonus scheme incentivizes teachers to engage in targeted instruction. Because effort is costly (i.e., $\partial C/\partial e_T > 0$), it follows from $FOC(e_T)$ that, if $b = 0$ then $e_T^* = 0$ for a rational teacher. A bonus ($b > 0$) makes the marginal benefit positive, thereby creating incentive for the teacher to utilize this focused instructional strategy ($e_T^* > 0$).

by pushing students at the cusp of passing, $\gamma(p - g(e_G))$, into the passing score range, net of the effect of targeted teaching, $(1 - f(e_T)$. Second, the density of students at the bottom of the bubble window, who are close enough to be promoted to the bubble through $e_G$, can then be converted into passes with the help of targeted teaching, $f(e_T)$.

**Proposition 3** *The optimal mix of instructional strategies depends on the teacher's student ability distribution.*

Note that, while marginal benefit of additional general instruction effort is sensitive to the density of students at the boundaries of the bubble window [i.e., $\gamma(P - g(e_G))$ and $\gamma(P - \Delta - g(e_G))$], the marginal benefit from targeted effort relies on the breadth of the bubble window. It follows that a utility maximizing teacher's ratio of optimal efforts will critically depend on the shape of the bubble window, or loosely, the baseline distribution (or variance) of student ability.

**Proposition 4** *The bonus provides lower incentive to teachers with high baseline performance.*

A teacher's 'high baseline performance' can be characterized by high baseline effectiveness (i.e., high ability to convert effort into student achievement) or by high baseline student achievement (i.e., a high pre-bonus pass rate). The bonus is less effective for these teachers due to two reasons. One, diminishing marginal returns to effort. The model assumes that the production functions for effort are concave ($g'' < 0,\ f'' < 0$). A teacher who is highly effective is likely already exerting high level of effort, placing them on a flatter part of the production curve. This means marginal product of their additional effort is lower ($g',\ f'$). As shown in the First order conditions, the marginal benefit of the bonus is directly proportional to these marginal products. A lower marginal product therefore translates into a lower marginal benefit, leading to a smaller increase in effort in response to the incentive. Two, proximity to the bonus cap. A high number of expected baseline passers ($\mathbb{E}[\Pi_0]$), driven either by a teacher's high baseline effectiveness or students' high baseline ability, reduces a their potential gain from bonus ($\bar{B} - b \cdot \mathbb{E}[\Pi_0]$) due to an existing cap and increases the likelihood that the teacher is in the capped out regime, where the marginal incentive is zero.

**Proposition 5** *The effect of student load (S) on teacher effort is theoretically ambiguous.*

The optimal effort, $e_k^*$ for $k \in \{G, T\}$, is defined implicitly by its FOC, which can be written as $H_k(e_k, S) = MB_k(e_k, S) - MC(e_k, S) = 0$. Here, MB and MC refer to Marginal Benefit and Marginal Cost, respectively. The implicit function theorem states that $de_k^*/dS = -(\partial H_k/\partial S)/(\partial H_k/\partial e_k)$. The denominator is negative by second order conditions; hence, the sign of the expression is determined by the sign of its numerator, $\partial H_k/\partial S = (\partial MB_k/\partial S) - (\partial MC_k/\partial S)$. From the model's assumptions both marginal benefit and marginal cost increase with student load ($\partial MB_k/\partial S > 0$, the economies of scale effect; and $\partial MC_k/\partial S > 0$, the diseconomies of scale effect). Thus, the effect of student load, $S$, on optimal effort is theoretically ambiguous.

**Proposition 6** *The bonus provides higher incentive to teachers with larger number of non-passing students.*

Denote the bubble window by $N_B$, such that $N_B = S\int_{P-\Delta-g(e_G)}^{P-g(e_G)} \gamma(a)da$. Let, $F_0$ denote the pool of non-passing students given baseline instructional effort. Because the bubble window is the upper boundary, $[P - \Delta, P)$, of this distribution, the absolute number of bubble students is an increasing function the size of the non-passing pool (i.e., $\frac{\partial N_B}{\partial F_0} > 0$).

For all $k \in \{G, T\}$, the implicit function theorem states that $de_k^*/dF_0 = -\frac{(\partial H_k/\partial F_0)}{(\partial H_k/\partial e_k)}$. The denominator is negative by the second order conditions. And given that the number of bubble students increase with the number of non-passers ($\frac{\partial N_B}{\partial F_0} > 0$), the numerator is positive. Therefore, the effect of the non-passing pool of students, $F_0$, incentivizes higher effort.

## Appendix References


Barlevy, G., & Neal, D. (2012). Pay for percentile. *American Economic Review*, *102*(5), 1805–1831 (cited on page A1).

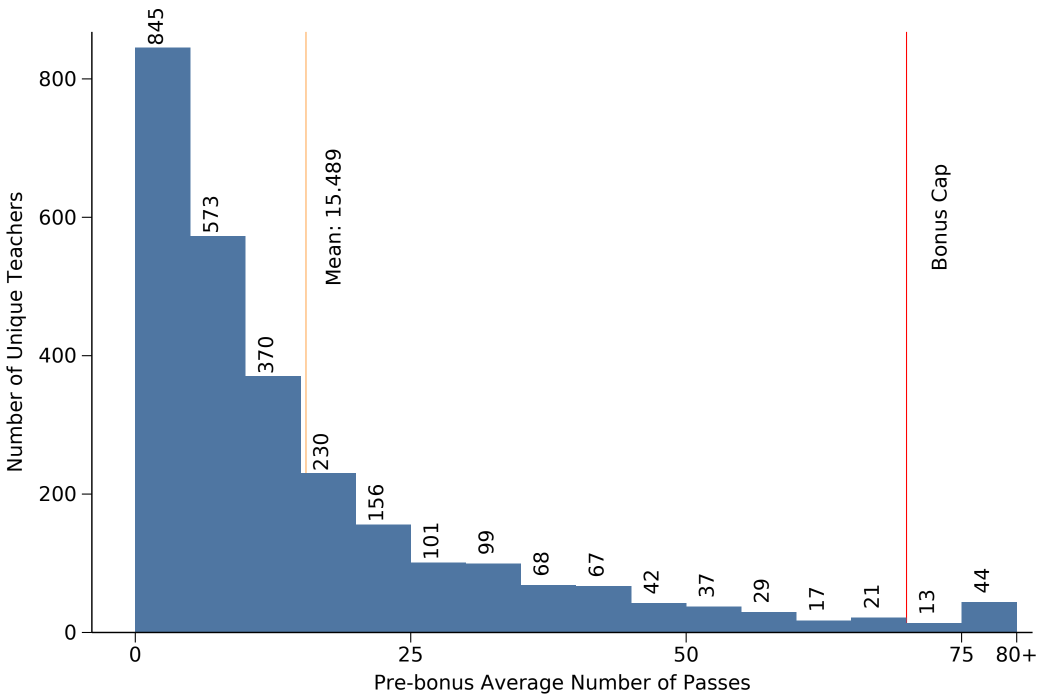
Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education* (1st). Columbia University Press. (Cited on page A1).

Gillborn, D., & Youdell, D. (1999). *Rationing education: Policy, practice, reform, and equity*. McGraw-Hill Education (UK). (Cited on page A2).

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, *7*(special_issue), 24–52 (cited on page A1).

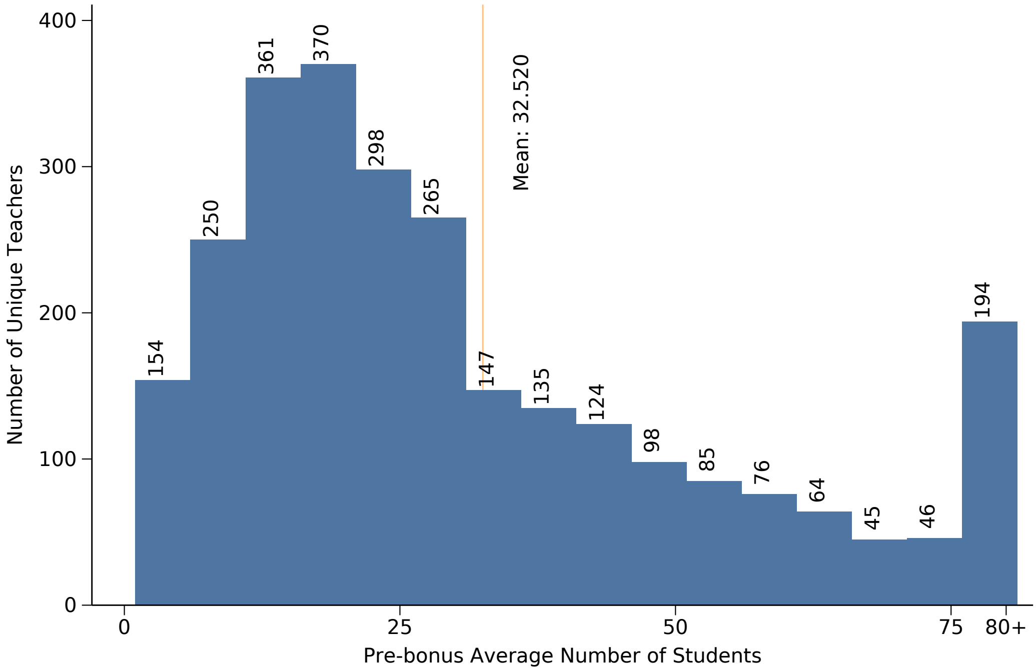
# B Supplementary Figures and Tables

Figure B.1—frequency distribution of teachers' average number of pre-bonus passers




Notes.- This figure plots the frequency distribution of the average number of passing students per teacher in the pre-bonus period (N = 2,712 continuing AP teachers). Data are grouped into 5-student bins, with the final bin representing teachers averaging 80 or more passing students.

Figure B.2—Frequency distribution of teachers' average number of pre-bonus students



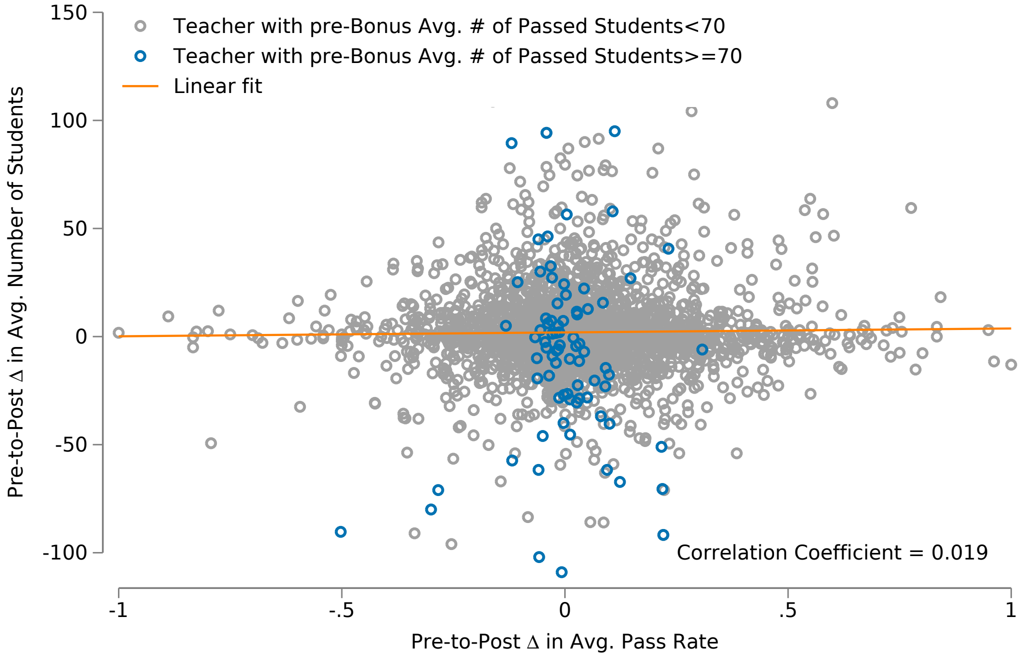

Notes.- This plot shows the frequency distribution of 2,712 continuing AP teachers' pre-bonus average student loads. Data are grouped into 5-student bins, with the final bin representing teachers averaging 80 or more average pre-bonus students assigned to a teacher.

Table B.1—robustness: treatment-cohort-specific estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\text{Post}_{t\geq 2016} \times \mathbb{I}(\text{NPass}_{j,pre} < 40)$ | 0.023*** | 0.025*** | 0.026*** | 0.019*** |
| | (0.007) | (0.007) | (0.008) | (0.007) |
| $\text{Post}_{t\geq 2017} \times \mathbb{I}(40 \leq \text{NPass}_{j,pre} < 70)$ | -0.014* | -0.010 | -0.010 | 0.002 |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| adj. $R^2$ | 0.20 | 0.21 | 0.21 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Teacher FE | ✓ | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Student FE | | | | ✓ |

Notes.– This table reports the treatment-cohort-specific estimates. Observations are at the student-course level. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.3—relationship between teachers' pre-to-post change in average student load and change in average pass rate



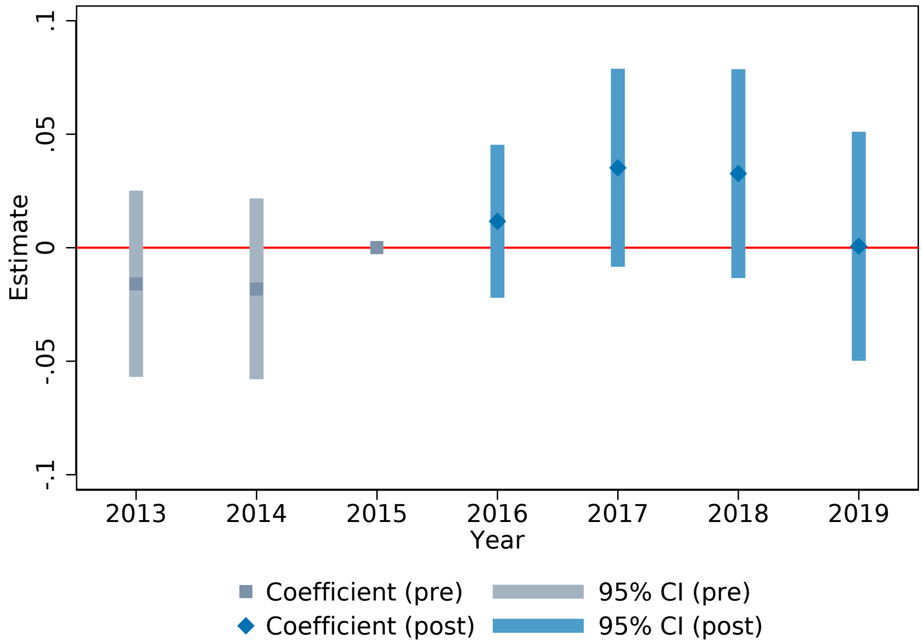

Notes.– This scatterplot uses teacher-level data to illustrate the relationship between the pre-to-post change in pass rate and the pre-to-post change in student load. The sample is restricted to the 2,712 teachers who have taught in both pre- and post-bonus periods.

Table B.2—NCAPP Robustness Check

| | Within-Teacher Model | | Difference-in-Differences Model | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Post | 0.025*** | 0.025*** | | |
| | (0.004) | (0.005) | | |
| Post×Incentivized$_j$ | | | 0.027*** | 0.028** |
| | | | (0.010) | (0.013) |
| adj. $R^2$ | 0.52 | 0.51 | 0.52 | 0.51 |
| Observations | 383,560 | 289,825 | 383,560 | 289,825 |
| Number of Unique teachers | 2,712 | 2,122 | 2,712 | 2,122 |
| NCAPP Target Indicator | No | No | No | No |
| NCAPP Non-Target Sample | No | Yes | No | Yes |

Notes.– This table reports the estimated coefficients of interest from the within-teacher and the difference-in-differences models, with full set of fixed effects. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The NCAPP Non-Target Sample excludes all observations from 34 school districts targeted by NCAPP at any year during the study period. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.4—event Coefficients from dynamic difference-in-differences specification without student fixed effects




Notes.– The coefficients are estimated using the dynamic difference-in-difference specification (Eq. 4), excluding student fixed effects and including teacher, grade, year, and course fixed effects. The sample includes all Spring-semester or year-long AP courses between AY 2013-14 and 2018-19, except AP Research and Capstone courses, and restricted to courses linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passers below 70.

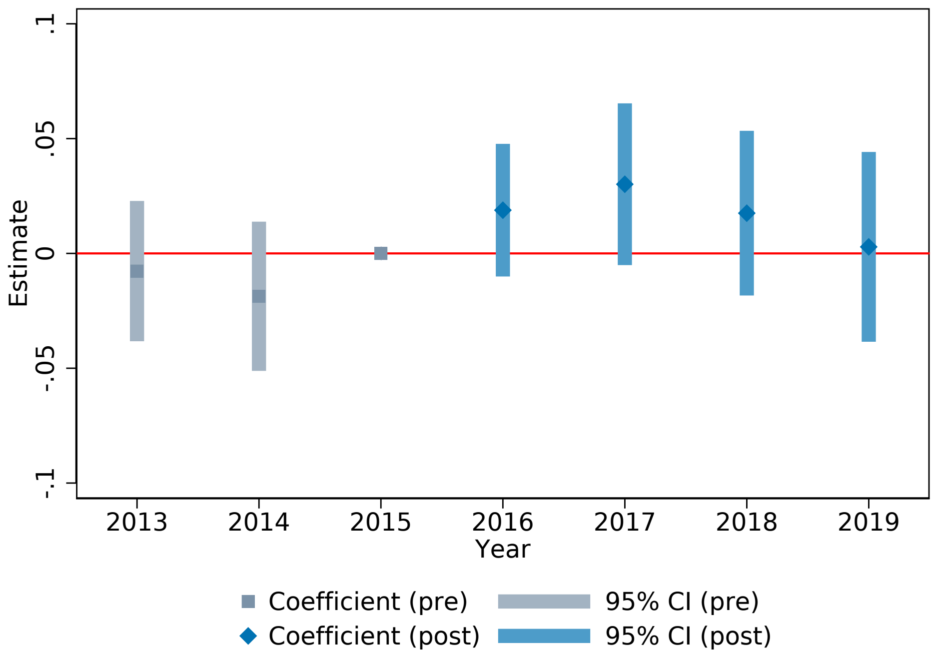
Table B.3—difference-in-differences model estimates using alternative thresholds to define unincentivized group

| | Outcome: AP Exam Pass [ $\mathbb{I}$(Exam Score$\geq$3) ] | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **A. Incentivized**$_j$ $= \mathbb{I}(\textbf{NPassed}_{j,pre} < 65)$ | | | | |
| Post×Incentivized$_j$ | 0.024** | 0.026*** | 0.029*** | 0.027*** |
| | (0.009) | (0.009) | (0.009) | (0.010) |
| Pre-bonus Mean | 0.51 | 0.51 | 0.51 | 0.51 |
| adj. $R^2$ | 0.20 | 0.20 | 0.21 | 0.52 |
| **B. Incentivized**$_j$ $= \mathbb{I}(\textbf{NPassed}_{j,pre} < 75)$ | | | | |
| Post×Incentivized$_j$ | 0.029*** | 0.036*** | 0.038*** | 0.036*** |
| | (0.007) | (0.007) | (0.008) | (0.009) |
| Pre-bonus Mean | 0.52 | 0.52 | 0.52 | 0.52 |
| adj. $R^2$ | 0.20 | 0.21 | 0.21 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Teacher FE | ✓ | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Student FE | | | | ✓ |

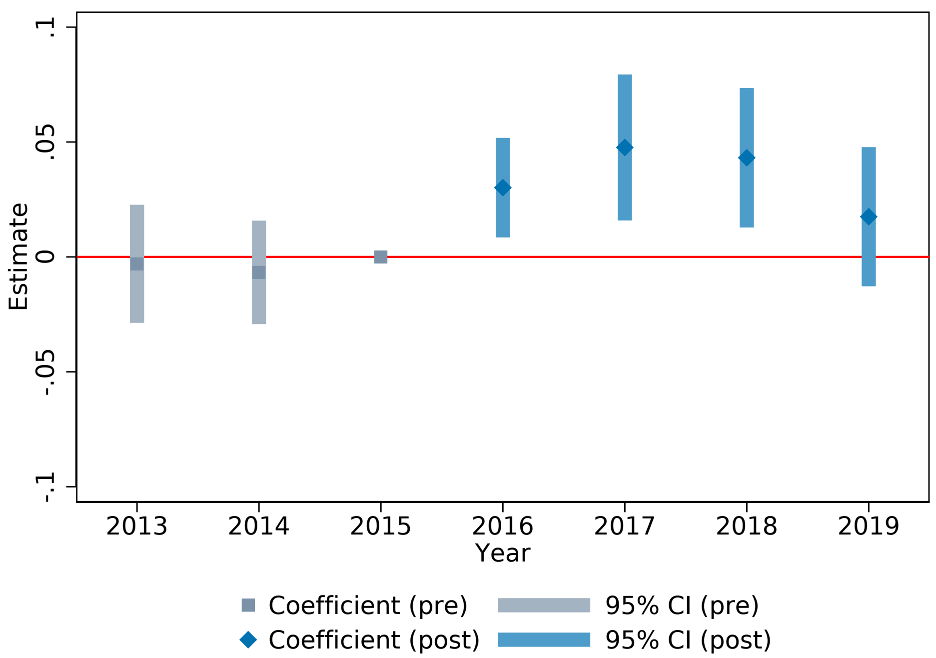
Notes.– This table reports regression estimates from the difference-in-differences model (Eq. 3). Observations are at the student-course level. The sample includes all Spring-semester or year-long AP courses between AY 2013-14 and 2018-19, except AP Research and Capstone courses, and restricted to student-course of multiple AP course enrollees linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passed exams below the indicated threshold in each panel. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.5—event coefficient estimates from dynamic difference-in-differences model using alternative thresholds to define unincentivized group

(a) Incentivized $= \mathbb{I}(\text{NPassed}_{j,pre} < 65)$




(b) Incentivized $= \mathbb{I}(\text{NPassed}_{j,pre} < 75)$




Notes.- The coefficients are estimated using Equation 4 and Linear probability model. The sample includes all Spring-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses, and restricted to courses linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passers below 60 in (a) and 70 in Panel (b).

Table B.4—difference-in-differences estimates from post-waiver sample (AY2014-15 through AY2018-19)

| | (1) | (2) | (3) |
|---|---|---|---|
| **A. Outcome: Exam Pass** | | | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 65)$ | 0.023 | | |
| | (0.015) | | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 70)$ | | 0.013 | |
| | | (0.019) | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 75)$ | | | 0.027** |
| | | | (0.012) |
| Pre-bonus Mean | 0.52 | 0.53 | 0.53 |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| **B. Outcome: Exam Participation** | | | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 65)$ | -0.008 | | |
| | (0.013) | | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 70)$ | | -0.017 | |
| | | (0.017) | |
| Post$\times$Incentivized$_j$ $= \mathbb{I}(\text{NPassed}_{j,pre} < 75)$ | | | 0.002 |
| | | | (0.005) |
| Pre-bonus Mean | 0.94 | 0.94 | 0.94 |
| adj. $R^2$ | 0.46 | 0.46 | 0.46 |
| Observations | 265,575 | 265,575 | 265,575 |
| Number of Unique teachers | 2,350 | 2,350 | 2,350 |
| Year FE | ✓ | ✓ | ✓ |
| Teacher FE | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ |
| Grade FE | | | |
| Student FE | ✓ | ✓ | ✓ |

Notes.– This table reports difference-in-differences estimates restricting the sample to the post fee waiver period. The sample includes all Spring-semester or year-long AP courses between AY 2014-15 and 2018-19, except AP Research and Capstone courses, and restricted to student-course of multiple AP course enrollees linked to 2,350 teachers – who are observed in both AY 2014-15 and later. Each column reports estimates based on different definition of 'incentivized' teachers. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table B.5—within-student model estimates

| | (1) | (2) | (3) |
|---|---|---|---|
| **A. Outcome: Exam Pass** | | | |
| Post | 0.020*** | 0.020*** | 0.021*** |
| | (0.005) | (0.004) | (0.005) |
| Pre-bonus Mean | 0.57 | 0.57 | 0.57 |
| adj. $R^2$ | 0.44 | 0.47 | 0.53 |
| **B. Outcome: Exam Participation** | | | |
| Post | 0.005** | 0.004 | 0.001 |
| | (0.002) | (0.002) | (0.003) |
| Pre-bonus Mean | 0.95 | 0.95 | 0.95 |
| adj. $R^2$ | 0.34 | 0.35 | 0.45 |
| Student FE | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ |
| Teacher FE | | | ✓ |
| Observations | 125,984 | 125,984 | 125,984 |
| Number of Students | 26,998 | 26,998 | 26,998 |

Notes.– The sample includes all Spring-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses, and restricted to courses taken by students – who are observed in both pre- and post-bonus periods. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

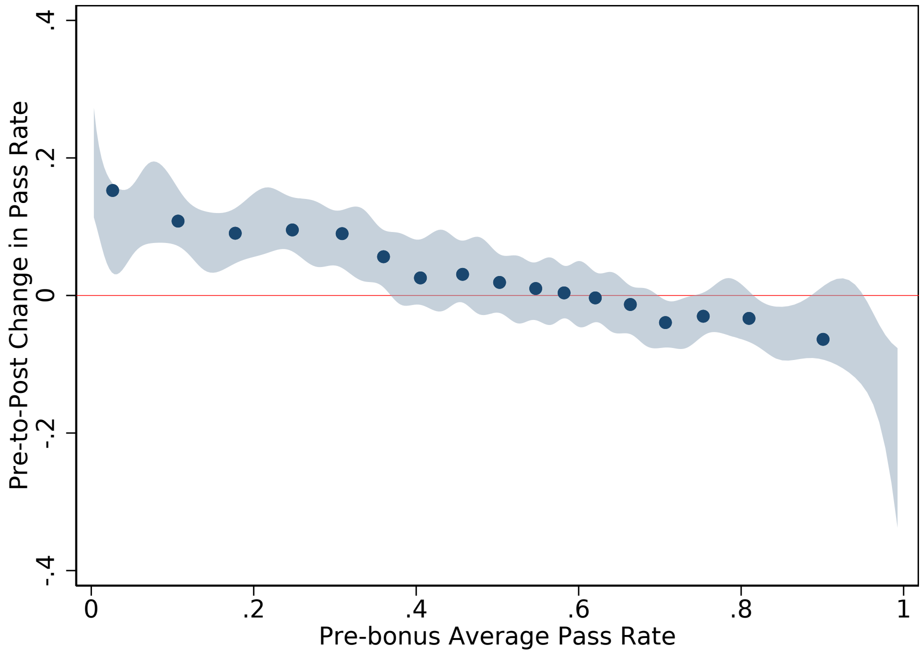
Table B.6—effects on ap score distribution (within-teacher model)

| | AP Exam Score: | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| **A. Non-cumulative**: Outcome = $\mathbb{I}$(Score= $k$) | | | | | |
| Post | -0.010*** | -0.003 | -0.003 | 0.013*** | 0.016*** |
| | (0.003) | (0.004) | (0.004) | (0.004) | (0.003) |
| Pre-bonus Mean | 0.15 | 0.20 | 0.23 | 0.19 | 0.12 |
| adj. $R^2$ | 0.45 | 0.22 | 0.13 | 0.15 | 0.37 |
| **B. Cumulative**: Outcome = $\mathbb{I}$(Score$\geq$ $k$) | | | | | |
| Post | 0.012*** | 0.022*** | 0.026*** | 0.029*** | 0.016*** |
| | (0.003) | (0.004) | (0.004) | (0.004) | (0.003) |
| Pre-bonus Mean | 0.89 | 0.75 | 0.54 | 0.31 | 0.12 |
| adj. $R^2$ | 0.48 | 0.49 | 0.52 | 0.48 | 0.37 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 | 383,560 |

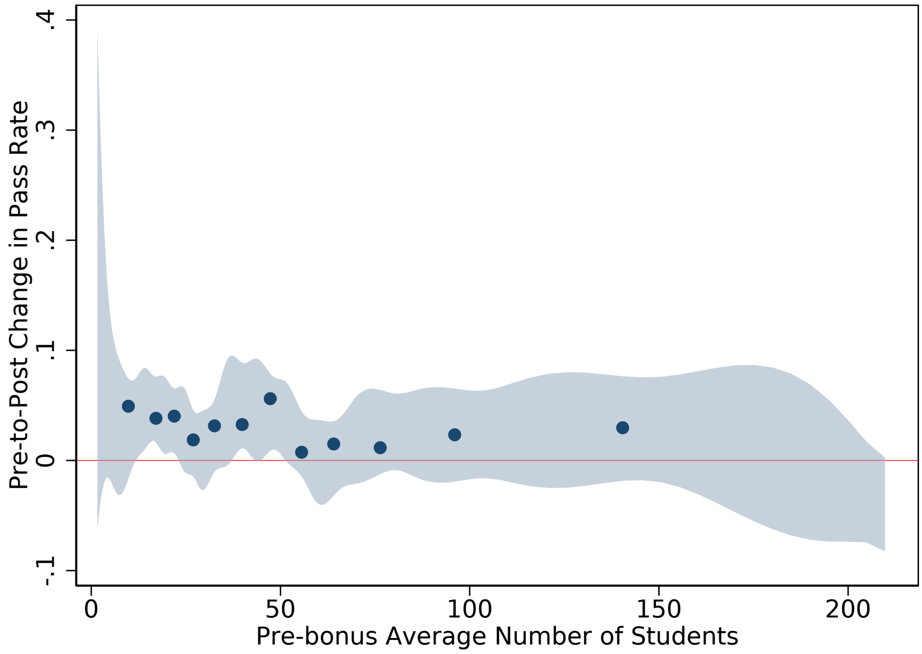
Notes.– This table reports regression estimates from the within-teacher model (Eq. 1). Observations are at the student-course (enrollment) level. The sample includes all Spring-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level.

Figure B.6—relationship between teachers' (alternative) baseline class characteristics and pre-to-post-bonus change in pass rate

(a) (Baseline) Pre-bonus Average Pass Rate




(b) (Baseline) Pre-bonus Average Number of Students




Notes.- This binned scatterplot illustrates the relationship between teachers' pre-bonus average number of passers and their pre-to-post change in pass rate. The plot uses teacher-level aggregate data from the sample of 2,712 continuing AP teachers. The number of bins and confidence intervals are selected using the data-driven procedure of Cattaneo et al. (2024).

Table B.7—distributional Effect Estimates from difference-in-differences model using continuing teacher sample

| | AP Exam Score: | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| **A. Non-cumulative**: Outcome = $\mathbb{I}$(Score= $k$) | | | | | |
| Post×Incentivized$_j$ | 0.000 | 0.004 | -0.004 | 0.020*** | 0.014* |
| | (0.006) | (0.005) | (0.007) | (0.007) | (0.007) |
| Pre-bonus Mean | 0.18 | 0.21 | 0.21 | 0.16 | 0.09 |
| adj. $R^2$ | 0.18 | 0.07 | 0.05 | 0.07 | 0.14 |
| **B. Cumulative**: Outcome = $\mathbb{I}$(Score$\geq$ $k$) | | | | | |
| Post×Incentivized$_j$ | 0.035*** | 0.035*** | 0.030*** | 0.034*** | 0.014* |
| | (0.010) | (0.011) | (0.011) | (0.010) | (0.007) |
| Pre-bonus Mean | 0.85 | 0.68 | 0.47 | 0.25 | 0.09 |
| adj. $R^2$ | 0.22 | 0.22 | 0.21 | 0.19 | 0.14 |
| Observations | 484,611 | 484,611 | 484,611 | 484,611 | 484,611 |

Notes.– This table reports regression estimates from the difference-in-differences model (Eq. 3), including teacher, grade, course, and year fixed effects. Observations are at the student-course level. The sample includes all Spring-semester or year-long AP courses between AY 2013-14 and 2018-19, except AP Research and Capstone courses, linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passed exams below 70. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

## Table B.8—heterogeneity by teachers' baseline (polynomial fit)

| | Outcome: AP Exam Pass= $\mathbb{I}$(Exam Score≥3) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Post×(NPass$_{j,pre}$/100) | -0.081*** | -0.198*** | -0.327*** |
| | (0.012) | (0.026) | (0.047) |
| Post×(NPass$_{j,pre}$/100)$^2$ | | 0.089*** | 0.328*** |
| | | (0.020) | (0.078) |
| Post×(NPass$_{j,pre}$/100)$^3$ | | | -0.107*** |
| | | | (0.032) |
| Post×(NFail$_{j,pre}$/100) | 0.090*** | 0.253*** | 0.296*** |
| | (0.021) | (0.033) | (0.085) |
| Post×(NFail$_{j,pre}$/100)$^2$ | | -0.182*** | -0.295 |
| | | (0.033) | (0.227) |
| Post×(NFail$_{j,pre}$/100)$^3$ | | | 0.073 |
| | | | (0.139) |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 |

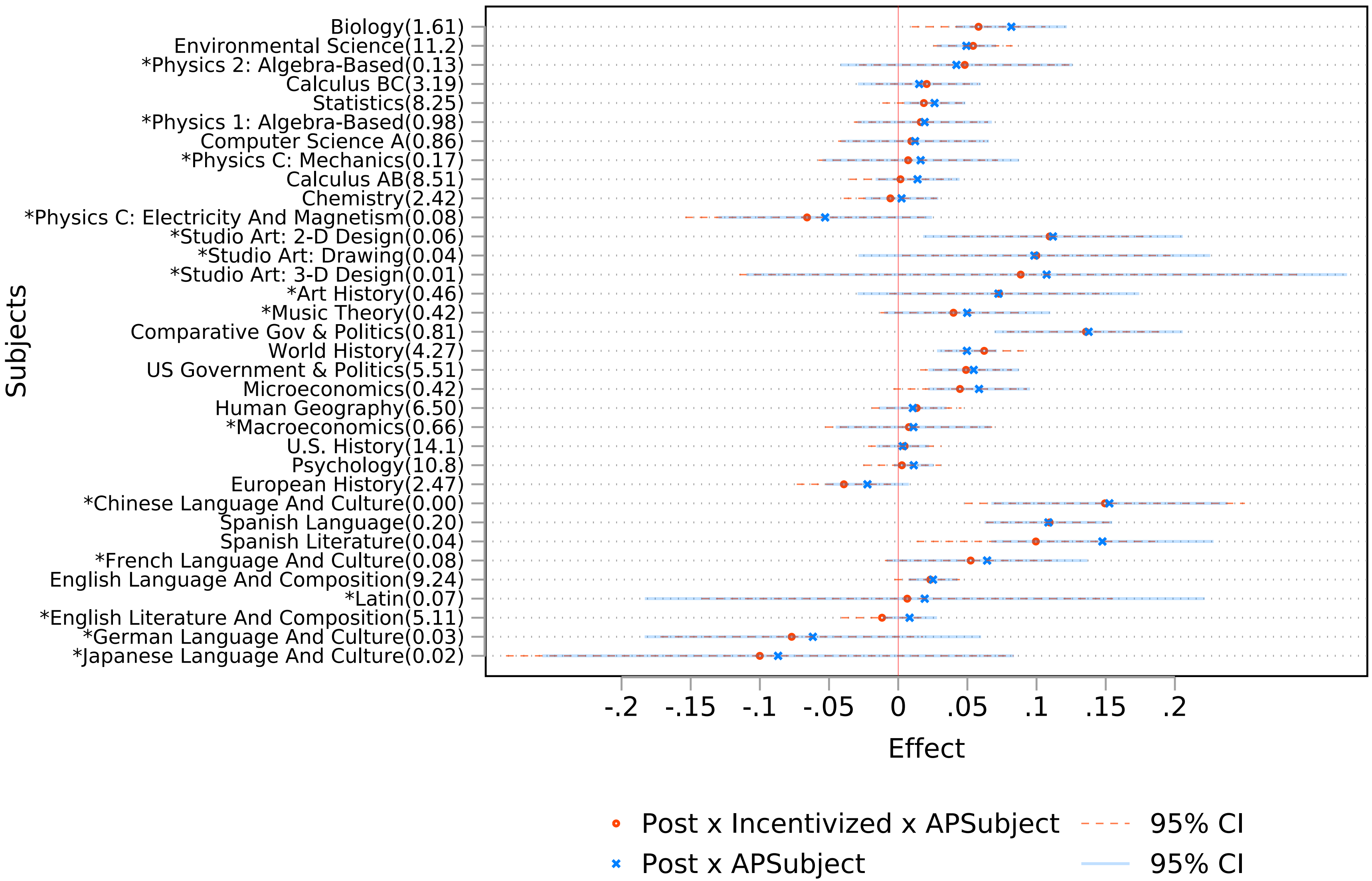
Notes.– This table reports regression estimates from the continuous difference-in-differences model, including higher-order polynomial of the interaction terms. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. Observations are at the student-course level. The sample includes all Spring-semester or year-long AP courses between AY 2013-14 and 2018-19, except AP Research and Capstone courses, and restricted to student-course of multiple AP course enrollees linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passed exams below the indicated threshold in each panel. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

## Table B.9—heterogeneity by teachers' baseline pass rate and student load

| | Outcome: AP Exam Pass= $\mathbb{I}$(Exam Score≥3) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Post×PassRate$_{j,pre}$ | -0.169*** | | -0.168*** |
| | (0.013) | | (0.013) |
| Post×(NStudents$_{j,pre}$/100) | | -0.018** | -0.010 |
| | | (0.009) | (0.009) |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 |

Notes.– This table reports regression estimates from the continuous difference-in-differences specification (Eq. 3), including teacher, course, year, and student fixed effects. The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. The sample includes all Spring-semester or year-long AP courses from AY 2012-13 through 2018-19, except AP Research and Capstone. The estimates are based on the 'Analysis Sample', which is restricted to students who have taken two or more AP courses with teachers who were observed teaching in both pre- and post-bonus periods. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.7—heterogeneity across AP subjects (Combined Graph)



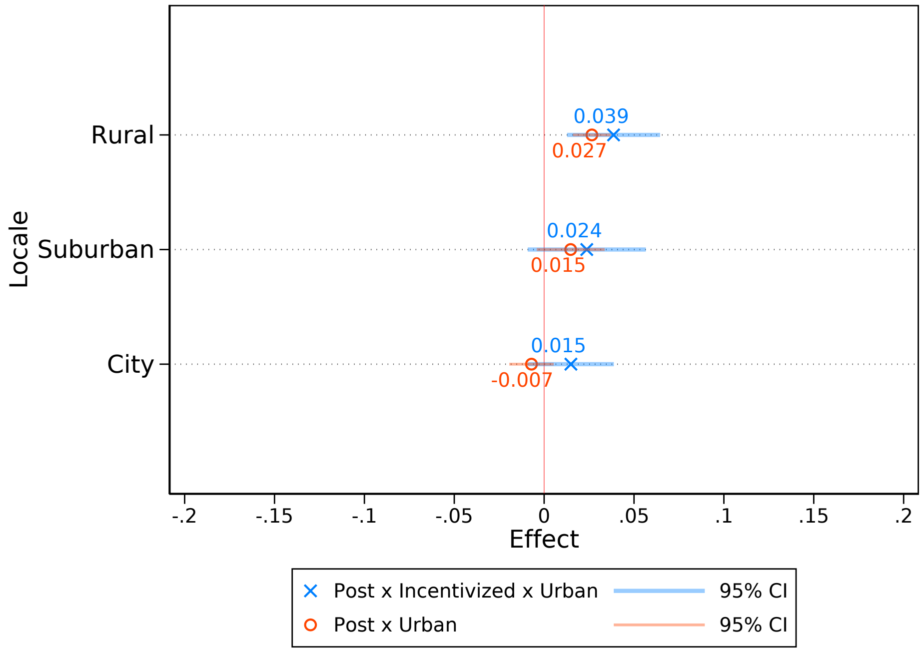

Notes.– See Figure 7b notes.

Table B.10—Estimates for Demographic Heterogeneity

| | (1) | (2) | (3) |
|---|---|---|---|
| **A. Difference-in-Differences** | | | |
| Post$\times$Incentivized$_j$ | 0.019 | 0.023* | 0.025** |
| | (0.012) | (0.012) | (0.012) |
| Post$\times$Incentivized$_j \times$Female$_i$ | 0.009 | | |
| | (0.006) | | |
| Post$\times$Incentivized$_j \times$Low-SES$_{it}$ | | 0.006 | |
| | | (0.007) | |
| Post$\times$Incentivized$_j \times$URM$_i$ | | | -0.004 |
| | | | (0.007) |
| Difference in Pre-bonus Mean | -0.05 | -0.23 | -0.23 |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| **B. Within-Teacher Model** | | | |
| Post | 0.010* | 0.025*** | 0.026*** |
| | (0.005) | (0.004) | (0.004) |
| Post$\times$Female$_i$ | 0.028*** | | |
| | (0.006) | | |
| Post$\times$Low-SES$_{it}$ | | 0.008 | |
| | | (0.007) | |
| Post$\times$URM$_i$ | | | 0.002 |
| | | | (0.005) |
| Difference in Pre-bonus Mean | -0.05 | -0.24 | -0.23 |
| adj. $R^2$ | 0.52 | 0.52 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 |
| Number of Unique teachers | 2,712 | 2,712 | 2,712 |

Notes.– This table reports regression estimates from the heterogeneous difference-in-differences model (Eq. 5). Observations are at the student-course level. The sample includes all Spring-semester or year-long AP courses between AY 2012-13 and 2018-19, except AP Research and Capstone courses, and restricted to student-course of multiple AP course enrollees linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passed exams below 70. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.8—heterogeneity by Locale (Combined Graph)



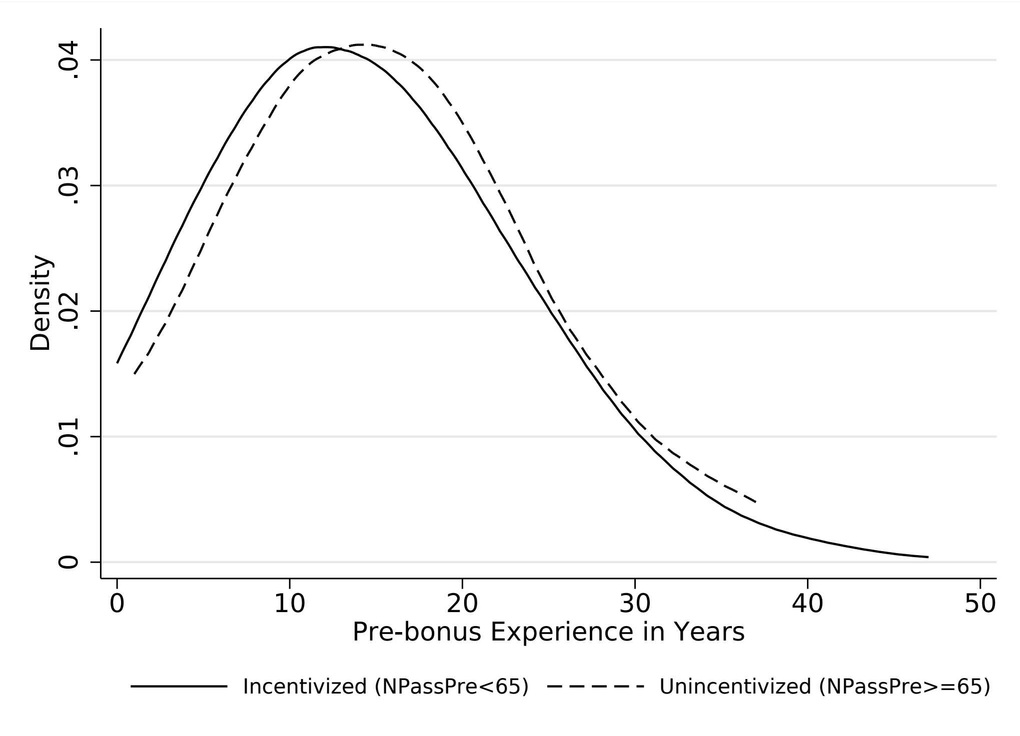

Notes.- The estimated effects are from heterogeneous within-teacher and difference-in-differences models, with full set of fixed effects. For further details see the text and Table 3 notes.

Table B.11—Heterogeneity by Teacher Experience

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Post×Incentivized$_j$ | 0.022** | 0.027** | 0.029** | 0.022* |
| | (0.011) | (0.011) | (0.011) | (0.013) |
| Post×Incentivized$_j$×Early-Career$_j$ | -0.003 | -0.005 | -0.006 | -0.000 |
| | (0.008) | (0.008) | (0.008) | (0.006) |
| Post×Incentivized$_j$×Novice$_j$ | 0.017 | 0.017 | 0.018* | 0.014* |
| | (0.011) | (0.011) | (0.011) | (0.007) |
| Pre-bonus Mean | 0.52 | 0.52 | 0.52 | 0.52 |
| adj. $R^2$ | 0.20 | 0.21 | 0.21 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Teacher FE | ✓ | ✓ | ✓ | ✓ |
| Subject FE | | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Student FE | | | | ✓ |

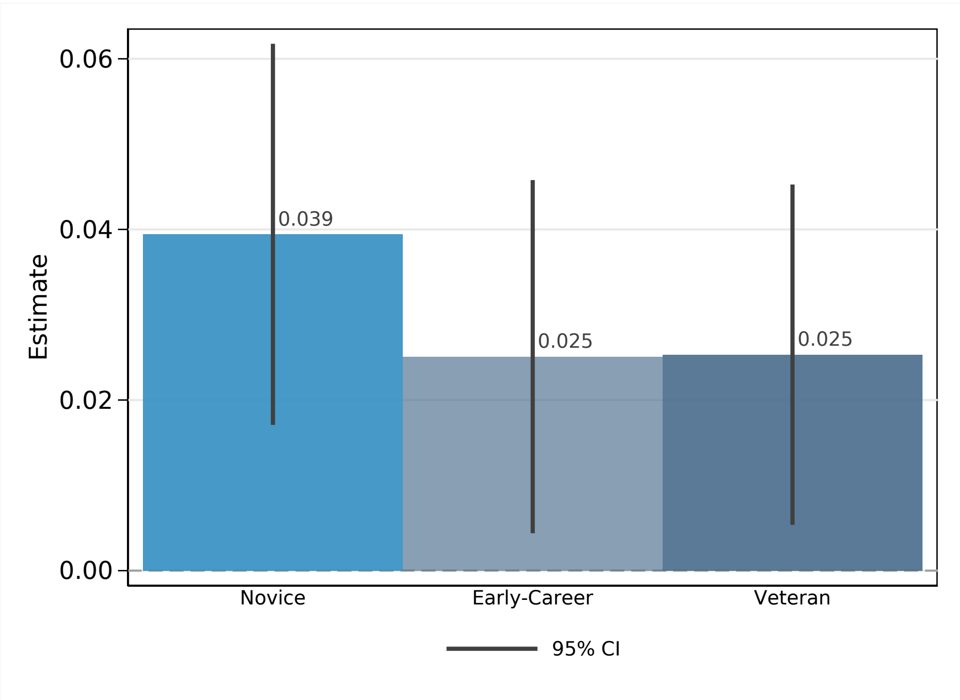
Notes.- This table reports regression estimates from the heterogeneous difference-in-differences model (Eq. 5). Observations are at the student-course level. The sample includes all Spring-semester or year-long AP courses between AY 2013-14 and 2018-19, except AP Research and Capstone courses, and restricted to student-course of multiple AP course enrollees linked to 2,712 teachers – who are observed in both pre- and post-bonus periods. The 'incentivized' teachers are those with pre-bonus average number of passed exams below 70. The coefficients are estimated using linear probability model; the standard errors are clustered at the teacher level. Based on their initial pre-bonus experience, teachers are assigned to Novice (less than 5 years), Early-Career (5 to 9 years), or Veteran (10 or more years) categories. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure B.9—Distribution of teacher experience
(incentivized versus unincentivized teachers, defined by alternative threshold)




Notes.– Kernel density estimate of pre-bonus teacher experience (in years), separated by incentivized and unincentivized status. The 'incentivized' teachers are those with pre-bonus average number of passed exams below 65. The sample includes 2,712 continuing AP teachers observed in both pre- and post-bonus periods. A bandwidth of 5 years was used for the estimation.

Figure B.10—heterogeneity by Teacher Experience (Incentivized group defining threshold: 65 average pre-bonus passers)




Notes.– This figure plots the regression estimates from the heterogeneous difference-in-differences model (Eq. 5), with teacher, grade, course, year, and student fixed effects. For further details see the text and Table 3 notes.

Table B.12—regression estimates

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **A. Full Sample** | | | | | | |
| Post | 0.039*** | 0.032*** | 0.027*** | - | - | - |
| | (0.006) | (0.006) | (0.004) | - | - | - |
| Pre-bonus Mean | 0.46 | 0.46 | 0.46 | - | - | - |
| adj. $R^2$ | 0.01 | 0.03 | 0.24 | - | - | - |
| Observations | 774,616 | 774,616 | 774,616 | - | - | - |
| **B. Matched Sample** | | | | | | |
| Post | 0.036*** | 0.029*** | 0.024*** | 0.022*** | 0.021*** | - |
| | (0.006) | (0.006) | (0.005) | (0.004) | (0.004) | - |
| Pre-bonus Mean | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | - |
| adj. $R^2$ | 0.00 | 0.02 | 0.24 | 0.23 | 0.35 | - |
| Observations | 693,634 | 693,634 | 693,634 | 693,634 | 693,634 | - |
| **C. Continuing Teacher Sample** | | | | | | |
| Post | 0.030*** | 0.024*** | 0.017*** | 0.022*** | 0.020*** | - |
| | (0.007) | (0.006) | (0.005) | (0.004) | (0.004) | - |
| Pre-bonus Mean | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | - |
| adj. $R^2$ | 0.00 | 0.03 | 0.24 | 0.21 | 0.34 | - |
| Observations | 484,611 | 484,611 | 484,611 | 484,611 | 484,611 | - |
| **D. Analysis Sample** | | | | | | |
| Post | 0.020*** | 0.020*** | 0.012** | 0.019*** | 0.017*** | 0.026*** |
| | (0.007) | (0.007) | (0.006) | (0.004) | (0.004) | (0.005) |
| Pre-bonus Mean | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 |
| adj. $R^2$ | 0.00 | 0.03 | 0.22 | 0.21 | 0.33 | 0.52 |
| Observations | 383,560 | 383,560 | 383,560 | 383,560 | 383,560 | 383,560 |
| Subject FE | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student Covariates | | | ✓ | | ✓ | |
| Teacher FE | | | | ✓ | ✓ | ✓ |
| Student FE | | | | | | ✓ |

Notes.- This table reports regression estimates from various specifications of within-teacher model (Eq. 1). The dependent variable is a binary indicator for achieving a (passing) score of 3 or higher on the AP exam for a given student-course. Observations are at the student-course level. Each panel reports estimates from a different sample of AP courses described in Section 4A. The coefficients are estimated using linear probability model, and for consistency, the standard errors are clustered at the school level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$