# Improving College Instruction through Incentives

Andy Brownback

*University of Arkansas*

Sally Sadoff

*University of California, San Diego*

In a field experiment, we examine the impact of performance-based incentives for community college instructors. Instructor incentives improve student exam scores, course grades, and credit accumulation while reducing course dropout. Effects are largest among part-time adjunct instructors. During the program, instructor incentives have large positive spillovers, increasing completion rates and grades in students' courses outside our study. One year after the program, instructor incentives increase transfer rates to 4-year colleges with no impact on 2-year college degrees. We find no evidence of complementarities between instructor incentives and student incentives. Finally, while instructors initially prefer gain-framed contracts over our loss-framed ones, preferences for loss-framed contracts significantly increase after experience with them.

We gratefully acknowledge Ivy Tech Community College and our invaluable partners there, Ronald Sloan and Neil Anthony. Eurika Bennett, Jaclyn Fisher, Ryan Johnson, Jacklyn Redwine, Tammy Jo Robbins, and Alisa Wells provided truly outstanding assistance with data collection and experimental execution. For thoughtful feedback on the paper, we thank Scott Carrell, Uri Gneezy, Alex Imas, Michael Kuhn, Steven Levitt, John List, and Jonah Rockoff. This project was made possible by the generous financial support of the Laura and John Arnold Foundation and the Yankelovich Center for Social Science Research. Our research was conducted with approval from the University of Arkansas and UCSD (University of California, San Diego) institutional review boards. The American Economic Association randomized controlled trials registry number is AEARCTR-0001411. Data are provided as supplementary material online.

Electronically published July 7, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 8]

## I. Introduction

Over the past several decades, returns to higher education have increased, as have college enrollment rates. However, much of the college premium eludes the many students who fail to adequately progress through their studies (Oreopoulos and Petronijevic 2013). Attainment is particularly low at 2-year community colleges, which serve about 40% of all undergraduates (Shapiro et al. 2017). These schools provide students with a low-cost entry point to accumulate college credits toward both 2-year and 4-year degrees. But they struggle with poor student performance, which hinders credit accumulation, degree completion, and transfers to 4-year schools (Snyder, De Brey, and Dillow 2018). In response, a growing literature examines interventions aimed at improving postsecondary performance. While wide-ranging, these policies share a common feature: they have generally targeted students, largely ignoring the role of college instructors.[1]

The lack of policy focus on college instructors is particularly surprising, given the rich literature demonstrating their importance. An increase of 1 standard deviation (SD) in college instructor quality improves student performance by an estimated 0.05–0.30 SD, with effects generally smaller at selective universities and larger at nonselective institutions similar to community colleges (Hoffmann and Oreopoulos 2009; Carrell and West 2010; Bettinger et al. 2014; Braga, Paccagnella, and Pellizzari 2016; Brodaty and Gurgand 2016; De Vlieger, Jacob, and Stange 2017). Recent work examines the extent to which postsecondary institutions adjust personnel policies—such as teaching allocations and salaries—in response to instructor productivity (Courant and Turner 2017; De Vlieger, Jacob, and Stange 2017). But we know little about whether college instructor quality can be improved.

To our knowledge, no prior study has explored whether an intervention targeting instructors can improve postsecondary student performance.[2] The dearth of interventions targeting college instructors stands in sharp contrast to the large body of literature on improving teacher effectiveness at the elementary and high school levels. This is a critical gap because the production function for postsecondary instruction differs in important ways from primary and secondary school teaching. Students

[1] These policies include lowering the costs of college attendance, providing students with information and support services, and offering students performance-based incentives (for reviews, see Deming and Dynarski 2009; Lavecchia, Liu, and Oreopoulos 2014; and Evans et al. 2017).

[2] Related work using event studies argues that an increased emphasis on teaching improved student course evaluations at a US business school (Brickley and Zimmerman 2001) and that an increased emphasis on faculty research decreased student grades at an Italian university (De Philippis 2015).

are generally independent adults, spend less time with their instructors, and can voluntarily withdraw from an individual course or the institution entirely. Thus, many of the mechanisms by which teachers can respond to incentives at the K–12 level are unavailable to college instructors. At the same time, there may be greater scope for innovative personnel policies in higher education, where teaching assignments and employment contracts are generally more flexible than in most K–12 settings. This flexibility has increased with the sharp rise in part-time adjunct instructors who work under short-term contracts and teach courses at lower cost than full-time faculty (Ehrenberg 2012; McFarland et al. 2017).[3]

We fill the gap in the literature by experimentally testing the impact of performance-based incentives for community college instructors. First, as our primary question, we investigate whether incentives for instructors can improve postsecondary student performance. Second, we test whether instructor incentives can be more effective in combination with incentives for students. Finally, in order to explore their feasibility as a personnel policy, we examine instructor preferences for the incentive contracts we offer.

We conducted our field experiment at several campuses of a statewide community college. Two-year community colleges are gaining increased attention from policy makers as a low-cost pathway to 4-year schools, particularly for underrepresented and nontraditional students (Bailey, Jaggars, and Jenkins 2015). They also offer potentially high returns, increasing earnings by an estimated 5%–11% for each year of completed coursework (Kane and Rouse 1995; Grubb 2002; Marcotte et al. 2005). Yet, for the majority of students, the promise of community college remains unfulfilled. While more than 80% enter with the intention of transferring to a 4-year school, only about a quarter succeed in doing so (Jenkins and Fink 2016).[4] Fewer than 40% of community college students earn a college degree within 6 years (Shapiro et al. 2017), and a large share drop out with debt.[5]

The poor outcomes of community college students stem from poor course performance—in particular, high rates of course dropout and low course grades (Ran and Xu 2017). Improving performance is critical for students to accumulate credits and establish the grade point average (GPA) needed to earn a degree or transfer to a 4-year school. Accordingly,

[3] There is ongoing debate about the impact of this shift on student achievement (Ehrenberg and Zhang 2005; Bettinger and Long 2006, 2010; Hoffmann and Oreopoulos 2009; Figlio, Schapiro, and Soter 2015; Rogers 2015; Ran and Xu 2017).

[4] Recent work examines whether expanding access to community colleges diverts students from earning 4-year degrees (Zimmerman 2014; Denning 2017; Goodman, Hurwitz, and Smith 2017; Mountjoy 2018).

[5] The five public postsecondary institutions producing the highest ratio of dropouts with debt to graduates include four community colleges (including our partner institution). See Barshay (2017).

we examine the impact of our intervention both on performance in targeted courses and on students' broader educational outcomes.[6]

Our study included 16 different departments with more than 6,000 student-course observations in the fall and spring semesters of the 2016–17 school year. In the fall semester, we randomly assigned instructors to one of two treatment groups: instructor incentives or control. In the instructor incentives group, instructors received performance bonuses of $50 per student who received 70% or higher on an objective, externally designed course exam ("passed," hereafter). We framed the incentives as losses, that is, bonuses that instructors would lose if they did not meet performance targets. To implement the loss framing, we gave instructors up-front bonuses at the beginning of the semester equivalent to the amount they would receive if half of their students passed the exam. At the end of the semester, if fewer than half of an instructor's students passed the exam, the instructor returned the difference between their final reward and the up-front bonus. If more than half of the students passed the exam, the instructor received additional rewards.

In the spring semester, we introduced combined incentives, which offered incentives to students in conjunction with incentives to instructors. Incentivized students received free tuition for one summer course (worth approximately $400) if they passed the exam. We assigned student incentives at the section level, cross-randomizing them with the existing assignment of instructor incentives. This yields four treatment groups: control, instructor incentives only, student incentives only, and combined incentives (incentives for both instructor and students). In order to explore potential complementarities between instructor and student incentives, we examine whether combined incentives are more effective than instructor incentives alone.

Finally, we used incentive-compatible mechanisms to elicit instructors' contract preferences, both at baseline when they enrolled in the study and at the end of the fall semester, after incentivized instructors had experienced the contracts. We compare the loss-framed contract to a more standard gain-framed contract, in which rewards are distributed at the end of the semester ("loss" and "gain" contracts, respectively). Our elicitation captures the difference in the per-student incentive amount that would make an instructor indifferent between working under the loss versus the gain contract.

As we discuss in more detail below, we find that instructor incentives have large impacts on student performance in the targeted course,

[6] Improved performance also helps students avoid debt. Course withdrawals provide neither credit nor refund. And federal financial aid generally requires students to maintain at least a 2.0 GPA and a 67% completion rate (Scott-Clayton and Schudde [2016] examine the impact of these requirements).

significantly improving course completion, exam performance, and course grades. Importantly, incentives for instructors also improve students' broader educational outcomes, including course completion and grades in unincentivized courses, overall credit accumulation, and transfer rates to 4-year schools.

In targeted courses, instructor incentives increase exam performance by an estimated 0.16–0.20 SD ($p < .01$) and increase exam pass rates by 19% ($p < .01$). The impact carries over to course performance, where grades improve by an estimated 0.10–0.11 SD ($p = .02$). The effects of incentives operate at both the extensive and intensive margins. Instructor incentives reduce course dropout rates by 17% ($p = .03$) and increase exam scores conditional on completion by 0.083 SD ($p = .04$). We find no evidence that instructor incentives are more effective in conjunction with student incentives.[7]

At the instructor level, the effects of incentives are largest among adjunct instructors (0.26 SD on exam scores, $p < .01$), with smaller effects among full-time faculty (0.13 SD, $p = .12$). We find no evidence that the effects are driven by a narrow set of departments or that they are dependent on novelty. On the contrary, the effects of incentives are consistently positive across a wide range of courses, and the effects are sustained across multiple semesters, with larger point estimates in the second semester that incentives are offered.

We next examine the impact of our intervention on students' broader educational outcomes. During the program, instructor incentives have large positive spillovers, significantly improving course completion and grades in students' unincentivized courses outside our study. Remarkably, the impact is similar in magnitude to the estimated effect in targeted courses. This suggests that the effects on incentivized courses are not due to gaming or to substitution of effort away from unincentivized courses. Instead, our findings suggest that instructor incentives foster general improvements in enrollment and academic focus during the treatment semester.

At the end of the program, we estimate that incentivized courses increase students' total credit accumulation by 18% ($p = .01$), with positive but not significant impacts on GPA. As noted above, credit accumulation and GPA are critical for students to qualify for transfer to 4-year schools. And indeed we find that 1 year after the program ends, instructor incentives increase transfer rates to 4-year colleges by 22%–28% ($p < .01$), with no effect on 2-year college degrees. Our findings demonstrate that instructor incentives have persistent and meaningful effects on students'

[7] We also find little evidence that student incentives have meaningful effects when offered alone. We note that we did not power the experiment to separately estimate the impact of student incentives.

educational outcomes. More broadly, the results suggest that instructor incentives can help community colleges better fulfill their mission as a low-cost pathway for students to pursue bachelor's degrees.

Finally, turning to instructors' contract preferences, we find two striking results. First, at baseline, instructors significantly prefer gain contracts to loss contracts. On average, they are willing to give up about 9% of the $50 per-student incentive payment in order to work under a gain contract rather than a loss contract. Second, after one semester of working under loss contracts, incentivized instructors significantly increase their preferences for them. The effects are large enough that instructors become (close to) indifferent between loss and gain contracts. This novel finding suggests that providing instructors experience with loss contracts could make them more attractive as a personnel policy.

To our knowledge, this is the first study to demonstrate that an intervention can improve instructor effectiveness at the postsecondary level. We show that instructor incentives substantially improve college instruction at relatively low cost. The effects of our incentives are similar in size to improving instructor quality by 1 SD and have an expected cost of about $25 per student-course. Based on the impact on credits alone, the intervention easily passes a cost-benefit analysis, with an estimated return of more than $40 per student per year or a 10-year net present value of more than $250.[8]

Our community college context may make incentives especially powerful. Community college instructors focus primarily on teaching (rather than research), most work under flexible contracts, and low-cost rewards can provide a substantial bonus relative to baseline pay. This is particularly true for adjunct instructors, for whom the expected incentive in our study was equivalent to approximately 20% of their salary. The dramatic impact of our incentives on adjunct instructors suggests that there could be substantial gains from reconsidering the contracts offered to part-time instructors. These changes not only could significantly improve student outcomes but are also feasible from a policy perspective, given the preferences of instructors, the low cost of the incentives, and the short-term contracting used to hire adjunct faculty.

In the remainder of the paper, section II discusses the related literature, section III describes the experimental design, section IV presents

[8] We calculate a yearly return for men (women) of $41.47 ($42.92) using our estimated treatment effects of 0.52 (0.39) credits divided by 30 credit hours (i.e., 1 year of credit hours), multiplied by the 6% (10.5%) increase in annual earnings from an additional year of completed community college coursework estimated by Marcotte et al. (2005) and the median yearly earnings of $39,950 ($31,400) for men (women) without any college—a conservative subgroup—estimated by the Bureau of Labor Statistics (2017). We use a conservative discount rate of 10% to calculate a net present value over 10 years of $263.72 ($254.82) for men (women).

the results for targeted courses, section V presents the effects on broader educational outcomes, section VI examines instructors' contract preferences, and section VII concludes.

## II. Literature

Our first contribution is to a growing body of literature examining teacher incentives, which has until now been limited to the elementary and high school levels. This prior work has found mixed results. While nonexperimental studies in the United States and experimental studies in developing countries have found that teacher incentives can improve performance, experimental studies in the United States have largely failed to demonstrate effectiveness.[9] We based the design of the incentives in our study on Fryer et al. (2012, 2018), which is the only prior experimental study in the United States to find a positive impact of teacher incentives. The authors test up-front, loss-framed incentives among elementary and middle school teachers and estimate effects of 0.12 SD on math test scores, pooling across 2 years of the experiment. Our finding that similarly structured incentives are effective among college instructors suggests that the impact of loss-framed incentives on teacher performance may replicate across contexts.[10]

We also add to a small set of existing studies that have found conflicting results when comparing incentives offered both alone and in combination.[11] In line with our results against complementarities between instructor and student incentives, List, Livingston, and Neckermann (2012) find little evidence of complementarities between incentives for students, parents, and tutors in an experiment in US elementary schools. In contrast, Behrman et al. (2015) find that incentives for teachers and students in

[9] Neal (2011) and Fryer (2017) provide reviews. For experimental studies in developing countries, see Glewwe, Ilias, and Kremer (2010), Muralidharan and Sundararaman (2011), Duflo et al. (2012), Loyalka et al. (2016), and Barrera-Osorio and Raju (2017). For experimental studies in the United States, see Glazerman, McKie, and Carey (2009), Springer et al. (2011, 2012), and Fryer (2013).

[10] Unlike Fryer et al. (2012, 2018), we do not attempt to compare loss- and gain-framed contracts. They find that gain-framed contracts have an estimated impact of 0.05 SD, pooling across 2 years of the experiment. Our incentives also differ. As discussed in sec. III.B, we base rewards on threshold achievement levels, while Fryer et al. (2012, 2018) used the pay-for-percentile structure developed by Barlevy and Neal (2012).

[11] A large body of literature examines student incentives alone and generally finds small effects (see reviews by Sadoff 2014 and Fryer 2017). In a community college context, Patel et al. (2013) and Barrow et al. (2014) find that performance-based scholarships for students modestly improve GPA, credit accumulation, and degree attainment. In contrast to our low-cost, short-term incentives, these scholarships were expensive ($1,000–$4,000) and long term. Other prior studies that find small overall effects of incentives for college students include Angrist, Lang, and Oreopoulos (2009), Leuven, Oosterbeek, and Van der Klaauw (2010), De Paola, Scoppa, and Nisticò (2012), and Angrist, Oreopoulos, and Williams (2014).

Mexican high schools were more effective when offered in combination than when offered separately.[12] The differing results across studies could be driven by differences in complementarities between instructor and student effort in the production function or could also be due to differences in the strategic response of instructors and students to each others' effort choices (Todd and Wolpin 2003; De Fraja, Oliveira, and Zanchi 2010).

Finally, our study contributes to personnel economics by examining employee preferences for loss contracts. The motivational power of loss contracts is consistent with a large literature in behavioral economics demonstrating loss aversion, under which having to pay back (or "lose") part of the bonus is more painful than failing to earn (or "gain") the equivalent amount at the end of the semester (Kahneman and Tversky 1979). A growing body of laboratory and field studies demonstrates that framing incentives as losses can increase worker effort compared to more traditional gain-framed incentives.[13] Despite their potential impact on productivity, however, explicit loss-framed contracts are not widely prevalent, raising questions about their effectiveness as a personnel policy.

One concern with loss contracts is that their motivational power may diminish after instructors experience these novel bonuses, making them decreasingly effective over time. The limited work on this question is mixed. Fryer et al. (2018) find that while up-front bonuses for teachers have large impacts in the first year they are offered, the incentives are not effective in the second year of their experiment.[14] Similarly, List (2003, 2004, 2011) finds that experience limits the impact of loss framing in trading markets. In contrast, Hossain and List (2012) conduct an experiment offering incentives to Chinese factory workers and find that the effects of loss framing are sustained over time. Our finding that the effects of incentives are as large, if not larger, the second time they are offered suggests that loss-framed incentives can have a sustainable impact.

A second concern is that workers may find the loss contracts aversive and prefer to work under gain contracts. If this is the case, employers may need to increase employee compensation in order to retain employees

[12] We note that the combined intervention in Behrman et al. (2015) had programmatic elements that were not included in the individual interventions. Using observational data, Geng (2017) finds evidence of complementarities between a grade-retention policy incentivizing students and an accountability scheme incentivizing teachers and schools.

[13] See Brooks, Stremitzer, and Tontrup (2012), Fryer et al. (2012, 2018), Hossain and List (2012), Armantier and Boly (2015), Hong, Hossain, and List (2015), and Imas, Sadoff, and Samek (2016). In online studies, DellaVigna and Pope (2018) and De Quidt et al. (2017) do not find significant differences between loss- and gain-framed incentives. Studies comparing loss- and gain-framed incentives outside of work settings find mixed results (e.g., List and Samek 2015; Levitt et al. 2016; Englmaier et al. 2018).

[14] Fryer et al. (2018) rerandomize teachers in the second year of their experiment so that instructors receive different treatments across years, and there is no group of never-incentivized teachers across both years.

who work under loss contracts, offsetting the improved productivity. While standard behavioral models predict that workers will prefer gain contracts, the limited empirical evidence from laboratory and online studies finds a preference for loss contracts (Imas, Sadoff, and Samek 2016; De Quidt 2018; Jie 2018).[15] Our study is the first to examine preferences for loss contracts in a high-stakes, natural environment using employee salaries, as well as the first to examine preferences both before and after working under loss-framed incentives. Interestingly for both theory and policy, our results suggest that people who experience loss contracts do not judge those experiences as negatively ex post as they did ex ante.[16]

## III. Experimental Design

### A. *Setting and Recruitment*

We conducted the experiment in the 2016–17 school year at Ivy Tech Community College of Indiana. Ivy Tech is Indiana's largest public postsecondary institution and the nation's largest singly accredited, statewide community college system, serving nearly 170,000 students annually. Our sample includes courses from several campuses in the East Central and Richmond regions: Anderson, Connorsville, Marion, Muncie, New Castle, and Richmond.

At the time of our experiment, the East Central and Richmond regions served communities in the 4th and 8th percentiles, respectively, of national median income. More than 60% of their student body was eligible for Pell Grants, placing them in the 90th percentile for community colleges. Their fall-to-fall retention rates of full-time students hovered around 40%, just above the bottom 10% of community colleges. Overall, only 24% of their full-time, first-time students would graduate or transfer to a 4-year institution within 3 years, also just above the bottom 10% of community colleges (NCCBP 2014).

Our study includes a broad range of departments: accounting, anatomy and physiology, art history, biology, business, business operations applications and technology, communications, criminology, English, health

[15] Models using the status quo as the reference point (e.g., Tversky and Kahneman 1991) predict that individuals will work harder under loss contracts conditional on the endowment (i.e., the up-front bonus) being incorporated as the status quo. If the distribution of possible outcomes (i.e., final rewards) is taken as the reference point (e.g., Kőszegi and Rabin 2006), then the contract framing should be irrelevant for both effort and preferences. Thus, our standard behavioral model refers to one assuming the status quo as the reference point. See Imas, Sadoff, and Samek (2016) for discussion of the theory.

[16] These findings are in line with Kermer et al. (2006), who argue that the affective experience of losses is less painful than people expect it to be. In contrast, Czibor et al. (2019) find evidence that workers in a laboratory experiment are less likely to want to participate in future studies after working under loss-framed incentives compared with gain-framed incentives.

sciences, math, nursing, physical science, psychology, sociology, and software development. We determined course eligibility based on whether the course included an objective course-wide exam (or objective portion of a larger exam). The exams were developed at the departmental or statewide level, tested key learning outcomes of the course, and could be graded objectively using Scantron answer sheets, computer-based testing, or a course-level answer key. To ensure instructors were not able to "game" our incentives, department heads agreed to maintain the confidentiality of the exam prior to its administration. Any instructor who taught at least one section of an eligible course was invited to participate.

Prior to and during our study, Ivy Tech offered no other performance pay and had no formal policies in place to determine hiring and retention of instructors based on student performance. Both full-time faculty and adjunct instructors work under nonpermanent contracts. The primary difference is that full-time faculty teach higher course loads, earn a salary (with benefits) rather than being paid on a per-course basis, and are assigned administrative tasks such as advising students. In addition, full-time faculty are hired on a yearly basis, whereas adjuncts are hired on a semester basis. In general, instructors are retained for the following terms if there are teaching needs, unless the administration receives reports of problematic behavior or student complaints. Neither full-time nor adjunct faculty were unionized at the time of our study.

In the fall 2016 semester, Ivy Tech identified approximately 150 eligible instructors. Ivy Tech administrators recruited these instructors by email and in person. We then enrolled interested instructors in the study through an online survey. The enrollment period began August 15, 2016, and ended September 6, 2016, with a final total of 108 enrolled instructors, 90% of our recruitment goal of 120 and 72% of all eligible instructors. The randomization (detailed in sec. III.D) was based on the students enrolled in a given course as of the Ivy Tech census date, September 2, 2016. The census date is at the end of the second week of courses and is the final date students can receive a refund for a course. By delaying the randomization, we can control for selective take-up or attrition resulting from treatment assignment. Additionally, we can ensure that our estimates of withdrawals are not influenced by the natural add and drop cycles during the first 2 weeks of class.[17]

Fall instructors teaching eligible courses in the spring 2017 semester were automatically reenrolled. Of the 108 participating instructors in the fall, 74 were eligible in the spring, and all but one elected to continue participation (as discussed in sec. III.D, there were no differences in eligibility by treatment group). We also recruited new instructors. The

[17] Adding a course after the census date is rare and requires special permission from the instructor of record and the regional academic officer.

recruitment followed the same procedure as in the fall 2016 semester, with Ivy Tech administrators emailing 74 eligible instructors who either had chosen not to participate in the fall semester or were newly eligible. The enrollment period began January 20, 2017, and ended February 3, 2017. An additional 26 instructors signed up, bringing the spring semester total to 99 participating instructors. Including continuing instructors from the fall, 66% of eligible instructors participated in the spring. As in the fall, the spring randomization was based on enrollment as of the spring semester census date, January 30, 2017. Over the two semesters, 134 instructors participated in the study, 93% of our recruitment goal of 144.

### *B. Treatments*

We test two cross-cutting incentive schemes, incentives for instructors and incentives for students, which yields four treatment groups: instructor incentives only, student incentives only, combined incentives, and control.

In the instructor incentives and combined incentives treatments, instructors received $50 per student who passed the objective course exam.[18] We chose our incentive structure for several reasons. First, it was simple for instructors to understand, which our pilot testing in spring 2016 suggested was critical for effectiveness. Second, our partners at Ivy Tech identified passing the exam as a critical measure of student learning and course success. Third, it was feasible to implement in the college context. We were not able to base incentives on an instructor's rank in terms of value added—as suggested by Barlevy and Neal (2012)—because we lacked sufficient baseline information about students, particularly at the subject-specific level. This uncertainty about baseline student information also makes it potentially more difficult for instructors to game incentives because, unlike in K–12 contexts, they have little information about which students are expected to be marginal. Our incentive scheme does share an important feature with Barlevy and Neal (2012), that there are continuous rewards for improvement regardless of the instructor's performance. This stands in contrast to incentive structures that reward, for example, only the instructors at the top end of the performance distribution (for discussion, see Neal 2011).

Instructors received incentives for all students in all of their eligible sections. At the beginning of the semester, the University of Arkansas distributed checks for up-front bonuses equivalent to the amount instructors would earn if 50% of their students passed the exam. For example,

[18] We defined passing as achieving a score of 70% or more, except in the health sciences and nursing courses, which had thresholds for passing that exceeded 70%. In these cases, we considered passing to be the preexisting requirement of 75%.

an instructor who taught one section with 20 students would receive an up-front check for $500. At the end of the semester, if fewer than 50% of the students passed the exam, the instructor was responsible for returning the difference between the final bonus amount and the up-front bonus. If more than 50% of the students passed, the instructor received an additional bonus check. Recent work demonstrates that under a prospect theory model with both loss aversion and diminishing sensitivity (i.e., utility is convex in losses and concave in gains), contracts like ours—offering both bonuses and penalties for performance above and below a threshold, respectively—can increase worker effort compared to pure bonus or pure penalty contracts (Armantier and Boly 2015).

At the beginning of the semester, we notified instructors of their treatment assignment. We emailed instructors assigned to instructor or combined incentives a list of their incentivized sections and an estimate of the up-front incentive payment they would receive. In order to clarify details and give instructors a chance to fill out the accounting forms in person, we held information sessions on each of the four primary campuses (Anderson, Marion, Muncie, and Richmond). One information session each semester was broadcast online for those who could not attend in person. Instructors who did not attend the session could sign the forms through our Ivy Tech partners or electronically. The up-front bonus payment was issued once the forms were signed.

The average up-front bonus was $726, and the average final bonus was $662. Fifty-five percent of instructors owed money back at the end of the semester, with an average repayment of $308. We had high rates of compliance for the up-front bonuses—98% of instructors in the fall and 94% of instructors in the spring complied with the up-front payments.[19] Compliance with repayments varied across the two semesters: in the fall, 93% of instructors who owed money complied with repayment (96% of money owed); and in the spring, 78% of instructors who owed money complied with repayment (83% of money owed). The lower repayment rate in the spring may have been due to instructors knowing that the study would not continue after the spring semester, a concern that would not be present if this were a system-wide policy. If instructors expected not to make repayments, this would likely lower the impact of incentives and thus our ability to detect treatment effects.

In the student incentives and combined incentives treatments, students received free tuition for one summer course if they passed their exam in the treated course. We designed the incentives in partnership with Ivy Tech to satisfy several administrative constraints. Offering cash incentives was not feasible, as cash rewards crowd out existing financial aid

[19] The remainder did not fill out the paperwork to receive payments (three instructors) or did not cash the up-front payment check (one instructor).

for certain students. Relatedly, because summer enrollment may not be covered by Pell Grants, summer scholarships can help lower a student's debt burden beyond what a fall or spring scholarship could do. The summer scholarship incentives were also attractive from a cost perspective. A summer scholarship had a face value of $400 but an expected marginal cost of only about $97, given realized pass rates of 44.7% and take-up rates of 54.4% in sections offering student incentives. Given summer enrollment rates of 26.8% among students in the control group, the ex ante expected value for students was about $107.

In spring 2017, we informed instructors which (if any) of their sections would receive incentives for students and outlined the basic design of the incentives. An Ivy Tech administrator described the incentives to students (in person for traditional classes and through a video for online classes). Participating students received a refrigerator magnet reminding them of the details (fig. A.1; figs. A.1–A.3 are available online). Students enrolled in the program by signing up in their class or through an online survey. Of the 1,035 students offered incentives, 772 (74.6%) actively consented to participate and 48 (4.6%) actively declined to participate. Our primary analysis is at the intent-to-treat level and does not depend on whether a student chose to participate in the program.[20]

While we randomized at the section level, we cannot fully rule out that there were interactions between students in treatment and control sections. However, our context may minimize concerns about spillovers: no campuses have on-campus housing in which students may develop relationships, and campuses are spread across several different cities. Finally, if students in incentivized sections did study with students in unincentivized sections, this would likely weaken our ability to detect treatment effects.

### *C. Survey*

All participating instructors filled out a short online survey in order to enroll in the experiment. We asked instructors participating in the fall semester to fill out a midyear survey at the end of the fall semester before they learned their final payment in December 2016 (instructors new to the program in the spring filled out the enrollment survey at this time). We asked instructors participating in the spring semester to fill out a year-end survey in May 2017. Response rates were 87% for the midyear survey (96% in instructor incentives and 77% in instructor control). Response rates were lower for the year-end survey at 67% (83% in instructor incentives and 49% in instructor control).

[20] Consent did not affect our access to anonymous student-level data but did affect whether we could distribute summer tuition vouchers to students.

In the enrollment and midyear surveys, we elicited instructors' preferences for loss-framed relative to gain-framed contracts (see app. B for preference elicitation questions; apps. A, B are available online). First, we asked instructors to choose which contract they would prefer to receive if both contracts paid \$50 per student. Then, we used a multiple price list in which instructors made a series of decisions between an advance bonus—a loss-framed contract that provided half of the total possible reward up-front—and an end bonus—a gain-framed contract that paid all rewards at the end of the semester. Our multiple price list elicited preferences between the loss-framed contract, with a bonus of \$50 per student, and 13 different gain-framed contracts, with a bonus of \$$X$ per student, with $X \in \{60, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 40\}$. In order to ensure that the surveys were incentive compatible, we randomly selected one choice from one respondent at random to determine that respondent's contract.[21]

Contract preferences may be confounded by time preferences because more impatient instructors may express a relatively stronger preference for loss framing due to the earlier arrival of the payments (and vice versa for more patient instructors and gain-framed contracts). In order to separately identify contract preferences from time preferences, we also elicited instructors' preferences over receiving unconditional cash awards at the beginning versus the end of the semester. Similar to the multiple price list for contracts, instructors made 13 decisions between a \$500 bonus at the beginning of the semester and a bonus of \$$B$ at the end of the semester, with $B \in \{600, 550, 540, 530, 520, 510, 500, 490, 480, 470, 460, 450, 400\}$. The large possible payments offered through this incentive-compatible elicitation also served as an incentive for all instructors to complete the survey.

In all surveys, we asked instructors about their subjective well-being and attitudes toward teaching. In the midyear and year-end surveys, we also asked about their time use and personal expenditures on instruction. For instructors in the treatment group, we additionally asked how they had used their up-front payments and their expectations about their final reward (e.g., whether they expected to receive additional rewards or owe money back).

### *D. Randomization*

We first describe the randomization of instructors to receive incentives in the fall and spring semesters (see fig. A.2 for a summary). We then

[21] In the randomly chosen decisions from both the fall and the spring, the instructor selected the loss contract and so received the same incentives as the other treatment instructors.

describe the randomization of individual sections to receive incentives for students in the spring semester. The randomization and analysis follow our preanalysis plan.[22]

In the fall 2016 semester, we assigned instructors to either instructor incentives or instructor control. We used a block randomized design, stratifying our instructors by department and instructor type (adjunct or full-time faculty).[23] We intended to stratify at a finer level, but that would have resulted in many strata having only one course, which precludes randomization within blocks. To ensure balance between treatment and control, we tested for significant differences in course-level characteristics: courses per instructor, students enrolled per instructor, and the percentage of courses with a corresponding remedial section (corequisite), as well as instructors' time preferences and instructors' contract preferences elicited in the enrollment survey. We also tested for significant differences in student-level characteristics: gender, age, race, accumulated credit hours, and GPA. For each characteristic, we specified that we would rerandomize in the event that differences were significant with $p < .15$.[24]

In spring 2017, we conducted the randomization in two stages. First, we determined whether an instructor would receive incentives. Next, we assigned which sections would receive student incentives. For the instructor incentive stage of the randomization, we independently assigned continuing instructors who participated in the fall and instructors who were new to the program. Of the 55 instructors assigned to instructor incentives in the fall, 37 taught eligible courses in the spring. Of the 53 instructors assigned to instructor control in the fall, 37 taught eligible courses in the spring. Continuing instructors were assigned to the same treatment they received in the fall. The exceptions to this are (1) one eligible instructor assigned to instructor incentives in the fall who opted out of the spring semester of the study and (2) two instructors assigned to instructor control in the fall who received instructor incentives in the spring. In order to encourage survey completion and continued participation among control instructors, we told them they would have a chance to receive incentives

[22] We preregistered our analysis plan; see https://osf.io/fbxpw/. We later note deviations from the preanalysis plan due to data or experimental constraints.

[23] For some departments, it was impossible to stratify on both instructor type and department. In these cases, we pooled courses across departments and stratified on instructor type.

[24] We used a probit regression and regressed the treatment assignment on the characteristics. We rerandomized if any coefficients were significant with $p < .15$. Bruhn and McKenzie (2009) examine various methods of randomization including stratification and rerandomization, both of which we employ. Following their recommendation, our preferred specification estimates treatment effects including both those covariates we stratified on and the individual student and teacher characteristics used to check balance. As shown in sec. IV, including only the sparse set of covariates we stratified on does not affect the results.

in the spring. Accordingly, we randomly reassigned two instructors. Therefore, we have 38 continuing instructors assigned to spring instructor incentives and 35 continuing instructors assigned to spring instructor control. Continuing fall instructors were well balanced on baseline characteristics. The only significant difference is that instructors continuing in instructor incentives taught 0.53 more sections, on average, than those continuing in instructor control, significant at the $p < .05$ level.

New instructors were assigned to instructor incentives or instructor control following the same procedure as in the fall randomization. While we checked the balance of these characteristics among the full sample of instructors, we ran the randomization for new instructors independently to ensure that new spring instructors underwent the exact same assignment process as new fall instructors.

For the student incentive stage of the randomization, we assigned sections to receive student incentives within each of the instructor incentive assignments. For instructors assigned to receive incentives, we selected half of all their sections to receive student incentives (making them combined incentives sections), while the other half remained instructor incentives only. In order to maximize within-instructor variation, any instructor with multiple sections had half of their sections assigned to receive combined incentives. Instructors with an odd number of sections were randomly rounded up or down. For instructors who taught one section, half were assigned to receive combined incentives, and half received instructor incentives only.

For instructors assigned to instructor control, we first randomized half the instructors to a pure control group (no instructor or student incentives). Among the other half of instructor control instructors, we selected half of their sections to receive student incentives only and the other half of their sections to remain as pure control, following the same procedure described for the instructor incentives group. This asymmetrical method of assigning student incentives to instructors based on their instructor incentive assignment preserves a pure control (no student or instructor incentives) group. It also allows for a more powerful within-instructor test of complementarity between instructor incentives and student incentives. We balanced the student incentives randomization on all of the same characteristics as the instructor incentive assignment.

### *E. Analysis*

We test two hypotheses: first, that instructor incentives improve student outcomes and, second, that instructor incentives have larger effects in combination with student incentives than they do alone.

Our primary estimating equation uses a random effects linear regression, with standard errors clustered by instructor, which is our unit of randomization. The random effects estimator is the efficient (i.e., minimum variance) estimator in our environment. We are able to employ this estimator because the random assignment of treatment satisfies the stringent independence requirement that instructor and student characteristics be uncorrelated with the explanatory variable of interest. In related work, Carrell and West (2010) provide a detailed discussion of their use of a random effects model to estimate instructor quality when students are randomly assigned to professors. For binary outcomes, we use a random effects probit model with clustered standard errors. We elect to use a probit model instead of a linear probability model (LPM) because the LPM estimation generates predicted probabilities greater than one for 1%–5% of our observations. Horrace and Oaxaca (2006) show how this leads to bias and inconsistency. We report the marginal effects whenever we conduct this analysis. As discussed below, our estimation is robust to using ordinary least squares (OLS) estimation.[25]

We estimate the following equation using a random effects linear regression model, with standard errors clustered at the instructor level:

$$Y_{i,j,s} = \beta_0 + \beta_1 Z^1_{s,j} + \beta_2 Z^2_{s,j} + \beta_3 Z^3_{s,j} + \beta_4 X_i + \beta_5 X_s + \beta_6 X_j + U_j + \epsilon_{i,s},$$

where $Y_{i,s,j}$ is the outcome for student $i$ in section $s$ taught by instructor $j$; $Z^t_{s,j}$ is an indicator variable for whether section $s$ taught by instructor $j$ is assigned to treatment $t = \{1, 2, 3\}$, with 1 = instructor incentives, 2 = student incentives, and 3 = combined incentives; $X_I$ represents a vector of student covariates (age, race, gender, baseline credits); $X_s$ represents section-specific covariates (semester, academic department, and whether it is a corequisite course); $X_j$ represents instructor-specific covariates (full-time or adjunct, time preference, and contract preference); $U_j$ represents the instructor-specific random effect; and $\epsilon_{i,s}$ is the error term, which, due to the randomization, is mechanically uncorrelated to the $Z^t_{s,j}$ terms.[26]

Since we partition our sections into the three treatments and control, $\beta_1$, $\beta_2$, and $\beta_3$ measure the full effects (rather than marginal effects) of

[25] Our preanalysis plan stated that our analysis would use OLS estimation. During the randomization, we realized that, given our data structure, we could increase our statistical power through a random effects model. We used the random effects model to estimate the minimum detectable effect sizes that we calculated prior to implementing the experiment. Accordingly, the random effects model is also our preferred specification for the analysis. We report OLS estimates for our main results alongside the random effects estimates in tables 2, A.3, and A.6.

[26] Contract and time preferences are included in our analysis as indicator variables for above- or below-the-median preference for loss-contract framing (relative to gain-contract framing) and end-of-semester payments (relative to start-of-semester payments), respectively. Indicator variables avoid the need to assign values to top- and bottom-coded data. If we cannot estimate an instructor's preference in the fall or spring semester due to missing or incomplete surveys, we substitute the value measured in the other semester. This affects contract-preference values for two fall and 11 spring instructors and time-preference values for two fall and 10 spring instructors.

instructor incentives alone, student incentives alone, and combined incentives, respectively. Based on our realized sample size and randomization, we estimate a minimum detectable effect size (MDES) for our primary outcome—performance on the objective exam—of 0.17 SD for instructor incentives, identical to our preanalysis plan, and of just under 0.22 SD for combined incentives, compared to an MDES of 0.2 SD in our preanalysis plan. We powered the study with a larger MDES for combined incentives, given their higher cost and our interest in testing the hypothesis that combined incentives have larger effects than instructor incentives alone. The pairwise test, $\beta_3 > \beta_1$, itself has an MDES of 0.25 SD. We did not have a large enough sample size to adequately power a test of student incentives alone, $\beta_2$, or the full test of complementarities between instructor and student incentives, which compares the effect of combined incentives to the sum of the effect of instructor incentives and the effect of student incentives, $\beta_3 > \beta_1 + \beta_2$.

For our primary treatment effect estimates (in tables 2–4), we calculate $p$-values using randomization inference. This procedure compares our observed treatment effect to hypothetical treatment effects for 5,000 simulated counterfactual random assignments using the randomization specification discussed in section III.D. Our $p$-value is then the percentage of counterfactual treatment effects that exceed our observed treatment effect. Due to computational constraints, we do not use randomization inference to test equality of the effects of instructor incentives and combined incentives.

### F. *Data and Baseline Characteristics*

We collected course data for 6,241 student-course observations in 383 sections. Our administrative data set does not have demographic characteristics for 175 student-course observations, leaving us with a final sample size of 6,066 student-course observations for 3,575 unique students. We are missing exam data from eight instructors in the fall semester and three instructors in the spring semester, yielding 5,839 student-course observations with valid exam data.[27] There are no differences by treatment in the rate of missing baseline characteristics or exam data (table A.2;

[27] Table A.1 shows the distribution of students, instructors, and courses with valid exam data and with course data across the two semesters by treatment. Exam scores are not individually recorded in the administrative data and had to be collected by our data-collection team for the study. One instructor left Ivy Tech in the middle of the fall semester, and the replacement instructor did not submit exams to our data-collection team. Also in the fall semester, six instructors (teaching seven courses) in the business, operations, applications, and technology departments recorded grades for their exams as pass or fail instead of recording scores. One additional instructor from the fall and three from the spring failed to submit their exams to our data-collection team for unknown reasons and were unavailable when we repeatedly attempted to follow up.

tables A.1–A.14 are available online). Nonetheless, to address concerns about missing data, we run our analysis on the exam data and the course data separately.

Table 1 reports means and proportions (with standard errors clustered by instructor) for baseline characteristics by semester and treatment. At the student level, these characteristics include age, gender, race/ethnicity, total credits accumulated at baseline, baseline GPA, and whether GPA is missing (all newly entering students and some students returning after long absences have missing baseline GPAs); at the instructor level, they are full-time or adjunct, total sections in the study, students per section, and elicited contract and time preferences; and at the section level, it is whether the course section is a corequisite (the Ivy Tech corequisite course model is a form of remedial education for underprepared students that operates concurrently with the enrolled course).

To show we are balanced on baseline characteristics, we report the $p$-value from a joint test of equality across all treatment groups within each semester using the same random effects specification we use in our analysis. For the spring semester, we also report significant differences of means from binary tests comparing each treatment group to the control group. Of the 48 pairwise tests of differences we conduct, one is significant at the 10% level and one at the 5% level, slightly less than what would be expected by chance.

## IV. Outcomes in Targeted Courses

### A. *Exam Performance*

We first examine the effect of treatment on the directly incentivized outcome: performance on the objective course exam. Figure 1 shows the distribution of test scores in the control group and in the instructor incentives group. The vertical line at a score of 70% indicates the threshold for passing the exam, which was the basis for instructors to receive the incentive. Students who withdrew from the course after the drop deadline are coded as having received a zero on the final exam, as are students who complete the course but do not take the exam.[28] The figures show that instructor incentives increase both student persistence in the course (i.e., nonzero exam scores) and scores among those who take the exam. Throughout the distribution, instructor incentives shift scores to the right.

[28] We note that our setting differs from most K–12 contexts in which all students are required to take the exam, and so those who do not take exams are considered attriters. In that context, missing scores should not necessarily be coded as zero. In our study, student withdrawal from the course—and therefore failure to take the exam—is an outcome of interest and not subject to the typical attrition concerns. As shown in table 3, our results are robust to examining alternative outcomes that do not depend on assigning a zero score to students who withdraw.

TABLE 1
BASELINE CHARACTERISTICS BY TREATMENT AND SEMESTER

| | Fall | | | Spring | | | | |
|---|---|---|---|---|---|---|---|---|
| | Control | Instructor Incentives | Fall *F*-Test | Control | Instructor Incentives | Combined Incentives | Student Incentives | Spring *F*-Test |
| Student-level characteristics: | | | | | | | | |
| Age | 24.385 | 24.594 | .343 | 25.228 | 24.758 | 24.787 | 24.659 | .839 |
| | (.423) | (.516) | | (.473) | (.545) | (.613) | (.642) | |
| Male | .342 | .340 | .926 | .326 | .274** | .322 | .325 | .135 |
| | (.028) | (.023) | | (.034) | (.029) | (.026) | (.044) | |
| White | .817 | .818 | .805 | .817 | .831 | .830 | .859* | .393 |
| | (.011) | (.015) | | (.015) | (.015) | (.018) | (.017) | |
| Baseline credits | 12.403 | 13.018 | .717 | 18.011 | 19.442 | 19.236 | 17.580 | .660 |
| | (1.012) | (1.665) | | (1.442) | (2.525) | (1.315) | (1.986) | |
| Baseline GPA | 2.843 | 2.804 | .298 | 2.934 | 2.925 | 2.950 | 2.790 | .327 |
| | (.037) | (.050) | | (.048) | (.068) | (.056) | (.066) | |
| Missing GPA | .372 | .371 | .487 | .498 | .446 | .416 | .514 | .395 |
| | (.030) | (.030) | | (.028) | (.039) | (.035) | (.041) | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Instructor-level characteristics: | | | | | | | | |
| Full-time | .377 | .345 | .733 | .442 | .514 | .378 | .368 | .284 |
| | (.067) | (.065) | | (.077) | (.083) | (.081) | (.114) | |
| Total sections | 1.981 | 1.727 | .273 | 2.093 | 2.000 | 2.000 | 2.211 | .898 |
| | (.190) | (.131) | | (.182) | (.182) | (.182) | (.224) | |
| Students per section | 16.467 | 16.342 | .906 | 14.200 | 15.893 | 15.164 | 14.421 | .520 |
| | (.749) | (.750) | | (.709) | (.866) | (.697) | (1.042) | |
| Below median | .509 | .455 | .572 | .558 | .378 | .459 | .684 | .151 |
| Contract value | (.069) | (.068) | | (.077) | (.081) | (.083) | (.110) | |
| Below median | .642 | .691 | .591 | .651 | .757 | .676 | .632 | .523 |
| Discount rate | (.067) | (.063) | | (.074) | (.072) | (.078) | (.114) | |
| Section-level characteristics: | | | | | | | | |
| Corequisite | .276 | .263 | .905 | .056 | .068 | .044 | .000 | .542 |
| | (.079) | (.075) | | (.055) | (.065) | (.044) | (.000) | |
| Observations | 1,776 | 1,592 | | 1,043 | 657 | 687 | 311 | |

Note.—Shown are means/proportions for each group, with standard errors (in parentheses) clustered at the level of randomization (instructor). At the student level, we conducted a random effects regression; *F*-tests report the *p*-value of a joint test that all treatment coefficients are zero. At the instructor and section levels, *p*-values are reported for a joint orthogonality test across treatment groups. Significance tests are against the control mean.

* $p < .10$.

** $p < .05$.

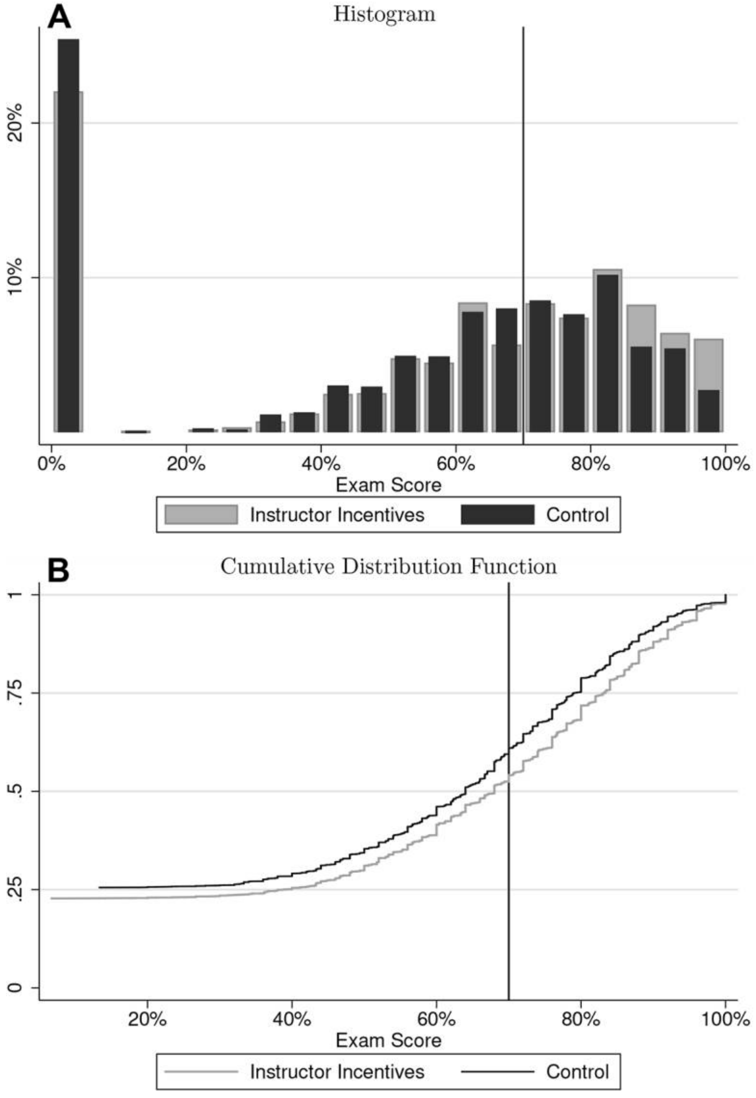


FIG. 1.—Distribution of exam scores in the instructor incentives and control groups. *A*, Histogram in 5 percentage point intervals. *B*, Cumulative distribution function of exam scores for all students in courses taught by instructors in the instructor incentives or control treatments. The vertical line at 70% represents the cutoff for "passing" the exam.

Figure 1*A* shows that, while there is some evidence of missing mass just below the 70% threshold, there is no evidence of a corresponding increase just above the threshold. Moreover, the treatment group shows a notable increase in the proportion of students scoring 90% or higher. As shown in figure 1*B*,

TABLE 2
EFFECTS OF INCENTIVES ON EXAM SCORES

| | Pooled Fall and Spring Semesters | | | | Semester | |
|---|---|---|---|---|---|---|
| | Ordinary Least Squares | | Random Effects | | Fall | Spring |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Instructor incentives | .172 | .161 | .202 | .204 | .113 | .247 |
| | (.063) | (.060) | (.057) | (.056) | (.057) | (.085) |
| *p*-value | .004 | .006 | <.001 | <.001 | .043 | .004 |
| Combined incentives | .111 | .099 | .162 | .156 | | .174 |
| | (.087) | (.083) | (.078) | (.075) | | (.085) |
| *p*-value | .086 | .107 | .018 | .019 | | .026 |
| Student incentives | −.042 | −.052 | .070 | .066 | | .067 |
| | (.089) | (.092) | (.074) | (.076) | | (.083) |
| *p*-value | .659 | .687 | .198 | .207 | | .233 |
| Department | Yes | Yes | Yes | Yes | Yes | Yes |
| Instructor type | Yes | Yes | Yes | Yes | Yes | Yes |
| Baseline characteristics | No | Yes | No | Yes | Yes | Yes |
| *p*-value (instructor = combined) | .384 | .308 | .573 | .485 | | .310 |
| Instructors | 127 | 127 | 127 | 127 | 100 | 96 |
| Observations | 5,839 | 5,839 | 5,839 | 5,839 | 3,189 | 2,650 |

NOTE.—Values in cols. 5 and 6 are random effects linear estimations. Standard errors (in parentheses) are clustered at the instructor level. Exam score standardized within department (mean: 0; SD: 1) is the dependent variable. Values in the pooled semester columns include semester fixed effects. All regressions control for the randomization strata: full-time status and department. Columns 2 and 4–6 add student covariates (age, gender, race/ethnicity, baseline credits), instructor covariates (contract value, discount rate), and course covariates (corequisite). Exact *p*-values are calculated based on randomization inference.

the instructor incentives distribution stochastically dominates the control distribution. A Kolmogorov-Smirnov test clearly rejects that the scores have equal distributions ($p < .01$).

In the regression analysis above, we examine test scores along several margins. We begin with the full effect on test performance by combining the extensive margin (taking the exam) and the intensive margin (test score). Unless otherwise noted, test scores are normalized within department to have a mean of zero and a standard deviation of one. Table 2 displays the results of the regression analysis first using OLS estimation in columns 1 and 2 and then using a random effects model in columns 3–6. We estimate treatment effects for the full year (cols. 1–4) and by semester (cols. 5 and 6 for fall and spring, respectively).[29] In columns 1 and 3, we include only indicators for treatment, semester, and the covariates used for stratification during the randomization: academic department and instructor type (adjunct or full-time). In all other columns, we add

[29] Columns 3 and 4 of table A.3 report the OLS estimates by semester. The results do not change.

controls for the following baseline characteristics reported in table 1: student age, gender, race, and credits accumulated; instructor contract and time preferences (using indicator variables for above/below median preference in the sample); and whether the course is a corequisite.[30] We report standard errors clustered by instructor in parentheses and $p$-values that are derived from exact tests using randomization inference (Gerber and Green 2012). We also report $p$-values at the bottom of the table from a test of equality of the effects of instructor incentives and combined incentives.[31]

In all specifications, instructor incentives have an economically meaningful and statistically significant impact on student outcomes. In the full-year sample, instructor incentives improve student exam performance by between 0.16 and 0.20 SD, or just more than 6 percentage points off a control group mean of 52% ($p < .01$ in all specifications). The estimated effects are smaller in the fall semester than in the spring semester, though we cannot reject that the effect sizes are equal ($p = .53$). The pattern of effects holds if we restrict the spring sample to the subset of returning instructors who also received incentives in the fall (table A.3, col. 8). These results suggest that the effects of the incentives sustain themselves beyond the first time they are offered. Through a series of quantile regressions (table A.4), we also confirm the broad distributional impact of instructor incentives shown in figure 1. Instructor incentives cause large and significant increases in exam performance at the 10th, 30th, 50th, 70th, and 90th percentiles of exam scores, both unconditional on exam taking and conditional on taking the exam.[32]

We find no evidence that student incentives increase the effect of instructor incentives. In all specifications, the estimated effect of combined incentives is economically meaningful, 0.10–0.16 SD, and statistically significant. However, the estimated effects are always smaller than those for instructor incentives—about half to three-quarters the size—though the

[30] Our analysis differs from our preanalysis plan in two ways. First, our preanalysis plan includes GPA as a student-level covariate. We exclude GPA from our main analysis because it is missing for a substantial fraction of students. Second, our preanalysis plan did not include corequisite classification as a course-level covariate because we were not aware of this classification at the time. Columns 6 and 7 of table A.3 repeat the analysis including GPA in two different ways. In col. 6, we impute missing GPAs as the mean GPA and include an indicator for whether GPA is missing. In col. 7, we run the analysis including GPA as a covariate and excluding students who are missing GPA. Neither specification affects the results. Column 5 of table A.3 repeats our analysis excluding the covariate for corequisite courses. No results are affected.

[31] As discussed in sec. III.E, due to computational constraints, we do not use randomization inference to calculate exact $p$-values for binary outcomes or to test equality of effects across treatments.

[32] We are unable to estimate the impact on the 10th percentile of exam scores unconditional on taking the exam because more than 10% of the sample in both the treatment and control groups received scores of zero. The 70% threshold for passing the exam lies at about the 60th percentile of the unconditional control distribution.

two treatments are never statistically distinguishable.[33] The estimated impact of student incentives is noisy, with a small point estimate that varies in sign across specifications and is never statistically significant. We note that, as discussed in section III.D, the student incentives treatment is underpowered, and so throughout our analysis, we interpret the estimates with caution (for a discussion of properly interpreting results from small samples, see Maniadis, Tufano, and List 2014).

We next explore the mechanisms behind the impact on exam scores by examining treatment effects on both the extensive and intensive margins. Table 3 estimates the effects of incentives in the full-year sample on exam pass rates (i.e., the threshold used for the incentives), exam-taking rates, and exam scores conditional on taking the exam. The extensive-margin results (exam-taking and pass rates) report marginal effects from a random effects probit regression. For all outcomes, we first present the results including only the subset of covariates in column 1 of table 2 and then with the full set of covariates used in column 2 of table 2.

Instructor incentives have large and statistically significant effects on all margins. Instructor incentives increase exam pass rates by 7.4–7.8 percentage points ($p < .01$), a 19% increase compared to the 40% pass rate in the control group. The increased pass rates reflect both higher persistence in the course and improved performance on the exam. Instructor incentives increase rates of taking the exam by 5.3 percentage points ($p < .01$). They also increase scores at the intensive margin, with exam scores improving an estimated 0.08 SD ($p = .04$) among the students who take the exam. This is particularly noteworthy given that the positive extensive-margin effect might suggest that instructor incentives induce more marginal students to take the exam, which could depress conditional exam scores. Turning to combined incentives, we find a large impact on pass rates of 8.0–8.2 percentage points ($p = .02$) and on scores among those who take the exam, which increase by an estimated 0.14 SD ($p < .01$). However, there is no effect on the extensive margin of taking the exam. Student incentives have little impact at either margin.[34]

[33] To address the concern that the estimated impact of instructor incentives is being buoyed by excluding less successful combined incentives sections, we also estimate effects on exam scores pooling instructor incentives with combined incentives. As shown in cols. 1 and 2 of table A.5, the estimated effects of pooled instructor incentives are large and statistically significant, 0.19 SD ($p < .01$), where col. 1 pools the student incentives and control groups and col. 2 includes a separate indicator for student incentives. We estimate an interaction specification in cols. 3 and 4 for the full-year sample and the spring semester (the only semester with all four treatments). The negative coefficients on the interaction term between instructor incentives and student incentives implies a subadditivity of combined incentives of between −0.12 SD ($p = .23$) and −0.15 SD ($p = .18$), which is economically meaningful but not statistically significant.

[34] We present the analogous results using OLS estimation in table A.6, cols. 1–3. No results are affected.

TABLE 3
Effects of Incentives on Extensive and Intensive Margins

| | Pass Exam | | Take Exam | | Score If Taken | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Instructor incentives | .074 | .078 | .053 | .053 | .079 | .083 |
| | (.030) | (.029) | (.019) | (.018) | (.043) | (.040) |
| *p*-value | .010 | .006 | .003 | .002 | .053 | .043 |
| Combined incentives | .080 | .082 | .016 | .013 | .136 | .136 |
| | (.036) | (.034) | (.030) | (.029) | (.047) | (.045) |
| *p*-value | .020 | .018 | .246 | .296 | .004 | .005 |
| Student incentives | .038 | .037 | .014 | .013 | .053 | .053 |
| | (.046) | (.048) | (.027) | (.028) | (.047) | (.047) |
| *p*-value | .203 | .208 | .329 | .334 | .149 | .144 |
| Control group mean | .400 | | .746 | | .450 | |
| | (.49) | | (.44) | | (.51) | |
| Department | Yes | Yes | Yes | Yes | Yes | Yes |
| Instructor type | Yes | Yes | Yes | Yes | Yes | Yes |
| Baseline characteristics | No | Yes | No | Yes | No | Yes |
| Instructors | 127 | 127 | 127 | 127 | 127 | 127 |
| Observations | 5,839 | 5,839 | 5,741 | 5,741 | 4,421 | 4,421 |

Note.—Columns 1–4 present marginal effects from random effects probit estimation. Standard errors (in parentheses) are clustered at the instructor level. Columns 5 and 6 use a random effects linear regression with robust standard errors. Standard deviation is reported for control group mean. Sample size is smaller for cols. 3 and 4 because one department perfectly predicts taking the exam. The dependent variable for cols. 5 and 6 is exam score standardized within department (mean: 0; SD: 1). All analysis includes semester and department fixed effects. Columns 2, 4, and 6 add student covariates (age, gender, race/ethnicity, baseline credits), instructor covariates (type, contract value, discount rate), and course covariates (corequisite). Exact *p*-values are calculated based on randomization inference.

## B. *Course Performance*

One concern with using exam scores to measure improvement is that instructors may “teach to the test” in ways that do not improve (or may even detract from) unincentivized elements of the course. To address this, we explore the impact of incentives on overall course outcomes. Course grades were collected at the administrative level and were not directly incentivized. Thus, they provide a robustness check for our exam score results. Course grades depend partly on final exam scores and so are not entirely independent. We address this potential confound by also estimating the impact of incentives on course performance excluding the exam.

Table 4 reports the impact of our treatments in the full-year sample. We first estimate the impact of incentives on course grades in grade points (col. 1) and then in standardized units (col. 2). We use the standard 0–4 scale of grade points corresponding to A–F grades, with withdrawals counting as zero grade points. We normalize grades within each department to have a mean of zero and a standard deviation of one. In column 3, we estimate effects on course grades excluding exam scores. We do this to address concerns that the impact on course grades is simply mechanical,

TABLE 4
EFFECTS OF INCENTIVES ON COURSE PERFORMANCE

| | COURSE GRADE | | | | |
|---|---|---|---|---|---|
| | Grade Points (1) | Standardized (2) | Excluding Exam (3) | COMPLETE COURSE (4) | GRADE IF COMPLETED (5) |
| Instructor incentives | .165 | .108 | .097 | .037 | .037 |
| | (.080) | (.052) | (.049) | (.019) | (.051) |
| $p$-value | .015 | .016 | .024 | .028 | .194 |
| Combined incentives | .113 | .071 | .041 | −.010 | .094 |
| | (.113) | (.072) | (.069) | (.029) | (.063) |
| $p$-value | .114 | .129 | .241 | .657 | .060 |
| Student Incentives | −.100 | −.058 | −.068 | −.002 | −.068 |
| | (.078) | (.051) | (.053) | (.025) | (.048) |
| $p$-value | .782 | .752 | .797 | .515 | .847 |
| Control group mean | 2.08 | −.02 | −.01 | .783 | .343 |
| | (1.55) | (.99) | (1.00) | (.41) | (.80) |
| $p$-value (instructor = combined) | .603 | .560 | .378 | .071 | .165 |
| Instructors | 134 | 134 | 127 | 130 | 134 |
| Observations | 6,066 | 6,066 | 5,839 | 5,951 | 4,797 |

NOTE.—Columns 1–3 and 5 use a random effects linear estimation; col. 4 presents marginal effects from random effects probit estimation. Standard errors (in parentheses) are clustered at the instructor level. Standard deviation is reported for control group mean. The dependent variable in cols. 2, 3, and 5 is course grade standardized within department (mean: 0; SD: 1). Column 3 calculates grade net of weighted exam score (weights estimated using course syllabi and discussion with administrators). The smaller sample size in col. 3 is due to inclusion of only observations with exam scores. The smaller sample size in col. 4 is due to one department perfectly predicting completion. All estimations include semester and department fixed effects and covariates for student (age, gender, race/ethnicity, baseline credits), instructor (type, contract value, discount rate), and course (corequisite). Exact $p$-values are calculated based on randomization inference.

that is, due to the increases in the exam scores on which teachers were incentivized. We calculate the nonexam grade by subtracting the exam score weighted by the percentage of the course grade that the syllabus attributes to the incentivized test (5%–25% across courses). We then reweight the remaining performance and standardize it within department.[35] Column 4 reports the marginal probability that a student completes the course, estimated using a random effects probit regression. Column 5 presents the impact of incentives on the course grade conditional on course completion. All estimations use the same controls as column 2 of table 2.

The effects of instructor incentives carry over to course outcomes, increasing course grades by 0.11 SD ($p = .02$) or 0.16 grade points ($p = .02$) off a control group mean of 2.08. Examining the effect on the nonexam course grade, we estimate that instructor incentives generate a

[35] We received letter grade data, so we assume that course performance is at the midpoint of the grade scale (e.g., a B is assumed to be 85%). When we were unable to acquire a course syllabus, we used 25% as an upper bound of the weight assigned to the test because this was the upper limit given to us by our administrative partners at Ivy Tech.

0.10 SD increase in course performance that is not attributable to the incentivized exam ($p = .02$). These results suggest that students experience broad learning gains that are not driven by instructors simply "teaching to the test." The impact on course grades is driven in part by course completion rates, which increase by 3.7 percentage points ($p = .03$). This represents a 17% reduction in the baseline dropout rate of about 22%. As noted above, course completion is a critical outcome for students, who receive no refund for the course if they withdraw and must meet a minimum completion rate to retain their financial aid. There is also a small increase at the intensive margin—course grades conditional on completion—that is not statistically significant. Combined incentives have no impact on course completion but do significantly improve course grades both unconditionally and conditional on completing the course. As with the exam results, we find little impact of student incentives (there is suggestive evidence of a small negative effect on course grades).[36]

### C. *Heterogeneity*

We stratified our randomization on instructor classification: full-time or adjunct faculty. This guarantees that we are balanced along this dimension and allows us to test for differential effects by instructor type—48% of our students (as well as 48% of sections) are instructed by full-time faculty, while 52% are instructed by adjunct faculty. Table 5 estimates the effects of incentives on exam score, exam passing, course grade, and course completion by instructor type. Our specification includes indicator variables for instructor type, the full set of interactions of each instructor type with each treatment, and the full set of covariates. We also report $p$-values from tests of equality of the marginal impact of instructor incentives for full-time versus adjunct faculty.

Our results suggest that there are heterogeneous effects by instructor type. Under instructors incentives, the exam scores and course grades of students taught by adjunct faculty improve by approximately 0.26 SD ($p < .01$) and 0.19 SD ($p < .01$), respectively. Instructor incentives also increase exam pass rates and course completion rates among adjunct faculty by an estimated 10 percentage points ($p < .01$) and 7 percentage points ($p < .01$), respectively.[37] For full-time faculty, the estimated effect

[36] We present the analogous results using OLS estimation in table A.6, cols. 4–6, and in table A.7 using random effects estimation with the sparse set of controls included in table 2, col. 1. The estimated effects are similar but generally estimated with less precision.

[37] The effects of instructor incentives among adjunct instructors are robust to corrections for multiple hypothesis testing using the method described in Anderson (2008). The estimated effects remain statistically significant for exam scores ($p < .01$), exam pass rates ($p = .05$), course grades ($p = .02$), and course completion ($p = .04$). The marginally significant impact of combined incentives on exam pass rates among full-time faculty does not survive the correction ($p = .24$).

TABLE 5
TREATMENT EFFECTS BY INSTRUCTOR TYPE

| | Exam Score (1) | Pass Exam (2) | Course Grade (3) | Complete Course (4) |
|---|---|---|---|---|
| Adjunct: | | | | |
| Instructor incentives | .262*** | .104*** | .186*** | .068*** |
| | (.073) | (.039) | (.062) | (.025) |
| Combined incentives | .126 | .072 | .031 | −.034 |
| | (.094) | (.044) | (.091) | (.035) |
| Student incentives | .100 | .081 | −.087 | .003 |
| | (.102) | (.061) | (.058) | (.031) |
| Full-time: | | | | |
| Instructor incentives | .128 | .041 | .002 | .002 |
| | (.082) | (.039) | (.084) | (.029) |
| Combined incentives | .189 | .092* | .113 | .014 |
| | (.119) | (.053) | (.114) | (.042) |
| Student incentives | .020 | −.036 | −.010 | −.010 |
| | (.091) | (.058) | (.081) | (.033) |
| Control group mean: | | | | |
| Adjunct | −.093 | .396 | −.026 | .772 |
| | (1.01) | (.49) | (1.00) | (.42) |
| Full-time | −.044 | .404 | −.015 | .793 |
| | (.98) | (.49) | (.98) | (.41) |
| Instructor incentives: *p*-value (adjunct = full-time) | .221 | .257 | .075 | .079 |
| Instructors | 127 | 127 | 134 | 130 |
| Observations | 5,839 | 5,839 | 6,066 | 5,951 |

NOTE.—Columns 1 and 3 use a random effects linear estimations; cols. 2 and 4 present marginal effects from a random effects probit estimation. Standard errors (in parentheses) are clustered at the instructor level. Standard deviations are reported for the control group mean. The dependent variable in cols. 1 and 3 are standardized within department (mean: 0; SD: 1). All estimations include covariates for instructor type and instructor type interacted with each treatment, semester and department fixed effects, and covariates for student (age, gender, race/ethnicity, baseline credits), instructor (contract value, discount rate), and course (corequisite). The smaller sample size in col. 4 is due to one department perfectly predicting completion.

* $p < .10$.

*** $p < .01$.

of instructor incentives on exam scores and pass rates are 0.13 SD and 4.2 percentage points, respectively, which are economically meaningful but not statistically significant. There is no discernible impact on course grades or course completion. The estimated effects of combined incentives and student incentives are similar to their effects in the full sample and do not appear to vary across instructor type.

We also stratified the randomization on department. Figure 2 presents the estimated within-department effects of instructor incentives on the normalized exam scores with 95% confidence intervals. The included departments are accounting (ACCT), anatomy and physiology (APHY), art history (ARTH), biology (BIOL), business operations applications and technology (BOAT), business (BUSN), communications (COMM),

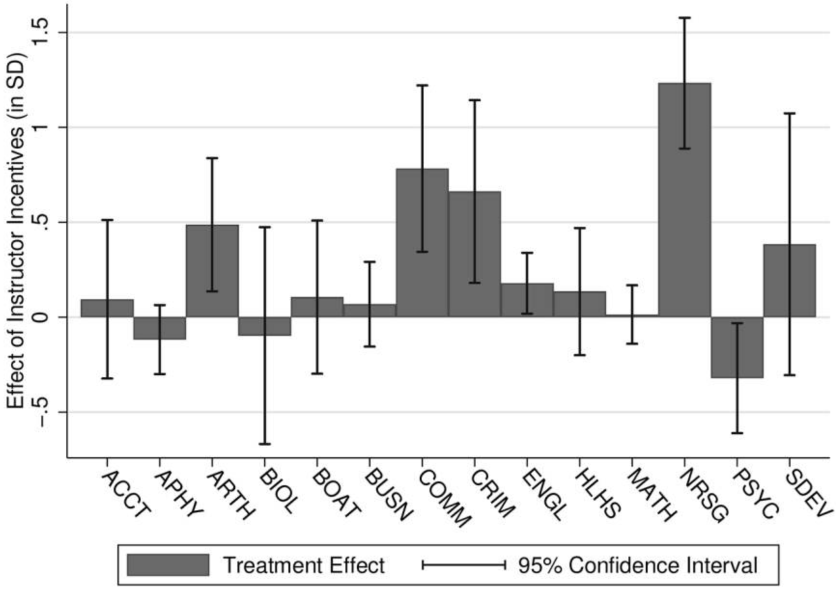


FIG. 2.—Effects of instructor incentives on standardized exam scores by department. Shown are coefficients and 95% confidence intervals for instructor incentives from a random effects linear estimation with standard errors clustered by instructor within the following departments: accounting (ACCT), anatomy and physiology (APHY), art history (ARTH), biology (BIOL), business operations applications and technology (BOAT), business (BUSN), communications (COMM), criminology (CRIM), English (ENGL), health sciences (HLHS), math (MATH), nursing (NRSG), psychology (PSYC), and software development (SDEV). The dependent variable is exam score standardized within department. All regressions include student-level covariates (age, gender, race/ethnicity, and baseline credits).

criminology (CRIM), English (ENGL), health sciences (HLHS), math (MATH), nursing (NRSG), psychology (PSYC), and software development (SDEV).[38] While the small sample sizes within each department increase the error in our estimates, we find positive effects across the vast majority of departments. Psychology is the only department where the estimated effects are even suggestively negative. These results suggest that instructor incentives can be effective across a wide range of departments.

In exploratory analysis (table A.8), we consider additional dimensions of instructor heterogeneity that may influence treatment effectiveness. We

[38] The physical science and sociology departments did not have enough variation in treatment to estimate effects within department. The estimates include student-level covariates (age, gender, race, baseline credits). There was not enough variation within department to include semester fixed effects or instructor-level or course-level covariates.

find no significant heterogeneity by gender, degree type, self-reported beliefs about how much students can improve, amount of the up-front bonus the instructor spent, online versus offline delivery, or preferences for loss contracts. These results suggest that the impact of instructor incentives is not limited to certain types of instructors and thus that incentives can be broadly effective.

### D. *Expenditures, Time Use, and Well-Being*

A potential mechanism for the impact of instructor incentives on student performance is through changes in instructors' financial expenditures or time use. As shown in table 6, we find little evidence of meaningful impacts on instructors' self-reported money spent on course materials or professional development. Similarly, we find no effect on self-reported time spent during a typical week on teaching-related activities or outside employment.[39] These results suggest that the impact of incentives on performance may arise from more subtle changes to teaching that are not captured by hours spent on particular tasks—such as learning students' names or instructors' sharing their personal phone number—or could be due to the difficulty of accurately measuring time use. We also surveyed instructors about their personal and professional well-being. We find little evidence that the incentives meaningfully affect well-being or stress or that instructors who work under them indicate a higher likelihood of leaving their job (table A.10). Similarly, we analyze administrative data on course evaluations for every course and find no evidence of treatment effects on a series of measures such as "I would recommend this instructor to others" (table A.11).

## V. Broad Educational Outcomes

In the section above, we show that instructor incentives have a large impact on student performance in targeted courses. We demonstrate improvements on the objectively measured, directly incentivized exam, as well as the overall course grade that was not directly incentivized (both inclusive and exclusive of the exam). In this section, we examine the impact of incentives on students' broader educational outcomes. We first

[39] We report means by treatment group and statistical differences from a random effects regression including controls for semester and instructor type (full-time or adjunct), with standard errors clustered at the instructor level. Instructors report expenditures for a \$0–\$500 range and time use for a range of 0–16 hours in each category. We note that half of the responses for outside employment are top coded at 16 hours. Table A.9 repeats the analysis with Lee (2009) bounds to correct for nonresponse on the survey.

TABLE 6
Self-Reported Expenditures and Time Use

| | Control | Instructor Incentives |
|---|---|---|
| Expenditure of personal funds (US$): | | |
| Class materials | 66.375 | 75.863 |
| | (19.966) | (13.201) |
| Professional development | 96.550 | 132.182 |
| | (24.170) | (23.056) |
| Time use (hours): | | |
| Teaching class | 9.727 | 9.483 |
| | (.763) | (.596) |
| Preparing for class | 5.273 | 5.944 |
| | (.542) | (.530) |
| Preparing assignments and exams | 4.268 | 4.079 |
| | (.473) | (.429) |
| Grading assignments and exams | 5.571 | 5.586 |
| | (.578) | (.490) |
| Holding office hours | 5.939 | 6.212 |
| | (.762) | (.785) |
| Helping students outside of office hours | 3.120 | 2.568 |
| | (.455) | (.316) |
| Advising students | 6.115 | 5.118 |
| | (.957) | (.830) |
| Administrative work | 3.264 | 3.977* |
| | (.451) | (.487) |
| Professional development | 2.426 | 2.389 |
| | (.502) | (.431) |
| Outside employment[a] | 8.818 | 11.132 |
| | (1.366) | (.886) |
| Observations | 59 | 91 |

Note.—Shown are means for each outcome by treatment group. Standard errors (in parentheses) are clustered at the instructor level. Observations are at the instructor-semester level. Significance tests were conducted using random effects regression, including controls for semester and instructor full-time status.

a Of 101 responses, 51 are top coded for outside employment (>16 hours).

* $p < .10$.

estimate treatment effects on outcomes during the program, including performance in courses outside our study, overall credit accumulation, and cumulative GPA. Examining courses outside our study allows us to measure impacts on outcomes for which instructors were not incentivized. It also addresses concerns that the impact of incentives in targeted courses may be partially driven by substitution of effort toward incentivized courses at the expense of unincentivized courses; if this is the case, the intervention may have little net impact on students' overall credit accumulation and GPA. Course performance, credit accumulation, and GPA are also the key requirements for students to maintain their federal financial aid, progress toward a degree, and qualify for transfer to 4-year schools. We then investigate students' longer-term outcomes directly. One year after the program ends, we estimate treatment effects on transfers to 4-year schools, degree receipt, and dropout status. This allows us to examine whether the impact of our intervention persists in ways that meaningfully improve student success.[40]

### A. *Educational Outcomes during the Program*

In this section, we estimate the impact of instructor incentives on overall student performance during the program. We focus on three outcomes: performance in courses outside our study, overall credit accumulation, and cumulative GPA. We conduct the analysis at the student level since aggregate treatment exposure does not vary within student and our outcomes of interest are measured at the student level. In order to focus our analysis on the impact of incentivizing instructors, we pool the instructor incentives and combined incentives groups into the treatment group (the control group pools student incentives with control). Any student who took at least one course in our study is in the analysis—a student may be enrolled in program courses in fall 2016, spring 2017, or both. Courses outside our study are those that were ineligible because they lacked an objective course exam or were eligible but the instructor chose not to participate in the study. Across the two semesters, students took an average 7.5 total courses, 1.7 courses in the study, and 0.8 treated courses, with 61% of all students having at least one treated course.

We estimate treatment effects in two ways. First, in our preferred estimation, we estimate the treatment effect per treated course using students' total treated courses (i.e., the total number of courses a student took in which the instructor received incentives). Second, we estimate treatment effects for any treated course, using an indicator for whether a student took at least one course that was exposed to incentives. This presents the difference in means between students who ever experienced a treatment course and those who were never exposed to treatment. In all regressions, we control for a student's total courses in the study. We do this because the treatment variables of interest—total exposure to treatment and the probability of having at least one treated course—are conditionally random after we control for the number of courses a student took that were part of our study (i.e., the number of courses that could have received incentives). We also control for the total number of courses a student was enrolled in during the study (total courses), including courses both in and outside of the study.

Table 7 presents the results. Columns 1–4 report estimates of the impact of incentives on course completion and grades (measured in grade

[40] Our diverse set of courses were not part of set tracks, and so we are not able to examine persistence in specific majors or follow-on courses, as related work on instructor quality has done (e.g., Carrell and West 2010; Ran and Xu 2017; De Vlieger, Jacob, and Stange 2017).

TABLE 7

TREATMENT EFFECTS ON NONPROGRAM COURSES AND OVERALL OUTCOMES

| | COURSES OUTSIDE THE STUDY | | | | OUTCOMES AT END OF PROGRAM | | | |
|---|---|---|---|---|---|---|---|---|
| | Completion | | Grades | | Credits | | GPA | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Total treated courses | .032*** | | .087*** | | .441*** | | .028 | |
| | (.008) | | (.032) | | (.171) | | (.026) | |
| Any treated course | | .021* | | .065 | | .531*** | | .007 |
| | | (.012) | | (.048) | | (.205) | | (.038) |
| Total courses in study | −.024*** | −.012** | −.069*** | −.038* | 2.420*** | 2.547*** | .025 | .038** |
| | (.006) | (.005) | (.026) | (.022) | (.152) | (.132) | (.022) | (.019) |
| Total courses | .014*** | .014*** | .048*** | .048*** | .928*** | .927*** | .036*** | .036*** |
| | (.002) | (.002) | (.007) | (.007) | (.041) | (.041) | (.005) | (.005) |
| Control mean | .778 | | 2.149 | | 24.492 | | 2.420 | |
| | (.342) | | (1.373) | | (22.016) | | (1.173) | |
| Students | 3,535 | 3,535 | 3,535 | 3,535 | 3,575 | 3,575 | 3,575 | 3,575 |

NOTE.—Shown are results from linear regression estimations with heteroskedasticity-robust standard errors (in parentheses). Pure control group mean is for students who have never attended a treated course. Standard deviations are reported for pure control group mean. Observations are at the student level. Total treated courses include both instructor incentives and combined incentives groups. All estimations include student-level covariates (age, gender, race/ethnicity, baseline credits, total courses).

* $p < .10$.

** $p < .05$.

*** $p < .01$.

points) for the courses students were taking outside of our study. In columns 5 and 6 and columns 7 and 8, respectively, we estimate effects on students' credit accumulation and cumulative GPA. These measures include all courses taken prior to and during the treatment semester. We report marginal effects from a probit regression in columns 1 and 2 and use OLS estimation in columns 3–8. Heteroskedasticity-robust standard errors are in parentheses. All regressions include student-level covariates (age, gender, race/ethnicity, baseline credits, total courses). We also report the mean and standard deviation for each outcome among control students who had no treated courses.

During the program, we find large positive spillovers of treatment to courses outside our study. For each course a student takes that is exposed to incentives, completion rates in unincentivized courses increase by an estimated 3.2 percentage points ($p < .01$). The estimated impact on courses outside our study is only slightly smaller than the 3.7 percentage point increase in completion rates in the incentivized courses themselves (table 4). Course grades also increase significantly, by an estimated 0.09 grade points ($p < .01$) per incentivized course. This magnitude is slightly more than half of the estimated impact of 0.17 grade points in the incentivized courses (table 4). The secondary specifications in columns 2 and 4 show similar results for the difference-in-means estimation, but with lower statistical significance.

Our results demonstrate that instructor incentives significantly improve student performance on measures for which instructors were not incentivized. The large impact on completion rates in both program and nonprogram courses suggests that incentives may be leading students to maintain their general enrollment levels during the treatment semester. Thus, instructor incentives may offer a tool for improving student retention throughout the term, which is a critical outcome for community colleges. Importantly, it is unlikely we would find such impacts if our main results were due to gaming (e.g., instructors teaching to the incentivized exam). The findings also show that students are not substituting attention or effort to incentivized courses at the expense of unincentivized courses.[41]

Taking more courses is associated with improvements in every outcome measure, as shown by the positive total courses coefficient. Conditional on total courses, we find a negative association between total courses in

[41] Similarly, in table A.12, we examine treatment effects in targeted courses by instructor and student exposure to incentivized courses. If substitution is driving our results, we would expect larger treatment effects for instructors or students who can concentrate their effort on only one incentivized course compared to instructors or students who have to spread their effort across multiple incentivized courses. In contrast, we find that the estimated effects of incentives on exam scores are larger (though not significantly so) among instructors and students exposed to more incentivized courses.

study and performance in outside courses. This correlation may reflect that the introductory-level courses in our study are more likely to be taken by students who are newer to college and have generally high levels of course dropout (these students also have fewer baseline credits accumulated). The positive causal impact of our intervention on course completion and grades suggests that instructor incentives can help address the low performance of students enrolled in these courses, which serve as the gateway to pursuing higher-level courses and bachelor's degrees.

Taking the impact on incentivized and unincentivized courses together, instructor incentives significantly increase overall credit accumulation at the end of the program. As we discussed above, credit accumulation has meaningful wage returns and is critical for students to progress toward degrees and qualify for transfer to 4-year schools. We estimate that students accumulate 0.44 additional credits ($p = .01$) per incentivized course. This represents an 18% increase in credit accumulation on top of the estimated 2.4 credits from an additional unincentivized course. In our secondary specification, the estimated impact of having any treated course is similarly large and statistically significant. We find no evidence that increased completion comes at the expense of course performance (the effects on cumulative GPA are positive but not statistically significant).

### B. *Postprogram Outcomes*

Finally, we examine students' longer-run outcomes 1 year after the program ends. These include transfer to a 4-year college, graduation with an associate degree, graduation with a certificate, and dropout, which is defined as students who have not transferred or earned a degree/certificate and are not currently enrolled. We report estimated treatment impacts in table 8, which has the same structure as table 7.

One year after the program ends, we find large and significant impacts of instructor incentives on transfers to 4-year schools. For every additional treated course, the transfer rate increases by 2.8 percentage points ($p < .01$). The estimated impact of having at least one treated course is a 2.2 percentage point increase in transfer rates. This represents a 22%–28% increase over a transfer rate of 9.9% among control students. We find no significant effects on graduation with an associate degree, graduation with a certificate, or dropout status.

Similar to the findings in table 7, we find that taking more courses is universally positive for long-run outcomes. Conditional on the total number of courses, however, taking more courses in our study (relative to courses outside of our study) is associated with lower transfer rates, as well as lower dropout rates, that is, still being enrolled without having transferred or earned a degree/certificate. Again, this may reflect that, at baseline, students in introductory courses struggle to accumulate the credits

TABLE 8

Effects of Incentives on Graduation, Transfer, and Dropout

| | Transfer to 4-Year College | | Associate Degree | | Graduate Certificate | | Dropout | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Total treated courses | .028*** | | −.008 | | −.005 | | −.012 | |
| | (.008) | | (.009) | | (.006) | | (.011) | |
| Any treated course | | .022** | | −.011 | | −.004 | | −.006 |
| | | (.011) | | (.013) | | (.008) | | (.016) |
| Total courses in study | −.023*** | −.012** | −.000 | −.002 | .007 | .005 | −.029*** | −.033*** |
| | (.007) | (.006) | (.007) | (.007) | (.005) | (.004) | (.010) | (.009) |
| Total courses | .007*** | .007*** | .028*** | .028*** | .006*** | .006*** | −.022*** | −.022*** |
| | (.002) | (.002) | (.002) | (.002) | (.001) | (.001) | (.002) | (.002) |
| Control mean | .099 | | .236 | | .054 | | .503 | |
| | (.299) | | (.425) | | (.226) | | (.500) | |
| Students | 3,575 | 3,575 | 3,575 | 3,575 | 3,575 | 3,575 | 3,575 | 3,575 |

Note.—Shown are marginal effect estimates from probit regressions with heteroskedasticity-robust standard errors (in parentheses). Eligible courses represent the total number of courses that could have been assigned to the treatment. The pure control mean is for students who never attended a treated course.

** $p < .05$.

*** $p < .01$.

and GPA needed to qualify for transfer to 4-year schools. The large impact of instructor incentives on transfer rates among these students suggests that our intervention helps them better progress through their studies. More broadly, these findings demonstrate that the impact of instructor incentives persists after the program ends and advances students along the pathway to pursue 4-year degrees.

## VI. Contract Preferences

The results in the previous section show that the incentives we offered instructors improved student performance. As we discussed previously, we chose to frame the incentives as losses based on prior work demonstrating the effectiveness of loss-framed incentives in other contexts, including among elementary and middle school teachers. However, there are several open questions about the usefulness of loss contracts as a personnel policy. In particular, these contracts need to be not only effective but also palatable to instructors who otherwise may select out of working at postsecondary institutions that offer loss-framed bonuses. Accordingly, we examine instructors' preferences for the loss contracts we offer. Standard behavioral models predict that people will prefer to work under gain contracts rather than loss contracts. In practice, there is limited empirical evidence on employee preferences between such contracts.

### A. *Baseline Preferences*

When instructors enrolled in our study (either before the fall semester or before the spring semester), we used the incentive-compatible multiple price list mechanism described in section III.C to elicit their baseline preferences between loss and gain contracts. For instructors who participated in the fall semester, we also elicited their contract preferences at the end of the fall semester (we could not incentivize end-line preferences for spring semester instructors because we did not continue the incentives after the spring semester).

From the multiple price list, instructors revealed the price ratio at which they preferred to receive the loss contract, which provides upfront bonuses, rather than a gain contract, which awards bonuses at the end of the semester. We then estimate the per-student bonus amount that an instructor is willing to sacrifice in order to receive a loss contract. Thus, positive values indicate a preference for loss contracts, and negative values indicate a preference for gain contracts; that is, an instructor needs to be paid a higher per-student bonus to work under a loss contract rather than a gain contract.

Figure 3 plots the histogram of baseline contract preferences elicited on the initial enrollment survey (either fall or spring). Preferences are

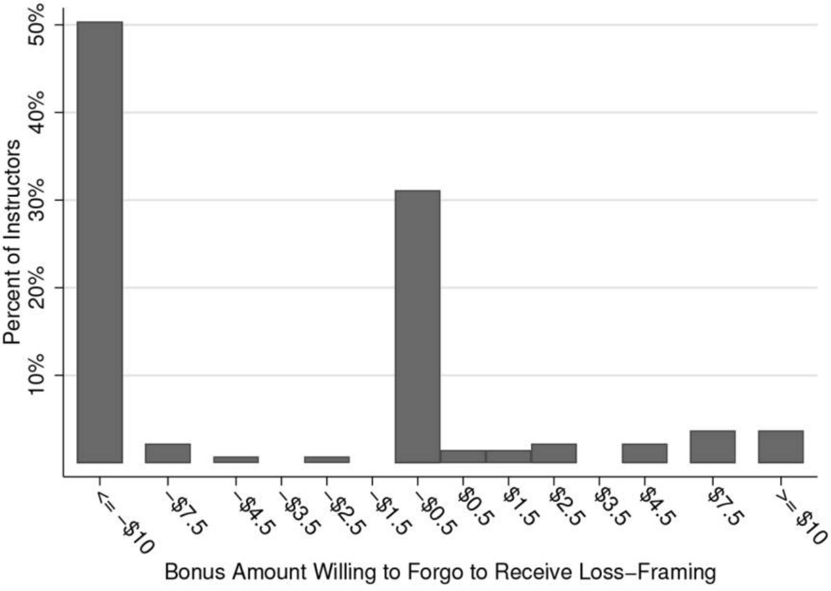


FIG. 3.—Distribution of baseline contract preferences for all instructors. At the far left of the *x*-axis, "<= −$10" indicates instructors who preferred a gain-framed bonus of $40 per student over a loss-framed bonus of $50 per student; at the far right, ">= $10" indicates instructors who preferred a loss-framed bonus of $50 per student over a gain-framed bonus of $60 per student. All other values represent the midpoint between the per-student bonus amounts over which the instructor switches from preferring the gain-framed bonus to preferring the loss-framed bonus.

calculated using midpoint estimation where possible.[42] We find a preference for gain contracts at baseline: on average, instructors prefer gain contracts until loss contracts offer $4.57 more per student, which is equivalent to 9.16% of the $50 per student incentives. For the average instructor, this represents a potential difference in incentive payments between the loss and gain contracts of $138 and an expected difference of $66 using average pass rates in the treatment group.[43] We estimate that this willingness to pay to avoid loss contracts corresponds to a loss-aversion parameter of 1.99, which is in line with the literature from laboratory

[42] When instructors switched only once across the multiple price list, we assigned them a value equal to the midpoint of the interval over which their preferences shifted. When they never switched, we assigned them the minimum or maximum value from the list, ensuring that their assigned preferences exceeded those of anyone who switched in the interior. We dropped anyone who switched multiple times in the list. This drops one instructor in the fall baseline and three in the fall end line.

[43] At a difference of $4.57 per student, the lower expected incentive payment under the gain contract is $4.57/\text{student} \times \Pr[\text{Pass}] \times N = 4.57 \times 0.478 \times 30.1 = 66$, where $\Pr[\text{Pass}] = 0.478$ is the pass rate in the treatment group and $N = 30.1$ is the average number of students per instructor per semester.

experiments (e.g., Tversky and Kahneman 1991; Abdellaoui, Bleichrodt, and L'Haridon 2008).[44]

Approximately half of respondents reveal the strongest preference for gain contracts (i.e., the minimum value), preferring gain contracts even if they offer \$10 less per student than the loss contract. For the average instructor, this represents a potential difference of \$301 and an expected difference of \$144, corresponding to an estimated loss-aversion parameter of at least 4.35. Such choices thus indicate substantial loss aversion. Or, alternatively, it could reflect confusion about the loss contract or lack of attention when taking the survey (i.e., filling in the same contract choice for every decision).

The second most common response, 31% of respondents, is a weak preference for gain contracts. These instructors prefer gain contracts if per-student bonuses are equal but will switch to preferring a loss contract if the gain contract offers \$1 less per student. We categorize these instructors' contract values as −\$0.50, which is the midpoint between −\$1 (when the instructor prefers the loss contract) and \$0 (when the instructor prefers the gain contract). This corresponds to a loss-aversion parameter between 0.43 and 0 (i.e., no reference dependence in preferences).

## B. *Effects of Experience with Incentives on Preferences*

We next examine the effect of treatment on instructor preferences. Figure 4 summarizes the changes in instructor preferences by treatment group between fall baseline and fall end line measured on the enrollment and midyear surveys, respectively.[45] The modal instructor in both the treatment and control groups has no change in preferences. However, treatment instructors are more likely to change their preference and do so toward preferring loss contracts: 38.0% of treatment instructors show increased preference toward loss contracts, compared to only 12.8% of control instructors. Of the treatment instructors who change their preferences, 28.6% move from the strongest preference for gain contracts to a

[44] We calculate the loss-aversion parameter, $\lambda$, such that the lower expected incentive payment under the gain contract equals the disutility from losses under the loss contract. We assume that instructors have rational expectations and anticipate a pass rate equal to the observed mean rate among treatment instructors; loss and gain contracts have identical motivating effects; there is no discounting between the payment dates; and disutility from losses equals $\lambda(x - r)$ for $x < r$, where $x = \Pr[\text{Pass}] \times \$50 \times N$ is the expected incentive payment for an instructor with $N$ students and $r = 0.5 \times \$50 \times N$ is the reference point (i.e., the up-front bonus). Thus, for indifference between a \$50/student loss contract and a \$45.4/student gain contract, an average of $N = 30.1$ students, and an expected pass rate $\Pr[\text{Pass}] = 0.478$, we set $\lambda(.478 \times 50 \times 30.1 - .5 \times 50 \times 30.1) = -66$.

[45] Unlike in the analysis of baseline preferences, we do not analyze changes in contract preferences during the spring semester because, as noted above, the end of the study meant that contract choices on the spring end-line survey could not be incentive compatible.

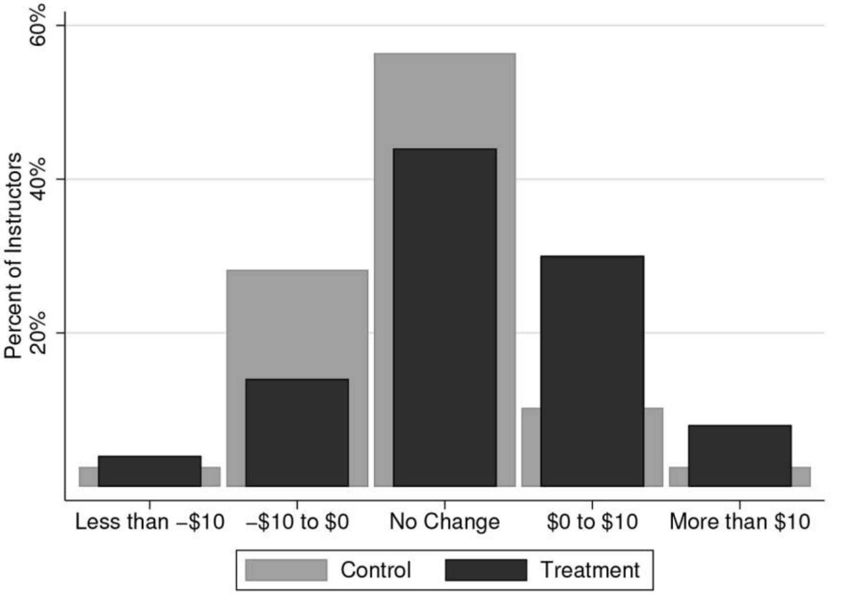


FIG. 4.—Distribution of instructors' changes in willingness to pay for loss-framed contracts between the fall baseline and fall end-line surveys. Positive values indicate increased willingness to pay for loss-framed contracts.

weak preference for gain contracts, making it the most common shift in preferences.[46]

In table 9, we estimate the effect of experience with loss contracts on contract preferences. In the analysis, we use a generalized tobit regression to correct for the interval censoring from our multiple price list. The outcome variable in column 1 is preference for loss contracts in the fall baseline survey, that is, a test of the balance of baseline preferences between treatment and control instructors (we restrict our sample to the instructors for whom we have both baseline and end-line preferences). The outcome variable in columns 2–4 is contract preference on the fall end-line survey. Columns 1–2 control only for instructor type. Column 3 adds controls for baseline contract preferences.[47] Column 4 additionally controls for the instructor's baseline discount rate. At the bottom of the table, we report the $p$-value from a test of whether the treatment group's value

[46] The individual-level changes in preferences for treatment and control instructors are plotted in fig. A.3.

[47] We use dummy variables for five categories of switching points: less than −$10 (i.e., always prefers the gain contract), −$10 to −$0.5, −$0.5 to $0, $0 to $10, and greater than $10 (always prefers the loss contract). Group 3 (−$0.5 to $0; i.e., the median instructor) is the omitted group.

TABLE 9

Treatment Effects on Preference for Loss-Framed Contracts

| | Baseline | Fall End-Line Preference | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Fall instructor incentives | .073 | 4.326** | 4.529** | 4.351*** |
| | (2.307) | (2.069) | (1.782) | (1.666) |
| Full-time | −.191 | −.065 | −.978 | .337 |
| | (2.343) | (1.955) | (1.667) | (1.633) |
| Constant | −6.134*** | −6.055*** | −3.579** | −4.636*** |
| | (2.060) | (1.884) | (1.418) | (1.434) |
| Group 1 (<−$10) | | | −6.492*** | −3.563* |
| | | | (2.022) | (1.894) |
| Group 2 (−$10 to −$.5) | | | −8.056** | −6.979** |
| | | | (3.417) | (3.485) |
| Group 4 ($0 to $10) | | | 2.897 | −1.960 |
| | | | (1.967) | (2.429) |
| Group 5 (>$10) | | | 6.115 | .976 |
| | | | (10.180) | (8.285) |
| Discount rate ($\delta$) | | | | −4.487*** |
| | | | | (1.305) |
| Pr(treatment group value = $0) | .001 | .291 | .499 | .850 |
| Instructors | 89 | 90 | 89 | 88 |

Note.—The dependent variable is the per-student bonus amount an instructor is willing to pay for a loss-framed contract. Estimates are from a generalized tobit regression to correct for interval-censored data. Heteroskedasticity-robust standard errors are shown in parentheses.

* $p < 0.10$

** $p < 0.05$

*** $p < 0.01$.

for the loss contract is equal to $0—that is, if treatment instructors are indifferent between loss and gain contracts of equal value.

Column 1 demonstrates that there are no baseline differences in contract preferences between treatment and control instructors. Both groups significantly prefer gain contracts: the average instructor would need to receive a little more than $6 more per student in order to prefer the loss contract ($p < .01$ for both the treatment and control groups).[48] Columns 2–4 estimate the impact of receiving (loss-framed) instructor incentives during the fall semester. As at baseline, control instructors continue to prefer gain contracts by a little more than $6 ($p < .01$). In contrast, assignment to instructor incentives significantly increases instructor preferences for loss contracts. The treatment effects of $4.33–$4.53 ($p < .05$ in all specifications) largely erase preferences for gain contracts. In the end-line survey, the treatment group's value for the loss contract is no longer statistically distinguishable from zero ($p > .20$ in all specifications). That is, after

[48] The difference between this value and the $4.57 preference stated above is due to our tobit estimation correcting for the interval censoring.

experiencing loss incentives, instructors become (close to) indifferent between the two contract types.[49]

This novel finding—that experience with loss contracts increases preferences for them—may be due to instructors learning that working under loss-framed incentives is less painful (or more beneficial) than they expected or could result from increased familiarity with these unusual contracts. Combined with the persistent impact over multiple semesters shown in table 2, our findings suggest that the motivation from loss contracts is not dependent on instructors finding the contracts unpleasant. These results are consistent with instructors being willing to work under loss contracts because they are motivating; that is, they act as a commitment device for instructors to work harder and earn more. Moreover, the increasing preference for and increasing impact of loss contracts provide support for the sustainability of their use as a policy tool.

## VII. Conclusion

Ours is the first study to test the effect on student performance of an intervention aimed at college instruction. We demonstrate that performance-based incentives for community college instructors have a large impact on student outcomes, equivalent to improving instructor quality by 1 SD. These impacts extend beyond directly incentivized course performance to nonexam grades, courses outside our study, credit accumulation, and post-treatment transfers to 4-year colleges a year after the program ends. At an expected cost of $25 per student-course (or $56 per accumulated credit-hour), instructor incentives represent a relatively low-cost option for improving student performance and encouraging student retention, both critical outcomes for community college students. Incentives may be particularly relevant for adjunct instructors, who experience the largest treatment effects, work under flexible contracts focused on teaching (rather than research), and now represent about 50% and 80% of 4-year and 2-year college instructors, respectively (Hurlburt and McGarrah 2016).

[49] As shown in table A.13, the effects on contract preferences are robust to corrections for differential attrition across treatment and control instructors in the end-line survey. Columns 1 and 2 estimate attrition. The estimated impact of incentives is similar if we assume that missing end-line preferences are the same as at baseline (col. 3) or if we estimate a Lee (2009) upper bound, assuming that aversion to loss contracts increases the likelihood of attrition (col. 4), which seems the likely direction for differential attrition. The effects are smaller and not statistically significant for the lower bound, which assumes that preference for loss contracts increases attrition (col. 5), but this direction of differential attrition seems less likely. As shown in table A.14, the treatment effect on contract preferences does not vary with the amount of money the instructor stands to gain or lose based on the number of students enrolled in their course(s) (cols. 1 and 2) and is unrelated to whether the instructor gained or lost money under the contract (col. 3).

The purpose of this study was to demonstrate that incentives can improve instructor effectiveness at the postsecondary level. We limited our focus to loss-framed incentives, but it may be the case that gain-framed incentives are also effective in this context. Because of instructors' baseline preferences as well as logistical concerns—for example, collecting repayments—gain-framed contracts are potentially preferable. However, it could also be the case that loss-framed contracts serve as a commitment device that instructors learn to prefer because they anticipate working harder under them. Given their demonstrated effectiveness and low cost, we believe that future work is warranted on the optimal design and implementation of incentive contracts for college instructors.

## References

Abdellaoui, M., H. Bleichrodt, and O. L'Haridon. 2008. "A Tractable Method to Measure Utility and Loss Aversion Under Prospect Theory." *J. Risk and Uncertainty* 36 (3): 245.

Anderson, M. L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *J. American Statis. Assoc*. 103 (484): 1481–95.

Angrist, J., D. Lang, and P. Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Econ. J. Appl. Econ*. 1 (1): 136–63.

Angrist, J., P. Oreopoulos, and T. Williams. 2014. "When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards." *J. Human Resources* 49 (3): 572–610.

Armantier, O., and A. Boly. 2015. "Framing of Incentives and Effort Provision." *Internat. Econ. Rev*. 56 (3): 917–38.

Bailey, T. R., S. S. Jaggars, and D. Jenkins. 2015. *Redesigning America's Community Colleges.* Cambridge, MA: Harvard Univ. Press.

Barlevy, G., and D. Neal. 2012. "Pay for Percentile." *A.E.R.* 102 (5): 1805–31.

Barrera-Osorio, F., and D. Raju. 2017. "Teacher Performance Pay: Experimental Evidence from Pakistan." *J. Public Econ*. 148:75–91.

Barrow, L., L. Richburg-Hayes, C. E. Rouse, and T. Brock. 2014. "Paying for Performance: The Education Impacts of a Community College Scholarship Program for Low-Income Adults." *J. Labor Econ*. 32 (3): 563–99.

Barshay, J. 2017. "3.9 Million Students Dropped Out of College with Debt in 2015 and 2016." https://hechingerreport.org/federal-data-shows-3-9-million-students-dropped-college-debt-2015-2016/.

Behrman, J. R., S. W. Parker, P. E., Todd, and K. Wolpin. 2015. "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools." *J.P.E*. 123 (2): 325–64.

Bettinger, E., L. Fox, S. Loeb, and E. Taylor. 2014. "Changing Distributions: How Online College Classes Alter Student and Professor Performance." Working Paper no. 10-15, Center Educ. Policy Analysis, Stanford, CA.

Bettinger, E. P., and B. T. Long. 2006. "The Increasing Use of Adjunct Instructors at Public Institutions: Are We Hurting Students?" In *What's Happening to Public Higher Education?*, edited by Ronald G. Ehrenberg, 51–69. Baltimore: Johns Hopkins Univ. Press.

———. 2010. “Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes.” *Rev. Econ. and Statis*. 92 (3): 598–613.

Braga, M., M. Paccagnella, and M. Pellizzari. 2016. “The Impact of College Teaching on Students’ Academic and Labor Market Outcomes.” *J. Labor Econ*. 34 (3): 781–822.

Brickley, J. A., and J. L. Zimmerman. 2001. “Changing Incentives in a Multitask Environment: Evidence from a Top-Tier Business School.” *J. Corp. Finance* 7 (4): 367–96.

Brodaty, T., and M. Gurgand. 2016. “Good Peers or Good Teachers? Evidence from a French University.” *Econ. Educ. Rev*. 54:62–78.

Brooks, R. R., A. Stremitzer, and S. Tontrup. 2012. “Framing Contracts: Why Loss Framing Increases Effort.” *J. Inst. and Theoretical Econ*. 168 (1): 62–82.

Bruhn, M., and D. McKenzie. 2009. “In Pursuit of Balance: Randomization in Practice in Development Field Experiments.” *American Econ. J. Appl. Econ*. 1 (4): 200–232.

Bureau of Labor Statistics. 2017. Labor Force Statistics from the Current Population Survey. US Dept. Labor. https://data.bls.gov/PDQWeb/le.

Carrell, S. E., and J. E. West. 2010. “Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors.” *J.P.E*. 118 (3): 409–32.

Courant, P. N., and S. Turner. 2017. “Faculty Deployment in Research Universities.” Working Paper no. 23025, NBER, Cambridge, MA.

Czibor, E., D. Hsu, S. Neckermann, and B. Subasi. 2019. “Loss Framed Incentives and Employee (Mis-)behavior.” Working paper. https://sites.google.com/site/czibore/research.

De Fraja, G., T. Oliveira, and L. Zanchi. 2010. “Must Try Harder: Evaluating the Role of Effort in Educational Attainment.” *Rev. Econ. and Statis*. 92 (3): 577–97.

DellaVigna, S., and D. Pope. 2018. “What Motivates Effort? Evidence and Expert Forecasts.” *Rev. Econ. Studies* 85 (2): 1029–69.

Deming, D., and S. Dynarski. 2009. “Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor.” Working Paper no. 15387, NBER, Cambridge, MA.

Denning, J. T. 2017. “College on the Cheap: Consequences of Community College Tuition Reductions.” *American Econ. J. Econ. Policy* 9 (2): 155–88.

De Paola, M., F. Scoppa, and R. Nisticò. 2012. “Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment.” *J. Human Capital* 6 (1): 56–85.

De Philippis, M. 2015. “Multitask Agents and Incentives: The Case of Teaching and Research for University Professors.” Temi di Discussione Working Paper no. 1042, Bank of Italy, Rome.

De Quidt, J. 2018. “Your Loss Is My Gain: A Recruitment Experiment with Framed Incentives.” *J. European Econ. Assoc*. 16(2): 522–59.

De Quidt, J., F. Fallucchi, F. Kölle, D. Nosenzo, and S. Quercia. 2017. “Bonus versus Penalty: How Robust Are the Effects of Contract Framing?” *J. Econ. Sci. Assoc*. 3 (2): 174–82.

De Vlieger, P., B. Jacob, and K. Stange. 2017. “Measuring Instructor Effectiveness in Higher Education.” In *Productivity in Higher Education*, edited by C. M. Hoxby and K. Stange, 209–58. Chicago: Univ. Chicago Press.

Duflo, E., R. Hanna, and S. P. Rya. 2012. “Incentives Work: Getting Teachers to Come to School.” *A.E.R.* 102 (4): 1241–78.

Ehrenberg, R. G. 2012. “American Higher Education in Transition.” *J. Econ. Perspectives* 26 (1): 193–216.

Ehrenberg, R. G., and L. Zhang. 2005. "Do Tenured and Tenure-Track Faculty Matter?" *J. Human Resources* 40 (3): 647–59.

Englmaier, F., S. Grimm, D. Schindler, and S. Schudy. 2018. "Effect of Incentives in Non-routine Analytical Team Tasks—Evidence from a Field Experiment." Rationality and Competition Discussion Paper 71, Collaborative Res. Center TRR 190, Munich.

Evans, W. N., M. S. Kearney, B. C. Perry, and J. X. Sullivan. 2017. "Increasing Community College Completion Rates among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case Management Intervention." Working Paper no. 24150, NBER, Cambridge, MA.

Figlio, D. N., M. O. Schapiro, and K. B. Soter. 2015. "Are Tenure Track Professors Better Teachers?" *Rev. Econ. and Statis*. 97 (4): 715–24.

Fryer, R. G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *J. Labor Econ*. 31 (2): 373–407.

———. 2017. "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments." In *Handbook of Economic Field Experiments*, vol. 2, edited by Esther Duflo and Abhijit Banerjee, 95–322. Amsterdam: North-Holland.

Fryer, R. G., Jr., S. D. Levitt, J. List, and S. Sadoff. 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." Working Paper no. 18237. NBER, Cambridge, MA.

———. 2018. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." Working paper, Univ. Chicago.

Geng, T. 2017. "The Complementarity of Incentive Policies in Education: Evidence from New York City." Working paper, Columbia Univ.

Gerber, A. S., and D. P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York: Norton.

Glazerman, S., A. McKie, and N. Carey. 2009. "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report. Final Report." Mathematica Policy Res., Washington, DC.

Glewwe, P., N. Ilias, and M. Kremer. 2010. "Teacher Incentives." *American Econ. J. Appl. Econ*. 2 (3): 205–27.

Goodman, J., M. Hurwitz, and J. Smith. 2017. "Access to 4-Year Public Colleges and Degree Completion." *J. Labor Econ*. 35 (3): 829–67.

Grubb, W. N. 2002. "Learning and Earning in the Middle, Part I: National Studies of Pre-baccalaureate Education." *Econ. Educ. Rev*. 21 (4): 299–321.

Hoffmann, F., and P. Oreopoulos. 2009. "Professor Qualities and Student Achievement." *Rev. Econ. and Statis*. 91 (1): 83–92.

Hong, F., T. Hossain, and J. A. List. 2015. "Framing Manipulations in Contests: A Natural Field Experiment." *J. Econ. Behavior and Org*. 118:372–82.

Horrace, W. C., and R. L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model." *Econ. Letters* 90 (3): 321–27.

Hossain, T., and J. A. List. 2012. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Sci*. 58 (12): 2151–67.

Hurlburt, S., and M. McGarrah. 2016. "The Shifting Academic Workforce: Where Are the Contingent Faculty?" Report, Delta Cost Project, American Inst. Res., TIAA Inst., New York. http://www.air.org/sites/default/files/downloads/report/Shifting-Academic-Workforce-November-2016.pdf.

Imas, A., S. Sadoff, and A. Samek. 2016. "Do People Anticipate Loss Aversion?" *Management Sci*. 63 (5): 1271–84.

Jenkins, P. D., and J. Fink. 2016. "Tracking Transfer: New Measures of Institutional and State Effectiveness in Helping Community College Students Attain Bachelor's Degrees." Technical report, Community Coll. Res. Center, New York.

Jie, Y. 2018. "Prepayment Effect: Prepayment with Clawback Increases Task Participation." *J. Bus. Res*. 92:210–18.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–91.

Kane, T. J., and C. E. Rouse. 1995. "Labor Market Returns to Two- and Four-Year Colleges." *A.E.R.* 85:600–614.

Kermer, D. A., E. Driver-Linn, T. D. Wilson, and D. T. Gilbert. 2006. "Loss Aversion Is an Affective Forecasting Error." *Psychological Sci*. 17 (8): 649–53.

Kőszegi, B., and M. Rabin. 2006. "A Model of Reference-Dependent Preferences." *Q.J.E.* 121 (4): 1133–65.

Lavecchia, A. M., H. Liu, and P. Oreopoulos. 2014. "Behavioral Economics of Education: Progress and Possibilities." Working Paper no. 20609. NBER, Cambridge, MA.

Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Rev. Econ. Studies* 76 (3): 1071–102.

Leuven, E., H. Oosterbeek, and B. Van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment." *J. European Econ. Assoc.* 8 (6): 1243–65.

Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff. 2016. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." *American Econ. J. Econ. Policy* 8 (4): 183–219.

List, J. A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Q.J.E.* 118 (1): 41–71.

———. 2004. "Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace." *Econometrica* 72 (2): 615–25.

———. 2011. "Does Market Experience Eliminate Market Anomalies? The Case of Exogenous Market Experience." *A.E.R.* 101 (3): 313–17.

List, J. A., J. A. Livingston, and S. Neckermann. 2012. "Harnessing Complementarities in the Education Production Function." Working paper, Univ. Chicago. https://drive.google.com/open?id=1MQCJ-9TmFxo2MS6YsDqtGe2TInAU_5mm.

List, J. A., and A. S. Samek. 2015. "The Behavioralist as Nutritionist: Leveraging Behavioral Economics to Improve Child Food Choice and Consumption." *J. Health Econ*. 39:135–46.

Loyalka, P. K., S. Sylvia, C. Liu, J. Chu, and Y. Shi. 2016. "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement." *J. Labor Econ*. 37(3): 621–62.

Maniadis, Z., F. Tufano, and J. A. List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *A.E.R.* 104 (1): 277–90.

Marcotte, D. E., T. Bailey, C. Borkoski, and G. S. Kienzl. 2005. "The Returns of a Community College Education: Evidence from the National Education Longitudinal Survey." *Educ. Evaluation and Policy Analysis* 27 (2): 157–75.

McFarland, J., B. Hussar, C. de Brey, et al. 2017. "The Condition of Education 2017." NCES 2017-144. US Dept. Educ., Washington, DC.

Mountjoy, J. 2018. "Community Colleges and Upward Mobility." Working paper, https://sites.google.com/site/jackmountjoyeconomics/.

Muralidharan, K., and V. Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *J.P.E.* 119 (1): 39–77.

NCCBP (National Community College Benchmark Project). 2014. "National Community College Benchmark Project Report." Technical report, Ivy Tech.

Neal, D. 2011. "The Design of Performance Pay in Education." In *Handbook of the Economics of Education*, vol. 4, edited by E. Hanushek, S. Machin, and L. Woessmann, 495–550. Amsterdam: Elsevier.

Oreopoulos, P., and U. Petronijevic. 2013. "Making College Worth It: A Review of the Returns to Higher Education." *Future Children* 23 (1): 41–65.

Patel, R., L. Richburg-Hayes, E. de la Campa, and T. Rudd. 2013. "Performance-Based Scholarships: What Have We Learned? Interim Findings from the PBS Demonstration." Policy brief, MDRC, New York.

Ran, X., and D. Xu. 2017. "How and Why Do Adjunct Instructors Affect Students' Academic Outcomes? Evidence from Two-Year and Four-Year Colleges." Working paper, Center Analysis Postsecondary Educ. and Employment, New York.

Rogers, G. S. 2015. "Part-Time Faculty and Community College Student Success." *Community Coll. J. Res. and Practice* 39 (7): 673–84.

Sadoff, S. 2014. "The Role of Experimentation in Education Policy." *Oxford Rev. Econ. Policy* 30 (4): 597–620.

Scott-Clayton, J., and L. Schudde. 2016. "Performance Standards in Need-Based Student Aid." Working Paper no. 22713, NBER, Cambridge, MA.

Shapiro, D., A. Dundar, F. Huie, P. K. Wakhungu, X. Yuan, A. Nathan, and A. Bhimdiwali. 2017. "Completing College: A National View of Student Completion Rates—Fall 2011 Cohort (Signature Report no. 14)." Nat. Student Clearinghouse Res. Center, Herndon, VA.

Snyder, T. D., C. De Brey, and S. A. Dillow. 2018. "Digest of Education Statistics 2016." NCES 2017-094. Nat. Center Educ. Statis., Washington, DC.

Springer, M. G., D. Ballou, L. Hamilton, V.-N. Le, J. Lockwood, D. F. McCaffrey, M. Pepper, and B. M. Stecher. 2011. "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (Point)." Soc. Res. Educ. Effectiveness, Evanston, IL.

Springer, M. G., J. F. Pane, V.-N. Le, et al. 2012. "Team Pay for Performance: Experimental Evidence from the Round Rock Pilot Project on Team Incentives." *Educ. Evaluation and Policy Analysis* 34 (4): 367–90.

Todd, P. E., and K. I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Econ. J.* 113 (485):F3–F33.

Tversky, A., and D. Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Q.J.E.* 106 (4): 1039–61.

Zimmerman, S. D. 2014. "The Returns to College Admission for Academically Marginal Students." *J. Labor Econ.* 32 (4): 711–54.