Improving College Instruction through Incentives

Andy Brownback
University of Arkansas

Sally Sadoff*
UC San Diego

October 4, 2019

Abstract

Prior work demonstrates the importance of college instructor quality, but little is known about whether college instruction can be improved. In a field experiment, we examine the impact of performance-based incentives for community college instructors. We estimate that instructor incentives improve student exam scores by 0.16 - 0.2 standard deviations (SD), increase course grades by 0.1 SD, reduce course dropout by 17 percent, and increase credit accumulation by 18 percent. The effects are largest among part-time adjunct instructors. During the program, instructor incentives have large positive spillovers to students' unincentivized courses, significantly increasing completion rates and grades in courses outside our study. One year after the program ends, instructor incentives increase transfer rates to four-year colleges by an estimated 22 - 28 percent, with no impact on two-year college degrees. To test for potential complementarities, we examine the impact of instructor incentives in conjunction with student incentives and find no evidence that the incentives are more effective in combination. Finally, we elicit contract preferences for the loss-framed incentives we offer. At baseline, instructors prefer gain-framed incentives. However, after experiencing loss-framed incentives, instructors significantly increase their preferences for them.

JEL codes: I23, M52, M55, C93
Keywords: college instruction, incentives, loss-framing, contract preferences, field experiment, community college

*Corresponding Author: Sally Sadoff, UC San Diego, Rady School of Management, 9500 Gilman Drive, La Jolla, CA 92093; sadoff@ucsd.edu. Andy Brownback, University of Arkansas, Department of Economics, 220 N. McIlroy Ave, Fayetteville, AR 72701; abrownback@walton.uark.edu. 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. The project was made possible by the generous financial support of the Laura and John Arnold Foundation and the Yankelovich Center for Social Science Research. This research was conducted with approval from the University of Arkansas and UC San Diego Institutional Review Boards. AEA RCT Registry number: AEARCTR-0001411

1 Introduction

Over the last several decades, the 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 two-year community colleges, which serve about forty percent of all undergraduates (Shapiro, 2017). These schools provide students with a low-cost entry point to accumulate college credits towards both two-year and four-year degrees. But they struggle with poor student performance, which hinders credit accumulation, degree completion, and transfers to four-year schools (Snyder et al., 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. A one standard deviation increase in college instructor quality improves student performance by an estimated $0.05 - 0.30$ standard deviations (SD, hereafter), with effects generally smaller at selective universities and larger at non-selective institutions similar to community colleges (Hoffmann and Oreopoulos, 2009; Carrell and West, 2010; Braga et al., 2016; Brodaty and Gurgand, 2016; Bettinger et al., 2014; De Vlieger et al., 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 et al., 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 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 are generally independent adults, spend less time with their instructors, and can voluntarily withdraw from an

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

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

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. Our primary question is whether incentives for instructors can improve postsecondary student performance. Secondarily, 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 policymakers as a low-cost pathway to four-year schools, particularly for underrepresented and non-traditional students (Bailey et al., 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 over eighty percent enter with the intention of transferring to a four-year school, only about a third succeed in doing so (Jenkins and Fink, 2016).[4] Fewer than forty percent of community college students earn a college degree within six years (Shapiro, 2017) and a large share dropout 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, 2016). 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 four-year school. Accordingly, we examine the impact of our intervention both on performance in targeted courses, as well as on students' broader educational outcomes.[6]

[3]There is ongoing debate about the impact of this shift on student achievement (Ehrenberg and Zhang, 2005; Bettinger and Long, 2006; Hoffmann and Oreopoulos, 2009; Bettinger and Long, 2010; Figlio et al., 2015; Rogers, 2015; Ran and Xu, 2016).

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

[5]The top five public postsecondary institutions producing the highest ratio of dropouts with debt to graduates features four community colleges (including our partner institution). See J. Barshay, "3.9 Million Students Dropped Out of College with Debt in 2015 and 2016," *U.S. News*, (2017).

[6]Improved performance also helps students avoid debt. Course withdrawals provide neither credit

Our study included sixteen different departments with over 6,000 student-course observations in the Fall and Spring semesters of the 2016-2017 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 a 70 percent or higher on an objective, externally-designed course exam ("passed," hereafter). We framed the incentives as *losses* – i.e., bonuses that instructors would lose if they did not meet performance targets. To implement the loss framing, we gave instructors upfront 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 upfront bonus. If over 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 (instructor incentives and student incentives). 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 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, 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

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).

four-year schools.

In targeted courses, Instructor Incentives increase exam performance by an estimated 0.16 - 0.20 SD ($p < 0.01$) and increase exam pass rates by 19 percent ($p < 0.01$). The impact carries over to course performance, where grades improve by an estimated 0.10 - 0.11 SD ($p = 0.02$). The effects of incentives operate at both the extensive and intensive margins. Instructor Incentives reduce course dropout rates by 17 percent ($p = 0.03$); and also increase exam scores conditional on completion by 0.083 SD ($p = 0.04$). 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, $p < 0.01$) with smaller effects among full-time faculty (0.13 SD, $p = 0.12$). We find no evidence that the effects are driven by a narrow set of departments nor 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 percent ($p = 0.01$) with positive but not significant impacts on GPA. As noted above, credit accumulation and GPA are critical for students to qualify for transfer to four-year schools. And indeed we find that one year after the program ends, instructor incentives increase transfer rates to four-year colleges by 22 - 28 percent ($p < 0.01$), with no effect on two-year college degrees. Our findings demonstrate that instructor incentives have persistent and meaningful effects on students' 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.

[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.

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 nine percent 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 one standard deviation 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 over $40 per student per year or a ten-year net present value of over $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 twenty percent 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 could not only 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 2 discusses the related literature, Section 3 describes the experimental design, Section 4 presents the results for targeted courses, Section 5 presents the effects on broader educational outcomes, Section 6 examines instructors' contract preferences, and Section 7 concludes.

[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., one 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 from the Bureau of Labor Statistics (2017). We use a conservative discount rate of 10% to calculate a net present value over ten years of $263.72 ($254.82) for men (women).

2 Literature

Our first contribution is to a growing 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 non-experimental studies in the U.S. and experimental studies in developing countries have found that teacher incentives can improve performance, experimental studies in the U.S. have largely failed to demonstrate effectiveness.[9] We based the design of the incentives in our study on Fryer Jr et al. (2012, 2018), which is the only prior experimental study in the U.S. to find a positive impact of teacher incentives. The authors test upfront, loss-framed incentives among elementary and middle school teachers and estimate effects of 0.12 SD on math test scores pooling across two years of the experiment. Our findings 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 et al. (2012) find little evidence of complementarities between incentives for students, parents, and tutors in an experiment in U.S. elementary schools. In contrast, Behrman et al. (2015) find that incentives for teachers and students in 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

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

[10] Unlike Fryer Jr 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 two years of the experiment. Our incentives also differ. As discussed in Section 3.2, we base rewards on threshold achievement levels, while Fryer Jr et al. (2012, 2018) used the pay-for-percentile structure developed by Barlevy and Neal (2012).

[11] A large literature examines student incentives alone and generally finds small effects (see reviews by Sadoff, 2014; Fryer, 2017). In a community college context, Barrow et al. (2014) and Patel et al. (2013) 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 colleges students include Leuven et al. (2010), Angrist et al. (2009, 2014) and De Paola et al. (2012).

[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.

differences in the strategic response of instructors and students to each others' effort choices (Todd and Wolpin, 2003; De Fraja et al., 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 Jr et al. (2018) find that while upfront 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 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 et al., 2016; De Quidt, 2017; Jie, 2018).[15]

[13]See Brooks et al. (2012); Hossain and List (2012); Fryer Jr et al. (2012, 2018); Hong et al. (2015); Armantier and Boly (2015); Imas et al. (2016). In online studies, DellaVigna and Pope (2016) 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]We note that Fryer Jr et al. (2018) re-randomize 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.

[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 upfront 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

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]

3 Experimental design

3.1 Setting and recruitment

We conducted the experiment in the 2016-2017 school year at Ivy Tech Community College of Indiana (Ivy Tech). 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.

The East-Central and Richmond regions respectively serve communities in the 4th and 8th percentile of national median income. Over 60 percent of their student body is eligible for Pell Grants, placing them in the 90th percentile for community colleges. Their fall-to-fall retention rates of full-time students hover around 40 percent, just above the bottom 10 percent of community colleges. Overall, only 24 percent of their full-time, first-time students will graduate or transfer to a four-year institution within 3 years, also just above the bottom 10 percent of community colleges (NCCBP, 2014).

Our study includes a broad range of departments: Accounting, Anatomy and Physiology, Art History, Biology, Business Operations Applications and Technology, Business, Communications, Criminology, English, Health Sciences, Math, Nursing, Psychology, Physical Science, Software Development, and Sociology. 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

should be irrelevant for both effort and preference. Thus, our "standard" behavioral model refers to one assuming the status quo as the reference point. See Imas et al. (2016) for discussion of the theory.

[16]These findings are in line with Kermer et al. (2006), which argues 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 gain-framed incentives.

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 non-permanent 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 15th, 2016 and ended September 6th, 2016 with a final total of 108 enrolled instructors, 90 percent of our recruitment goal of 120 and 72 percent of all eligible instructors. The randomization (detailed in Section 3.4) was based on the students enrolled in a given course as of the Ivy Tech census date, September 2nd, 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 two weeks of class.[17]

Fall instructors teaching eligible courses in the Spring 2017 semester were automatically re-enrolled. 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 Section 3.4, there were no differences in eligibility by treatment group). We also recruited new instructors. The 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 20th, 2017 and ended February 3rd, 2017. An additional 26 instructors

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

signed up, bringing the spring semester total to 99 participating instructors. Including continuing instructors from the fall, 66 percent 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 30th, 2017. Over both semesters, 134 instructors participated in the study, 93 percent of our recruitment goal of 144.

3.2 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 scored a 70 percent or higher on the objective course exam ("passed," hereafter).[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 (see Neal, 2011, for discussion).

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 upfront bonuses equivalent to the amount instructors would earn if 50 percent of their students passed the exam. For example, an instructor who taught one section with 20 students would receive an upfront check for $500. At the end of the semester, if fewer than 50 percent of the students passed the exam, the instructor was responsible for returning the difference between the final bonus amount and the upfront bonus. If more than

[18] The health sciences and nursing courses had thresholds for passing that exceeded 70 percent. In these cases, we considered "passing" to be the pre-existing requirement of 75 percent.

50 percent 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 upfront 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 upfront bonus payment was issued once the forms were signed.

The average upfront 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 upfront bonuses – 98 percent of instructors in the fall and 94 percent of instructors in the spring complied with the upfront payments.[19] Compliance with repayments varied across the two semesters – in the fall, 93 percent of instructors who owed money complied with repayment (96 percent of money owed), and in the spring 78 percent of instructors who owed money complied with repayment (83 percent 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 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

[19] The remainder did not fill out the paperwork to receive payments (3 instructors) or did not cash the upfront payment check (1 instructor).

were also attractive from a cost perspective. A summer scholarship had a face value of $400 but had an expected marginal cost of only about $97, given realized pass rates of 44.7 percent and take up rates of 54.4 percent in sections offering student incentives. Given summer enrollment rates of 26.8 percent 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 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 (Appendix Figure A.1). Students enrolled in the program by signing up in their class or through an online survey. Of the 1035 students offered incentives, 772 (74.6 percent) actively consented to participate and 48 (4.6 percent) 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.

3.3 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 mid-year 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 percent for the mid-year survey (96 percent in Instructor Incentives and 77 percent in Instructor Control). Response rates were lower for the year-end survey: 67 percent (83 percent in Instructor Incentives and 49 percent in Instructor Control).

In the enrollment and mid-year surveys, we elicited instructors' preferences for loss-framed relative to gain-framed contracts (see Appendix B for preference elicitation

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

questions). 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 upfront – 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 thirteen 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 towards teaching. In the mid-year 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 upfront payments and their expectations about their final reward (e.g., whether they expected to receive additional rewards or owe money back).

3.4 Randomization

We first describe the randomization of instructors to receive incentives in the fall and spring semesters (See Appendix Figure A.2 for a summary). We then describe the randomization of individual sections to receive incentives for students in the spring

[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.

semester. The randomization and analysis follows our pre-analysis 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, the percentage of courses with a corresponding remedial section (Co-Req), 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 Grade Point Average (GPA). For each characteristic, we specified that we would re-randomize in the event that differences were significant with a p-value < 0.15.[24]

In Spring 2017, we conducted the randomization in two stages. First, we determined if 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 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 opted out of the spring semester of the study; and (2) two instructors assigned to Instructor Control in the fall 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 in the spring. Accordingly, we randomly re-assigned two instructors. Therefore, we have 38 continuing instructors as-

[22]We pre-registered our analysis plan. See https://osf.io/fbxpw/. We note below deviations from the pre-analysis 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 re-randomized if any coefficients were significant with $p < 0.15$. Bruhn and McKenzie (2009) examine various methods of randomization including stratification and re-randomization, 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 Section 4, including only the sparse set of covariates we stratified on does not affect the results.

signed 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 $p < 0.05$ 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 as Instructor Incentives only sections. 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 Incentive 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.

3.5 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 1 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 co-requisite 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 or control, β_1, β_2, and β_3

[25]Our pre-analysis 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 (MDES) 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 Table 2 and in Appendix Tables 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 2 fall and 11 spring instructors, and time preference values for 2 fall and 10 spring instructors.

measure the full effects (rather than marginal effects) of 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 pre-analysis plan; and of just under 0.22 SD for Combined Incentives, compared to a MDES of 0.2 SD in our pre-analysis 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 a MDES of 0.25 SD. We did not have a large enough sample size to adequately power a test of Student Incentives alone, β_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 below), 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 3.4. 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.

3.6 Data and baseline characteristics

We collected course data for 6,241 student-course observations in 383 sections. Our administrative dataset 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 (Appendix Table A.2). Nonetheless,

[27] Appendix 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 department 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.

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 for the following student-level characteristics: 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); instructor-level characteristics: full-time or adjunct, total sections in the study, students per section, and elicited contract and time preferences; and section-level characteristics: whether the course section is a co-requisite (the Ivy Tech co-requisite course model is a form of remedial education for under-prepared 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 as 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 forty-eight 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.

4 Outcomes in targeted courses

4.1 Exam performance

We first examine the effect of treatment on the directly incentivized outcome: performance on the objective course exam. Figures 1a and 1b show the distribution of test scores in the control group and in the Instructor Incentives group. The vertical line at a score of 70 percent 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 student persistence in the course (i.e., non-zero exam scores) and

[28] We note that our setting differs from most K-12 contexts in which all students are required to take the exam and so non-exam takers are considered attriters. In that context, missing scores should not necessarily be coded as having a zero score on the exam. 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.

also increase exam scores among those who take the exam. Throughout the distribution, Instructor Incentives shift scores to the right. Figure 1a shows that, while there is some evidence of missing mass just below the 70 percent 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 percent or higher. As shown in Figure 1b, the Instructor Incentives distribution stochastically dominates the Control distribution. A Kolmogorov-Smirnov test clearly rejects that the scores have equal distributions ($p < 0.01$).

In the regression analysis below, we examine tests 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 Ordinary Least Squares (OLS) estimation in column 1 - 2 and then using a random effects model in columns 3 - 6. We estimate treatment effects for the full year (columns 1-4) and by semester (columns 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 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 co-requisite.[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

[29] Appendix Table A.3 columns 3 - 4 report the OLS estimates by semester. The results do not change.

[30] Our analysis differs from our pre-analysis plan in two ways. First, our pre-analysis 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 pre-analysis plan did not include co-requisite classification as a course-level covariate because we were not aware of this classification at the time. Columns 6 and 7 of Appendix Table A.3 repeat the analysis including GPA in two different ways. In column 6, we impute missing GPAs as the mean GPA and include an indicator for whether GPA is missing. In column 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 Appendix Table A.3 repeats our analysis excluding the covariate for co-requisite courses. No results are affected.

[31] As discussed in Section 3.5, 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.

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 over 6 percentage points off a control group mean of 52% ($p < 0.01$ 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 ($p = 0.53$). The pattern of effects holds if we restrict the spring sample to the subset of returning instructors who also received incentives in the fall (Appendix Table A.3 column 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 (Appendix Table A.4), we also confirm the broad distributional impact of Instructor Incentives shown in Figures 1a and 1b. 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, as well as 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 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 3.4, the Student Incentives treatment is under-powered, and so throughout our analysis we interpret the estimates with caution (see Maniadis et al. (2014) for a discussion of properly interpreting results from small samples).

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

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

[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 Appendix Table A.5 columns 1 - 2, the estimated effects of pooled instructor incentives are large and statistically significant, 0.19 SD ($p < 0.01$), where column 1 pools Student Incentives with Control and column 2 includes a separate indicator for Student Incentives. We estimate an interaction specification in columns 3 - 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 ($p = 0.23$) and -0.15 SD ($p = 0.18$), which is economically meaningful but not statistically significant.

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 ($p < 0.01$), a 19 percent increase compared to the 40 percent 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 ($p < 0.01$). They also increase scores at the intensive margin, with exam scores improving an estimated 0.08 SD ($p = 0.04$) 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 ($p = 0.02$); as well as on scores among those who take the exam, which increase by an estimated 0.14 SD ($p < 0.01$). However, there is no effect on the extensive margin of taking the exam. Student Incentives have little impact at either margin.[34]

4.2 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 (column 1) and then in standardized units (column 2). We use the standard 0-4 scale of grade points corresponding to A-F grades with withdrawals counting as 0 grade points. We normalize grades within each department to have a mean of zero and a standard deviation of

[34]We present the analogous results using OLS estimation in Appendix Table A.6 columns 1- 3. No results are affected.

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 – i.e., due to the increases in the exam scores on which teachers were incentivized. We calculate the non-exam grade by subtracting the exam score weighted by the percentage of the course grade that the syllabus attributes to the incentivized test (between 5-25% across courses). We then re-weight 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 ($p = 0.02$) or 0.16 grade points ($p = 0.02$) off a control group mean of 2.08. Examining the effect on the non-exam course grade, we estimate that Instructor Incentives generate a 0.10 SD increase in course performance that is not attributable to the incentivized exam ($p = 0.02$). 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 ($p = 0.03$). This represents a 17 percent reduction in the baseline dropout rate of about 22 percent. As previously noted, 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]

4.3 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

[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 an 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.

[36] We present the analogous results using OLS estimation in Appendix Table A.6 columns 4 - 6; and, in Appendix Table A.7, using random effects estimation with the sparse set of controls included in Table 2 column 1. The estimated effects are similar but generally estimated with less precision.

for differential effects by instructor type – 48 percent of our students (as well as 48 percent of sections) are instructed by full-time faculty while 52 percent 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 ($p < 0.01$) and 0.19 SD ($p < 0.01$), respectively. Instructor Incentives also increase exam pass rates and course completion rates among adjunct faculty by an estimated 10 percentage points ($p < 0.01$) and 7 percentage points ($p < 0.01$), respectively.[37] For full-time faculty, the estimated effect 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), 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.

[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 ($p < 0.01$), exam pass rates ($p = 0.05$), course grades ($p = 0.02$), and course completion ($p = 0.04$). The marginally significant impact of Combined Incentives on exam pass rates among full-time faculty does not survive the correction ($p = 0.24$).

[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, instructor-level, or course-level covariates.

In exploratory analysis (Appendix Table A.8), we consider additional dimensions of instructor heterogeneity that may influence treatment effectiveness. We find no significant heterogeneity by gender, degree type, self-reported beliefs about how much students can improve, amount of the upfront bonus the instructor spent, online vs 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 incentives can be broadly effective.

4.4 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 (Appendix 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" (Appendix Table A.11).

5 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

[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. Appendix Table A.9 repeats the analysis with Lee (2009) bounds to correct for non-response on the survey.

examine the impact of incentives on students' broader educational outcomes. We first 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 towards 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 towards a degree, and qualify for transfer to four-year schools. We then investigate students' longer-term outcomes directly. One year after the program ends, we estimate treatment effects on transfers to four-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]

5.1 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 change 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 Instructor Incentives and Combined Incentives into "treatment" (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 on 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"

[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; De Vlieger et al., 2017; Ran and Xu, 2016).

(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 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 controlling 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") which includes 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 points) for the courses students were taking outside of our study. In columns 5 - 6 and columns 7 - 8, we estimate effects on students' credit accumulation and cumulative GPA, respectively. These measures include all courses taken prior to and during the treatment semester. We report marginal effects from a probit regression in columns 1 - 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 ($p < 0.01$). 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 ($p < 0.01$) 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 non-program 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 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 towards degrees and qualify for transfer to four-year schools. We estimate that students accumulate 0.44 additional credits ($p = 0.01$) per incentivized course. This represents a 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).

[41] Similarly, in Appendix 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.

5.2 Post-program outcomes

Finally, we examine students' longer-run outcomes one year after the program ends. These include transfer to a four-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 four-year schools. For every additional treated course, the transfer rate increases by 2.8 percentage points ($p < 0.01$). The estimated impact of having at least one treated course is a 2.2 percentage point increase in transfer rates. This represents a $22 - 28$ percent 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, nor on 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 – i.e., 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 and GPA needed to qualify for transfer to four-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 four-year degrees.

6 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.

6.1 Baseline preferences

When instructors enrolled in our study (either before the Fall semester or before the Spring semester), we used an incentive-compatible multiple price list mechanism described in Section 3.3 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 endline 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 – i.e., 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 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 experiments (e.g. Tversky and Kahneman, 1991; Abdellaoui et al., 2008).[44]

[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 anyone who switched in the interior. We dropped anyone who switched multiple times in the list. This drops 1 instructor in the Fall baseline and 3 in the Fall endline.

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

[44] We calculate the loss aversion parameter, λ, such that the lower expected incentive payment under

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, 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 $-\$.50$, 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).

6.2 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 endline measured on the enrollment and mid-year 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 towards preferring loss contracts: 38.0% of treatment instructors show increased preference towards 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 weak preference for gain contracts, making it the most

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 $x = \Pr[Pass](\$50N)$ is the expected incentive payment for an instructor with N students and $r = 0.5(\$50N)$ is the reference point (i.e., the upfront 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 $\Pr[Pass] = 0.478$, 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 endline survey could not be incentive-compatible.

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 – i.e., 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 endline preferences). The outcome variable in columns 2-4 is contract preference on the Fall endline survey. Columns 1-2 only control 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 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 over \$6 *more* per student in order to prefer the loss contract ($p < 0.01$ for both the treatment and control group).[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 over \$6 ($p < 0.01$). In contrast, assignment to Instructor Incentives significantly increases instructor preferences for loss contracts. The treatment effects of \$4.33 - \$4.53 ($p < 0.05$ in all specifications) largely erase preferences for gain contracts. In the endline survey, the treatment group's value for the loss contract is no longer statistically distinguishable from zero ($p > 0.20$ in all specifications). That is, after experiencing loss incentives, instructors become (close to) indifferent between the two contract types.[49]

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

[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.

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

[49]As shown in Appendix Table A.13, the effects on contract preferences are robust to corrections for differential attrition across treatment and control instructors in the endline survey. Columns 1-2 estimate attrition. The estimated impact of incentives is similar if we assume missing endline preferences are the same as at baseline (column 3) or if we estimate a Lee (2009) upper bound assuming that aversion to loss contracts increases the likelihood of attrition (column 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 (column 5), but this direction of differential attrition seems less likely. As shown in Appendix Table A.14, the

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 – i.e., 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 provides support for the sustainability of their use as a policy tool.

7 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 one standard deviation. These impacts extend beyond directly incentivized course performance to non-exam grades, courses outside our study, credit accumulation, and post-treatment transfers to four-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 fifty and eighty percent of four-year and two-year college instructors, respectively (Hurlburt and McGarrah, 2016).

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

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) (columns 1-2); and is also unrelated to whether the instructor gained or lost money under the contract (column 3).

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., Bleichrodt, H., and L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of 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. *Journal of the American statistical Association*, 103(484):1481–1495.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.

Angrist, J., Oreopoulos, P., and Williams, T. (2014). When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources*, 49(3):572–610.

Armantier, O. and Boly, A. (2015). Framing of incentives and effort provision. *International Economic Review*, 56(3):917–938.

Bailey, T. R., Jaggars, S. S., and Jenkins, D. (2015). *Redesigning America's community colleges*. Harvard University Press.

Barlevy, G. and Neal, D. (2012). Pay for percentile. *The American Economic Review*, 102(5):1805–1831.

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

Barrow, L., Richburg-Hayes, L., Rouse, C. E., and Brock, T. (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics*, 32(3):563–599.

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

Bettinger, E., Fox, L., Loeb, S., and Taylor, E. (2014). Changing distributions: How online college classes alter student and professor performance. Technical report, Working Paper, Stanford University.

Bettinger, E. P. and Long, B. T. (2006). The increasing use of adjunct instructors at public institutions: Are we hurting students. *What's happening to public higher education*, pages 51–69.

Bettinger, E. P. and Long, B. T. (2010). Does cheaper mean better? the impact of using adjunct instructors on student outcomes. *The Review of Economics and Statistics*, 92(3):598–613.

Braga, M., Paccagnella, M., and Pellizzari, M. (2016). The impact of college teaching on students' academic and labor market outcomes. *Journal of Labor Economics*, 34(3):781–822.

Brickley, J. A. and Zimmerman, J. L. (2001). Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate finance*, 7(4):367–396.

Brodaty, T. and Gurgand, M. (2016). Good peers or good teachers? evidence from a french university. *Economics of Education Review*, 54:62–78.

Brooks, R. R., Stremitzer, A., and Tontrup, S. (2012). Framing contracts: Why loss framing increases effort. *Journal of Institutional and Theoretical Economics JITE*, 168(1):62–82.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232.

Carrell, S. E. and West, J. E. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.

Courant, P. N. and Turner, S. (2017). Faculty deployment in research universities. Technical report, National Bureau of Economic Research.

Czibor, E., Hsu, D., Neckermann, S., and Subasi, B. (2019). Loss framed incentives and employee (mis-)behavior. Technical report.

De Fraja, G., Oliveira, T., and Zanchi, L. (2010). Must try harder: Evaluating the role of effort in educational attainment. *The Review of Economics and Statistics*, 92(3):577–597.

De Paola, M., Scoppa, V., and Nisticò, R. (2012). Monetary incentives and student achievement in a depressed labor market: results from a randomized experiment. *Journal of Human Capital*, 6(1):56–85.

De Philippis, M. (2015). Multitask agents and incentives: the case of teaching and research for university professors.

De Quidt, J. (2017). Your loss is my gain: a recruitment experiment with framed incentives. *Journal of the European Economic Association*.

De Quidt, J., Fallucchi, F., Kölle, F., Nosenzo, D., and Quercia, S. (2017). Bonus versus penalty: How robust are the effects of contract framing? *Journal of the Economic Science Association*, 3(2):174–182.

De Vlieger, P., Jacob, B., and Stange, K. (2017). Measuring instructor effectiveness in higher education. In *Productivity in Higher Education*. University of Chicago Press.

DellaVigna, S. and Pope, D. (2016). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*.

Deming, D. and Dynarski, S. (2009). Into college, out of poverty? policies to increase the postsecondary attainment of the poor. Technical report, National Bureau of Economic Research.

Denning, J. T. (2017). College on the cheap: Consequences of community college tuition reductions. *American Economic Journal: Economic Policy*, 9(2):155–88.

Duflo, E., Hanna, R., and Rya, S. P. (2012). Incentives work: Getting teachers to come to school. *The American Economic Review*, 102(4):1241–1278.

Ehrenberg, R. G. (2012). American higher education in transition. *The Journal of Economic Perspectives*, 26(1):193–216.

Ehrenberg, R. G. and Zhang, L. (2005). Do tenured and tenure-track faculty matter? *Journal of Human Resources*, 40(3):647–659.

Englmaier, F., Grimm, S., Schindler, D., and Schudy, S. (2018). Effect of incentives in non-routine analytical team tasks – evidence from a field experiment.

Evans, W. N., Kearney, M. S., Perry, B. C., and Sullivan, J. X. (2017). Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case management intervention. Technical report, National Bureau of Economic Research.

Figlio, D. N., Schapiro, M. O., and Soter, K. B. (2015). Are tenure track professors better teachers? *Review of Economics and Statistics*, 97(4):715–724.

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

Fryer, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experimentsa. *Handbook of Economic Field Experiments*, 2:95–322.

Fryer Jr, R. G., Levitt, S. D., List, J., and Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. Technical report, National Bureau of Economic Research.

Fryer Jr, R. G., Levitt, S. D., List, J., and Sadoff, S. (2018). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. Technical report, Working paper.

Geng, T. (2017). The complementarity of incentive policies in education: Evidence from new york city.

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Glazerman, S., McKie, A., and Carey, N. (2009). An evaluation of the teacher advancement program (tap) in chicago: Year one impact report. final report. *Mathematica Policy Research, Inc.*

Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–227.

Goodman, J., Hurwitz, M., and Smith, J. (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35(3):829–867.

Grubb, W. N. (2002). Learning and earning in the middle, part i: National studies of pre-baccalaureate education. *Economics of Education Review*, 21(4):299–321.

Hoffmann, F. and Oreopoulos, P. (2009). Professor qualities and student achievement. *The Review of Economics and Statistics*, 91(1):83–92.

Hong, F., Hossain, T., and List, J. A. (2015). Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization*, 118:372–382.

Horrace, W. C. and Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327.

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

Hurlburt, S. and McGarrah, M. (2016). The shifting academic workforce: Where are the contingent faculty. *Delta Cost Project at American Institutes for Research. Retrieved from http://www. air. org/sites/default/files/downloads/report/Shifting-Academic-Workforce-November-2016. pdf.*

Imas, A., Sadoff, S., and Samek, A. (2016). Do people anticipate loss aversion? *Management Science*, 63(5):1271–1284.

Jenkins, P. D. and Fink, J. (2016). Tracking transfer: New measures of institutional and state effectiveness in helping community college students attain bachelor's degrees.

Jie, Y. (2018). Prepayment effect: Prepayment with clawback increases task participation. *Journal of Business Research*, 92:210–218.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, pages 263–291.

Kane, T. J. and Rouse, C. E. (1995). Labor market returns to two-and four-year colleges. *American Economic Review*, 85:600–614.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., and Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological science*, 17(8):649–653.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lavecchia, A. M., Liu, H., and Oreopoulos, P. (2014). Behavioral economics of education: Progress and possibilities. Technical report, National Bureau of Economic Research.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Leuven, E., Oosterbeek, H., and Van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4):183–219.

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, 118(1):41–71.

List, J. A. (2004). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica*, 72(2):615–625.

List, J. A. (2011). Does market experience eliminate market anomalies? the case of exogenous market experience. *The American Economic Review*, 101(3):313–317.

List, J. A., Livingston, J. A., and Neckermann, S. (2012). Harnessing complementarities in the education production function. Technical report, Working Paper, available at https://econresearch. uchicago. edu/sites/econresearch. uchicago. edu/files/Livingston_AWFE% 20version. pdf.

List, J. A. and Samek, A. S. (2015). The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption. *Journal of health economics*, 39:135–146.

Loyalka, P. K., Sylvia, S., Liu, C., Chu, J., and Shi, Y. (2016). Pay by design: Teacher performance pay design and the distribution of student achievement.

Maniadis, Z., Tufano, F., and List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1):277–90.

Marcotte, D. E., Bailey, T., Borkoski, C., and Kienzl, G. S. (2005). The returns of a community college education: Evidence from the national education longitudinal survey. *Educational Evaluation and Policy Analysis*, 27(2):157–175.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., et al. (2017). The condition of education 2017. nces 2017-144. *National Center for Education Statistics*.

Mountjoy, J. (2018). Community colleges and upward mobility.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of political Economy*, 119(1):39–77.

NCCBP (2014). *National Community College Benchmark Project Report*.

Neal, D. (2011). The design of performance pay in education. In *Handbook of the Economics of Education*, volume 4, pages 495–550. Elsevier.

Oreopoulos, P. and Petronijevic, U. (2013). Making college worth it: A review of the returns to higher education. *The Future of Children*, 23(1):41–65.

Patel, R., Richburg-Hayes, L., de la Campa, E., and Rudd, T. (2013). Performance-based scholarships: What have we learned? interim findings from the pbs demonstration.

Ran, X. and Xu, D. (2016). How and why do adjunct instructors affect students' academic outcomes? evidence from two-year and four-year colleges.

Rogers, G. S. (2015). Part-time faculty and community college student success. *Community College Journal of Research and Practice*, 39(7):673–684.

Sadoff, S. (2014). The role of experimentation in education policy. *Oxford Review of Economic Policy*, 30(4):597–620.

Scott-Clayton, J. and Schudde, L. (2016). Performance standards in need-based student aid. Technical report, National Bureau of Economic Research.

Shapiro, D., D. A. H. F. W. P. Y. X. N. A. . H. Y. A. (2017). Completing college: A national view of student attainment rates (signature report no. 12b). *National Student Clearinghouse*.

Snyder, T. D., De Brey, C., and Dillow, S. A. (2018). Digest of education statistics 2016, nces 2017-094. *National Center for Education Statistics*.

Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J., McCaffrey, D. F., Pepper, M., and Stecher, B. M. (2011). Teacher pay for performance: Experimental evidence from the project on incentives in teaching (point). *Society for Research on Educational Effectiveness*.

Springer, M. G., Pane, J. F., Le, V.-N., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., and Stecher, B. (2012). Team pay for performance: Experimental evidence from the round rock pilot project on team incentives. *Educational Evaluation and Policy Analysis*, 34(4):367–390.

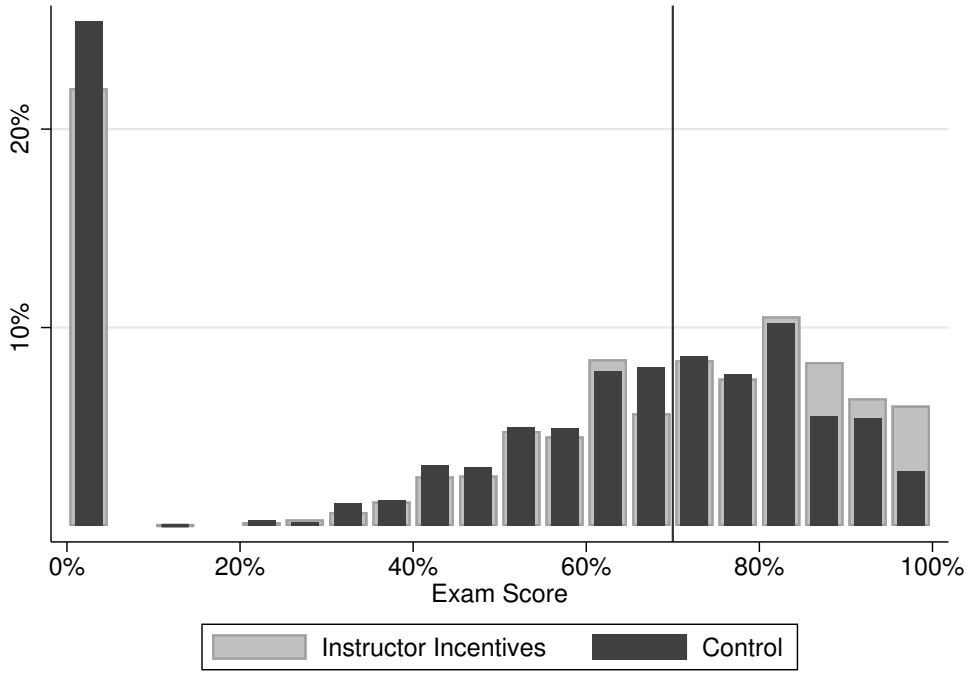
Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485).

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 106(4):1039–1061.

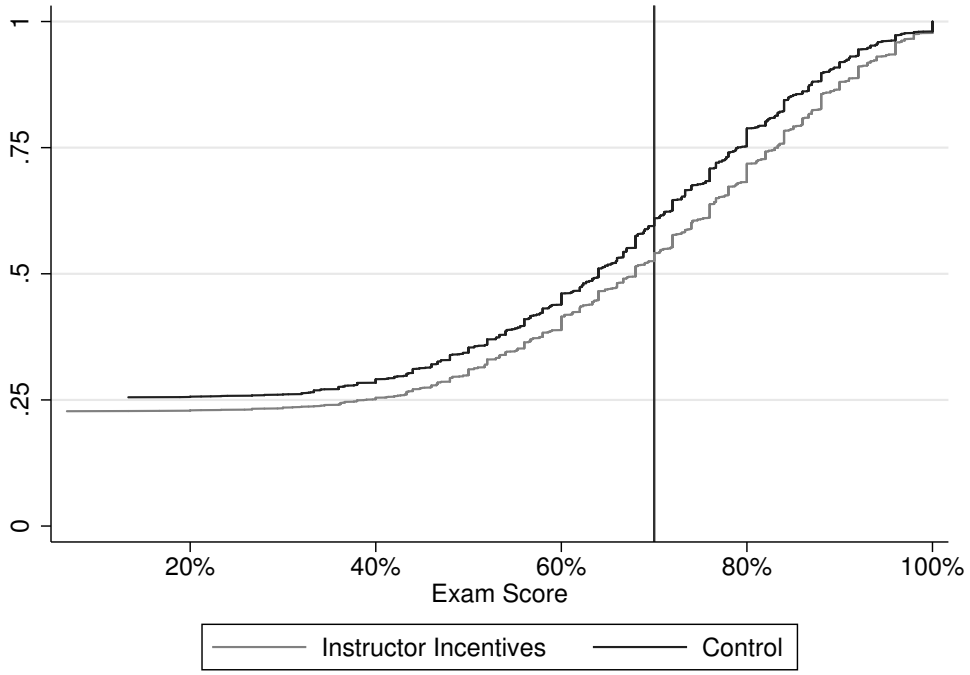
Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

Figure 1: Distribution of Exam Scores in Instructor Incentives and Control Groups

(a) Histogram

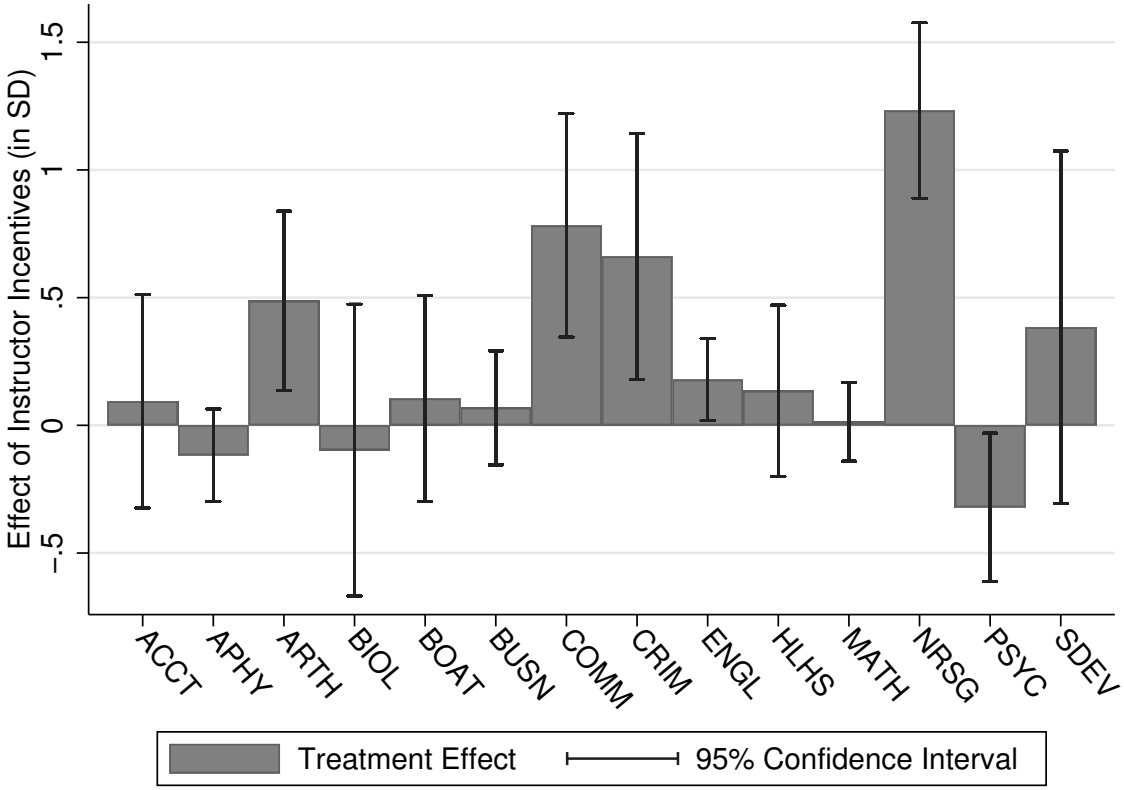


(b) Cumulative Distribution Function



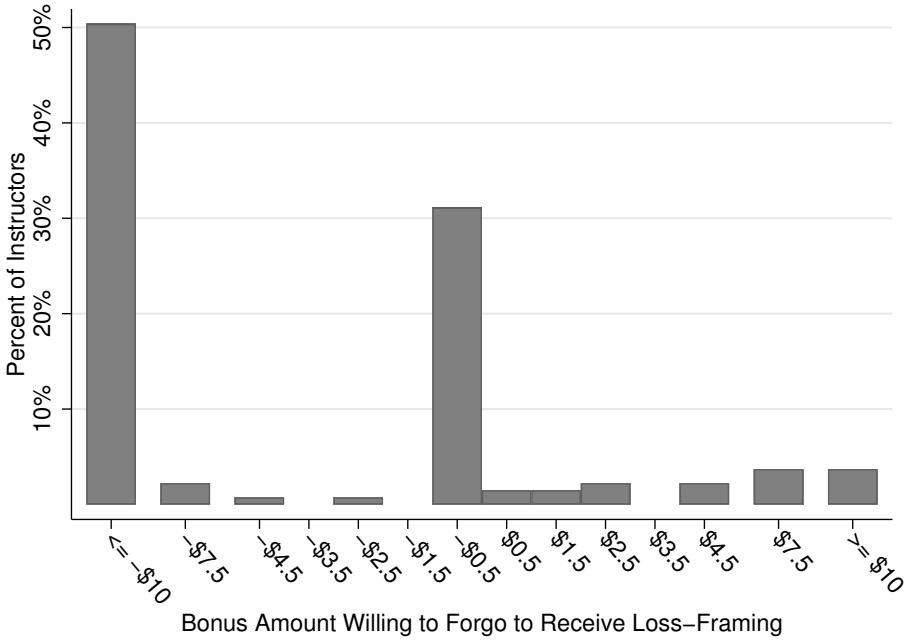
Notes: The figure presents (a) a histogram in 5 percentage point intervals and (b) a cumulative distribution function of exam scores for all students in courses taught by instructors in the Instructor Incentives or Control treatments. The vertical line at 70% represents the cutoff for "passing" the exam.

Figure 2: Effects of Instructor Incentives on Normalized Exam Scores by Department



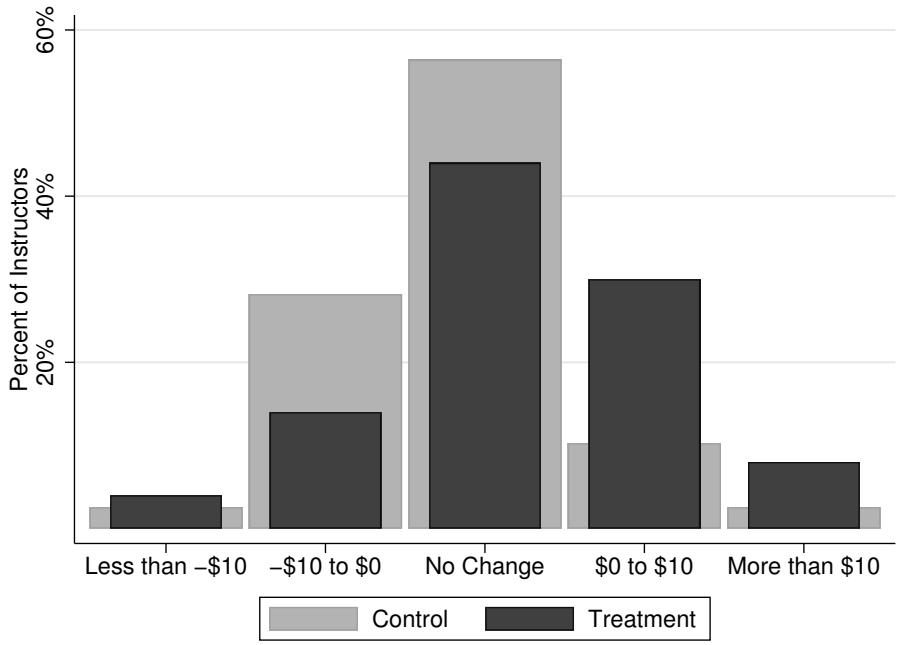
Notes: The figure presents coefficients and 95% confidence intervals for Instructor Incentives from random effects linear estimation with standard errors clustered by instructor within the following departments: Accounting (ACCT), Anatomy and Physiology (APHY), Art History (ARTH), Biology (BIOL), Business 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, baseline credits).

Figure 3: Baseline Contract Preferences



Notes: The figure presents the distribution of baseline contract preferences for all instructors. < -$10 indicates instructors who preferred a gain-framed bonus of $40 per student over a loss-framed bonus of $50 per student. > $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 are the mid-point between the per-student bonus amounts over which the instructor "switches" from preferring the gain-framed bonus to preferring the loss-framed bonus.

Figure 4: Change in Contract Preferences by Treatment



Notes: The figure presents the distribution of instructors' changes in willingness to pay for loss-framed contracts between the fall baseline and fall endline surveys. Positive values indicate increased willingness to pay for loss-framed contracts.

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	**0.343**	25.228	24.758	24.787	24.659	**0.839**
	(0.423)	(0.516)		(0.473)	(0.545)	(0.613)	(0.642)	
Male	0.342	0.340	**0.926**	0.326	0.274**	0.322	0.325	**0.135**
	(0.028)	(0.023)		(0.034)	(0.029)	(0.026)	(0.044)	
White	0.817	0.818	**0.805**	0.817	0.831	0.830	0.859*	**0.393**
	(0.011)	(0.015)		(0.015)	(0.015)	(0.018)	(0.017)	
Baseline Credits	12.403	13.018	**0.717**	18.011	19.442	19.236	17.580	**0.660**
	(1.012)	(1.665)		(1.442)	(2.525)	(1.315)	(1.986)	
Baseline GPA	2.843	2.804	**0.298**	2.934	2.925	2.950	2.790	**0.327**
	(0.037)	(0.050)		(0.048)	(0.068)	(0.056)	(0.066)	
Missing GPA	0.372	0.371	**0.487**	0.498	0.446	0.416	0.514	**0.395**
	(0.030)	(0.030)		(0.028)	(0.039)	(0.035)	(0.041)	
Instructor-Level Characteristics								
Full-Time	0.377	0.345	**0.733**	0.442	0.514	0.378	0.368	**0.284**
	(0.067)	(0.065)		(0.077)	(0.083)	(0.081)	(0.114)	
Total Sections	1.981	1.727	**0.273**	2.093	2.000	2.000	2.211	**0.898**
	(0.190)	(0.131)		(0.182)	(0.182)	(0.182)	(0.224)	
Students Per Section	16.467	16.342	**0.906**	14.200	15.893	15.164	14.421	**0.520**
	(0.749)	(0.750)		(0.709)	(0.866)	(0.697)	(1.042)	
Below-Median	0.509	0.455	**0.572**	0.558	0.378	0.459	0.684	**0.151**
Contract Value	(0.069)	(0.068)		(0.077)	(0.081)	(0.083)	(0.110)	
Below-Median	0.642	0.691	**0.591**	0.651	0.757	0.676	0.632	**0.523**
Discount Rate	(0.067)	(0.063)		(0.074)	(0.072)	(0.078)	(0.114)	
Section-Level Characteristics								
Co-Requisite	0.276	0.263	**0.905**	0.056	0.068	0.044	0.000	**0.542**
	(0.079)	(0.075)		(0.055)	(0.065)	(0.044)	(0.000)	
Observations	1776	1592		1043	657	687	311	

Table reports means/proportions for each group with standard errors in parentheses, are clustered at the level of randomization (instructor).
Student-level: we conduct a random effects regression. F-tests report the p-value of a joint test that all treatment coefficients are zero.
Instructor-level and section-level: p-values are reported for joint orthogonality test across treatment groups.
Significance tests against control mean: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Effects of Incentives on Exam Scores

	Pooled Fall and Spring Semesters				Semester	
	OLS		Random Effects		Fall	Spring
Instructor Incentives	0.172	0.161	0.202	0.204	0.113	0.247
	(0.063)	(0.060)	(0.057)	(0.056)	(0.057)	(0.085)
	$p = 0.004$	$p = 0.006$	$p < 0.001$	$p < 0.001$	$p = 0.043$	$p = 0.004$
Combined Incentives	0.111	0.099	0.162	0.156		0.174
	(0.087)	(0.083)	(0.078)	(0.075)		(0.085)
	$p = 0.086$	$p = 0.107$	$p = 0.018$	$p = 0.019$		$p = 0.026$
Student Incentives	-0.042	-0.052	0.070	0.066		0.067
	(0.089)	(0.092)	(0.074)	(0.076)		(0.083)
	$p = 0.659$	$p = 0.687$	$p = 0.198$	$p = 0.207$		$p = 0.233$
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)	0.384	0.308	0.573	0.485		0.310
Instructors	127	127	127	127	100	96
Observations	5839	5839	5839	5839	3189	2650

Columns 1 & 2: OLS estimation.

Columns 3-6: Random effects linear estimation.

Standard errors in parentheses clustered at the instructor level.

Dependent variable: exam score standardized within dept. (mean 0, s.d. 1) except in column 4.

Columns 1-4 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 (co-requisite).

Exact p-values calculated based on randomization inference.

Table 3: Effects of Incentives on Extensive and Intensive Margins

	Pass Exam		Take Exam		Score if Take	
Instructor Incentives	0.074	0.078	0.053	0.053	0.079	0.083
	(0.030)	(0.029)	(0.019)	(0.018)	(0.043)	(0.040)
	$p = 0.010$	$p = 0.006$	$p = 0.003$	$p = 0.002$	$p = 0.053$	$p = 0.043$
Combined Incentives	0.080	0.082	0.016	0.013	0.136	0.136
	(0.036)	(0.034)	(0.030)	(0.029)	(0.047)	(0.045)
	$p = 0.020$	$p = 0.018$	$p = 0.246$	$p = 0.296$	$p = 0.004$	$p = 0.005$
Student Incentives	0.038	0.037	0.014	0.013	0.053	0.053
	(0.046)	(0.048)	(0.027)	(0.028)	(0.047)	(0.047)
	$p = 0.203$	$p = 0.208$	$p = 0.329$	$p = 0.334$	$p = 0.149$	$p = 0.144$
Control Group Mean	0.400		0.746		0.450	
	(0.49)		(0.44)		(0.51)	
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	5839	5839	5741	5741	4421	4421

Columns 1-4: Marginal effects from random effects probit estimation. Robust standard errors
Columns 5 & 6: Random effects linear estimation. Standard errors clustered at the instructor level.
Standard deviation reported for control group mean.
Sample size is smaller for columns 3 & 4 because one department perfectly predicts taking the exam.
Dependent variable for columns 5 & 6: exam score standardized within dept. (mean 0, SD 1).
All analysis includes semester and department fixed effects.
Columns 2, 4, & 6 add student covariates (age, gender, race/ethnicity, baseline credits),
instructor covariates (type, contract value, discount rate) and course covariates (co-requisite).
Exact p-values calculated based on randomization inference.

Table 4: Effects of Incentives on Course Performance

	Course Grade			Complete	Grade if
	Grade Points	Standardized	Excluding Exam	Course	Complete
Instructor Incentives	0.165	0.108	0.097	0.037	0.037
	(0.080)	(0.052)	(0.049)	(0.019)	(0.051)
	$p = 0.015$	$p = 0.016$	$p = 0.024$	$p = 0.028$	$p = 0.194$
Combined Incentives	0.113	0.071	0.041	-0.010	0.094
	(0.113)	(0.072)	(0.069)	(0.029)	(0.063)
	$p = 0.114$	$p = 0.129$	$p = 0.241$	$p = 0.657$	$p = 0.060$
Student Incentives	-0.100	-0.058	-0.068	-0.002	-0.068
	(0.078)	(0.051)	(0.053)	(0.025)	(0.048)
	$p = 0.782$	$p = 0.752$	$p = 0.797$	$p = 0.515$	$p = 0.847$
Control group mean	2.08	-0.02	-0.01	0.783	0.343
	(1.55)	(0.99)	(1.00)	(0.41)	(0.80)
p-value (Instructor = Combined)	0.603	0.560	0.378	0.071	0.165
Instructors	134	134	127	130	134
Observations	6066	6066	5839	5951	4797

Column 1-3 and 5: Random effects linear estimation.

Column 4: Marginal effects from random effects probit estimation.

Standard errors in parentheses clustered at the instructor level.

Standard deviation reported for control group mean.

Dependent variable column 2, 3 and 5: course grade standardized within dept. (mean 0, s.d. 1).

Column 3 calculates grade net of weighted exam score (weights estimated using course syllabai & discussion with administrators).

Smaller sample size in column 3 due to only including observations with exam scores.

Smaller sample size in column 4 due to one department perfectly predicting completion.

Includes semester and department fixed effects and covariates for student (age, gender, race/ethnicity, baseline credits), instructor (type, contract value, discount rate) and course (co-requisite).

Exact p-values calculated based on randomization inference.

Table 5: Treatment Effects by Instructor Type

		Exam Score	Pass Exam	Course Grade	Complete Course
Adjunct	Instructor Incentives	0.262***	0.104***	0.186***	0.068***
		(0.073)	(0.039)	(0.062)	(0.025)
	Combined Incentives	0.126	0.072	0.031	-0.034
		(0.094)	(0.044)	(0.091)	(0.035)
	Student Incentives	0.100	0.081	-0.087	0.003
		(0.102)	(0.061)	(0.058)	(0.031)
Full-time	Instructor Incentives	0.128	0.041	0.002	0.002
		(0.082)	(0.039)	(0.084)	(0.029)
	Combined Incentives	0.189	0.092*	0.113	0.014
		(0.119)	(0.053)	(0.114)	(0.042)
	Student Incentives	0.020	-0.036	-0.010	-0.010
		(0.091)	(0.058)	(0.081)	(0.033)
Control group mean	Adjunct	-0.093	0.396	-0.026	0.772
		(1.01)	(0.49)	(1.00)	(0.42)
	Full-time	-0.044	0.404	-0.015	0.793
		(0.98)	(0.49)	(0.98)	(0.41)
Instructor Incentives: p-value(Adjunct = Full-time)		0.221	0.257	0.075	0.079
Instructors		127	127	134	130
Observations		5839	5839	6066	5951

Column 1 & 3: Random effects linear estimation.

Columns 2 & 4: Marginal effects from random effects probit estimation.

Standard errors in parentheses clustered at the instructor level.

Standard deviation reported for control group mean.

Dependent variable in columns 1 and 3: standardized within dept. (mean 0, s.d. 1).

Includes instructor type and instructor type interacted with each treatment.

Includes semester and department fixed effects and covariates for student (age, gender, race/ethnicity, baseline credits), instructor (contract value, discount rate) and course (co-requisite).

Smaller sample size in column 4 due to one department perfectly predicting completion.

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

Table 6: Self-reported expenditures and time use

	Control	Instructor Incentives
Panel A: Expenditure of personal funds (dollars)		
Class materials	66.375	75.863
	(19.966)	(13.201)
Professional Development	96.550	132.182
	(24.170)	(23.056)
Panel B: Time use (hours)		
Teaching class	9.727	9.483
	(0.763)	(0.596)
Preparing for class	5.273	5.944
	(0.542)	(0.530)
Preparing assignments & exams	4.268	4.079
	(0.473)	(0.429)
Grading assignments & exams	5.571	5.586
	(0.578)	(0.490)
Holding office hours	5.939	6.212
	(0.762)	(0.785)
Helping students outside of office hours	3.120	2.568
	(0.455)	(0.316)
Advising students	6.115	5.118
	(0.957)	(0.830)
Administrative work	3.264	3.977*
	(0.451)	(0.487)
Professional development	2.426	2.389
	(0.502)	(0.431)
Outside employment†	8.818	11.132
	(1.366)	(0.886)
Observations	**59**	**91**

Table reports means for each outcome by treatment group.
Standard errors in parentheses clustered at the instructor level.
Observations are at the instructor-semester level.
Significance tests conducted using random effects regression.
including controls for semester and instructor full-time status.
† 51 of 101 responses are top-coded for outside employment ($>$ 16 hours).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Treatment Effects on Non-Program Courses and Overall Outcomes

	Courses Outside the Study				Outcomes at End of Program			
	Completion		Grades		Credits		GPA	
Total Treated Courses	0.032***		0.087***		0.441***		0.028	
	(0.008)		(0.032)		(0.171)		(0.026)	
Any Treated Course		0.021*		0.065		0.531***		0.007
		(0.012)		(0.048)		(0.205)		(0.038)
Total Courses in Study	-0.024***	-0.012**	-0.069***	-0.038*	2.420***	2.547***	0.025	0.038**
	(0.006)	(0.005)	(0.026)	(0.022)	(0.152)	(0.132)	(0.022)	(0.019)
Total Courses	0.014***	0.014***	0.048***	0.048***	0.928***	0.927***	0.036***	0.036***
	(0.002)	(0.002)	(0.007)	(0.007)	(0.041)	(0.041)	(0.005)	(0.005)
Control Mean	0.778		2.149		24.492		2.420	
	(0.342)		(1.373)		(22.016)		(1.173)	
Students	3535	3535	3535	3535	3575	3575	3575	3575

Linear regression estimation.

Heteroskedasticity robust standard errors in parentheses.

"Pure-Control" group mean is for students never attending a treated course.

Standard deviation reported for group mean.

Observations at the student level.

Total Treatment sections include both Instructor Incentives & Combined Incentives sections.

Includes covariates for student (age, gender, race/ethnicity, baseline credits, total courses).

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

Table 8: Effects of Incentives on Graduation, Transfer, and Dropout

	Transfer to 4-Year		Associate Degree		Graduate Certificate		Dropout	
Total Treated Courses	0.028***		-0.008		-0.005		-0.012	
	(0.008)		(0.009)		(0.006)		(0.011)	
Any Treated Course		0.022**		-0.011		-0.004		-0.006
		(0.011)		(0.013)		(0.008)		(0.016)
Total Courses in Study	-0.023***	-0.012**	-0.000	-0.002	0.007	0.005	-0.029***	-0.033***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.005)	(0.004)	(0.010)	(0.009)
Total Courses	0.007***	0.007***	0.028***	0.028***	0.006***	0.006***	-0.022***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Control Mean	0.099		0.236		0.054		0.503	
	(0.299)		(0.425)		(0.226)		(0.500)	
Students	3575	3575	3575	3575	3575	3575	3575	3575

All estimates are marginal effects from probit regression.

Standard errors in parentheses clustered at the instructor level.

"Eligible Courses" represents the total number of courses that could have been assigned to the treatment.

Full-Time student is defined as 2 or more courses in a semester.

"Pure-Control Mean" is for all students (full- or part-time) who never received a treated course.

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

Table 9: Treatment Effects on Preference for Loss-Framed Contracts

	Baseline	Fall Endline Preference		
Fall Instructor Incentives	0.073	4.326**	4.529**	4.351***
	(2.307)	(2.069)	(1.782)	(1.666)
Full-Time	-0.191	-0.065	-0.978	0.337
	(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 -$0.5)			-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	0.976
			(10.180)	(8.285)
Discount Rate (δ)				-4.487***
				(1.305)
Pr(Treatment Group Value = $0)	0.001	0.291	0.499	0.850
Instructors	89	90	89	88

Dependent variable: per-student bonus willing to pay for loss-framed contract.
Generalized tobit regression to correct for interval-censored data.
Heteroskedasticity-robust standard errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.