HOW DOES PEER PRESSURE AFFECT EDUCATIONAL INVESTMENTS?*

Leonardo Bursztyn
Anderson School of Management, UCLA
and
NBER

Robert Jensen
Wharton School, University of Pennsylvania
and
NBER

Abstract: When effort or investment is observable to peers, students may act to avoid social penalties by conforming to prevailing norms. We explore this hypothesis in two settings. We first consider a natural experiment that newly introduced a leaderboard into existing computer-based remedial high school courses, revealing top performers to the rest of the class. The result was a 23 percent decline in performance. The decline appears to be driven by a desire to avoid the leaderboard; for example, students performing at the top of the class prior to the change, those most at risk of appearing on the leaderboard, had a 38 percent decline in performance, while those previously at the bottom improved slightly. Our second setting involves a field experiment that offered students complimentary access to an online SAT preparatory course. Signup forms differed randomly across students only in the extent to which they emphasized that their decision would be kept private from classmates. In non-honors classes, the signup rate was 11 percentage points lower when decisions were public rather than private. Signup in honors classes was unaffected. For students taking both honors and non-honors classes, the response differed greatly based on which peers they happened to be sitting with at the time of the offer, and thus to whom their decision would be revealed. When offered the course in one of their non-honors classes (where peer signup rates are low), they were 15 percentage points less likely to sign up if the decision was public rather than private. But when offered the course in an honors class (where peer signup rates are high), they were 8 percentage points more likely to sign up if the decision was public. These results show that students are highly responsive to who their peers are and what the prevailing norm is when they make decisions.

* We would like to thank Nava Ashraf, Ernesto Dal Bó, Leigh Linden, Aprajit Mahajan, Torsten Persson, Bruce Sacerdote, Noam Yuchtman, numerous seminar participants and Lawrence Katz, Andrei Shleifer and four anonymous referees for comments and suggestions, and Pedro Aratanha, Andrea Di Miceli, Stefano Fiorin, Craig Jones, Vasily Korovkin, Matthew Miller and Benjamin Smith for excellent research assistance. We are grateful to the UCLA Anderson Price Center and the California Center for Population Research for financial support. Our study was approved by the UCLA Institutional Review Board and the Los Angeles Unified School District Committee on External Research Review.

I. INTRODUCTION

It has long been suggested that students may be motivated as much by the desire to gain social approval (e.g., being popular or fitting in) or avoid social sanctions (e.g., being teased, made fun of or ostracized) as they are by the future benefits of education (Coleman 1961). An important question then arises as to whether, and how, student effort or investments are affected by such peer social concerns or peer pressure.[1] Are students willing to deviate from what they privately believe to be the optimal scholastic effort or investment decision just because of such social concerns? In this paper, we test this hypothesis using both a natural experiment and a field experiment.

Despite the perception that peer pressure is widespread, there is very little direct empirical evidence of its effects.[2] Adolescence is believed to be the period of greatest vulnerability to peer pressure, during which the desire to be popular or fit in is felt most acutely (Brown 2004). Adolescents may also be more likely to give in to such pressure and engage in behaviors that can have long term effects. Though many studies have found peer effects in education,[3] there are many mechanisms through which they might occur.[4] A few studies have found peer social pressures for adults in the workplace. Mas and Moretti (2009) find that the productivity of supermarket cashiers is affected by coworkers who can see them (particularly those they interact with more), but not those who can't. Bandiera, Barankay and Rasul (2010) find that the productivity of fruit pickers is affected by those working alongside them, but only when they are friends.

We begin by examining how the introduction of a system that revealed top performers affected subsequent performance. The natural experiment we consider was applied to a computer-based learning system used in over 100 high schools located predominantly in one American state. The system is used primarily for remedial English and math courses, particularly in preparation for a high-stakes, statewide high school exit exam. Prior to the change, students would answer

[1] We define peer pressure as students taking actions that deviate from what they privately consider to be the optimal action (what they would do if others would not observe their actions) in order to achieve social gains or avoid social costs. Peer pressure thus doesn't just refer to active efforts by peers to persuade others, but could also include passive effects such as not undertaking an action for fear of peer social sanctions.
[2] Some studies in social psychology measure peer pressure through direct survey questions, such as by asking whether a student has faced pressure from others to undertake certain actions (Brown 1982, Brown et al. 1986 and Santor et al. 2000). However, there is some concern with using such subjective self-reports, and further, it is difficult to link these responses directly and causally to specific behaviors.
[3] Sacerdote (2001), Zimmerman (2003), Carrell, Fullerton and West (2009), Duflo, Dupas and Kremer (2011) and Carrel, Sacerdote and West (2013). See Epple and Romano (2011) for a summary.
[4] Bursztyn et al. (2014) and Cai et al. (2012) examine channels of peer influence in financial settings.

multiple-choice questions and receive private feedback on whether their answers were correct. About one month into the 2011−12 school year, without any advanced notice or explanation, the system was changed slightly. Students were now awarded points for correct answers. Simultaneously, home screens provided tabs that revealed the names of the top three scorers in the classroom, the school and among all users of the system, as measured by cumulative points received for the past week, month and all time. Finally, each tab also showed students their own rank (in the classroom, school and among all users, for the past week, month and all-time). There were no other changes to the system.

We find that the introduction of the leaderboard led to a 23 percent decline in performance. We also provide evidence that these results are driven by an aversion to being on the leaderboard. Because students had already been using the system for over a month before the change, they would have had some private information about their own performance, and thus their risk of showing up on the leaderboard if they continued performing at their previous level. We find that students in the top quartile of the pre-change distribution of performance, those most at-risk of showing up on the leaderboard, on average had a 38 percent decline in performance. Further, this decline comes primarily through attempting fewer questions (not getting fewer questions correct), and includes reduced discretionary use of the system outside of school; both are consistent with an active choice to reduce effort. Further, these students cut back almost immediately, with declines evident on the very first day of the change. By contrast, students at the bottom of the pre-change performance distribution did slightly better following the change. The pattern across the distribution of pre-change performance is monotonic; on average, the better you were performing before the leaderboard was in place (and thus the greater your risk of being in the top three), the more your performance declined afterwards. In other words, it appears that at least some students were willing to work hard and do well only as long as their classmates wouldn't know about it.

To further isolate and test for the effects of peer pressure, and to see whether these effects apply beyond remedial students, we next turn to our field experiment. In four low-income Los Angeles high schools, we offered 11th grade students complimentary access to an online SAT preparatory course from a well-known test preparation company. Across students within classrooms, we randomly varied whether the signup forms emphasized that the decision to enroll would be kept private from the other students in the classroom. In particular, students were either told that their decision to enroll would be kept completely private from everyone *including* the other students in the room, or *except* those students. Notably, the sole difference between signup forms in our "private" and "public" treatments was a single word (including vs. except).

We use both honors and non-honors classes for the experiment. The prep class is an educational investment, and making it observable to peers could carry different social costs in settings where the norms on the acceptability of effort differ, such as in the two types of classes. We find that observability has a large impact on signup rates, and that the effects do differ across settings. In non-honors classes, signup is 11 percentage points lower when students believe others in the class will know whether they signed up, compared to when they believed it would be kept private. In honors classes, there is no difference in signup rates under the two conditions.

Consistent with these results being driven by peer social concerns, in non-honors classes, students who say that it is important to be popular are less likely to sign up when the decision is public rather than private, whereas students who say it is not important are not affected at all. In honors classes, students who say that it is important to be popular are slightly more likely to sign up when the decision is public, whereas those who say it is not important are again unaffected. In both cases, students concerned with popularity move in the direction of the locally prevailing norm when the decision is public, whereas those unconcerned with popularity are unaffected.

The differential response to observability by class type could still be consistent with explanations other than peer social concerns. For example, students in honors and non-honors classes may differ from each other in many ways, which may affect how much they care about privacy or how they respond when decisions are observable. Though this would not change the policy implication that observability affects signup in non-honors classes, to test peer pressure even more cleanly we can address this selection problem and make the set of students we examine in honors and non-honors classes more similar by focusing on students taking both honors and non-honors classes. Students are free to choose whether to take an honors or non-honors class for every subject for which both are available. To fix ideas, consider the set of students taking exactly two honors classes (hereafter, "two-honors" students). Honors classes are spread throughout the day, but our team showed up for just two periods. The timing of our arrival will find some two-honors students in an honors class and others in a non-honors class. Just as important, the timing of our visit, and therefore which type of class we find them in, will be uncorrelated with student characteristics. Thus, though this approach does not explicitly randomize peers, the set of two-honors students who happen to be sitting in one of their honors classes when we arrive and conduct our experiment should be similar in expectation to those who happen to be sitting in one of their non-honors classes – all that will differ is whether they are at that moment sitting with their honors or non-honors peers. Further, because we are not actually changing a student's peers at all (or their teachers, schools, neighborhoods or anything else in their environment), we can rule out most other channels through which peers may influence each other.[5] We will capture the effect of varying just to which of a student's peers the signup decision will be revealed, and thus whether and how those peers punish or reward observable effort.

Examining the set of all students taking some (but not all) honors classes (hereafter, "some-honors" students), we find that making the decision to enroll public rather than private decreases signup rates by 15 percentage points when they are in one of their non-honors classes (where the signup rate among their classmates is low). In stark contrast, making the decision public *increases* sign up rates by 8 percentage points when they are in one of their honors classes (where the signup rate among their classmates is high). Viewed another way, when decisions are observable to peers, the signup rate for these students is over 20 percentage points lower when they are with their non-honors peers rather than their honors peers.

Both settings suggest peer social concerns affect educational investment and effort. When faced with the tradeoff between the future benefits of academic effort and the present potential social costs, some students choose to reduce effort and performance (though conforming to the locally prevailing norm can also induce greater effort or investment, as observed in honors classes under the public treatment). The fact that we find similar effects in two entirely independent settings suggests that these effects may be more widespread.

Beyond providing insights into behavior, we believe these results carry important policy lessons. In both cases, peer awareness has dramatic effects on investment or performance. Our natural experiment reveals that some students are willing to do well only as long as their classmates don't know about it. In our field experiment, in non-honors classes, even very low-income students are willing to forgo free access to an SAT prep course that could improve their educational and possibly later life outcomes, solely in order to avoid having their peers know about it. Changing either norms or peers is likely to be quite difficult, particularly on a large scale;[6] changing the

[5] We can also rule out social learning since signup is done before students know what their peers did.
[6] And Carrel, Sacerdote and West (2013) show that even when you construct peer groups, students may re-sort into more homogenous subgroups. Also, the extent to which changing peer groups can help is limited by the fact that if enough students are moved across groups the dominant norm may change.

extent to which behavior is observable to peers is likely to be less so. This is particularly important given that many efforts students can make are observable, such as participating in class, joining study groups or seeking extra credit or extra help. Further, awards, honors or other forms of recognition that publicly reveal high performance may discourage effort for some students.

The finding from our field experiment that our sample of predominantly Hispanic students in non-honors classes are less likely to sign up for the prep course when it is observable is consistent with the "Acting White" hypothesis, whereby minorities may face peer sanctions for engaging in certain behaviors (Fordham and Ogbu 1986, Austen-Smith and Fryer 2005, Fryer 2011 and Fryer and Torelli 2010).[7] Our results are also relevant to other models of social interactions,[8] including those that examine the role of culture in shaping interactions or the tradeoff between intra-group participation and economic achievement in society as a whole. Such effects have been explored for a variety of ethnic and cultural groups (Berman 2000, Gans 1962, Lee and Warren 1991, Ausubel 1977). Several studies have also documented the benefits that such networks provide, including helping members find a job or providing informal insurance.[9] The broader issue of group loyalty versus economic success has also been modeled by Austen-Smith and Fryer (2005) and Fryer (2007). For example, when students face a tension between investing in activities rewarded by the labor market and signaling loyalty to a group, one possible equilibrium involves sorting wherein higher ability individuals invest in the labor market oriented activities rather than those likely to increase acceptance by the group, and lower ability individuals choose the reverse. The differential responses by class type we observe would be consistent with such sorting, with social penalties only for students in non-honors classes. Further, for some-honors students, the different response based on which peers may observe their behavior demonstrates this tension between a desire to invest and the costs of peer sanctions. The fact that the investment was an SAT prep course is particularly relevant given that it signals, more than many other actions, a likelihood that the individual will leave the community or group (since you only take the SAT if you plan to go to college) and is thus the very type of behavior we expect to be sanctioned under these models.

[7] However, since the schools we study are 96% Hispanic, we cannot provide a more complete test of this hypothesis. And unfortunately, the data from our natural experiment do not include student ethnicity.
[8] See Akerlof (1997), Becker and Murphy (2000), Durlauf and Ioannides (2010), and Postlewaite (2011).
[9] Townsend (1994), Ligon (1998), Topa (2001), Munshi (2003), Magruder (2010), Wang (2013). See Munshi (2014) for a summary. Thus, we do not take a stand on whether conformity may be privately or socially welfare enhancing.

Beyond this, the results showing how differences in peers and locally prevailing norms regarding accepted vs. sanctioned behavior can affect investments is also relevant to the literature examining the role of schools and neighborhoods in the educational outcomes of poor and minority students (Dobbie and Fryer 2011, Fryer and Katz 2013, Jacob 2004, Kling, Liebman and Katz 2007 and Oreopoulos 2003).

The remainder of this paper proceeds as follows. In section II, we discuss the natural experiment in more detail. Section III discusses the field experiment, and section IV concludes.

II. LEADERBOARD NATURAL EXPERMENT

II.A. Background and Policy Change

Many schools use in-class, computer-based learning materials created and operated by private companies. The company responsible for the software we consider was operating in over 200 high schools across several states (though primarily in just one). The sample of schools we examine is not random; they are simply the ones that chose to use software from this company.[10]

Though many courses are available, the most widely used are 10th and 11th grade remedial English and Math, including courses designed for statewide high school exit exams. Schools require students to take these courses if they scored in the lowest proficiency levels on the previous year's statewide standardized test or if they failed the high school exit exam.[11] The fact that students taking these courses are low-performing is relevant for generalizability; these students might for example be more sensitive to peer social stigma than the average student.

Students are given individual online accounts. When logged in, they have access to a large database of questions. Questions are multiple choice, and after each question students receive private feedback on whether their answer is correct. The questions are organized into modules that typically follow along with in-class instruction. But students have some discretion in how many questions they choose to answer (the database was sufficiently large that students will not run out of questions). Students can also access the system at any time from any computer or device with an internet connection for additional, voluntary learning opportunities.

[10] Students in the schools using this software are on average poorer and more likely to be minorities when compared to other schools in the state where the system is most widely used.

[11] Students can first take the exam in 10th grade, giving them several chances to pass before graduation.

On September 20, 2011, without any prior notice or any explanation, the company introduced a point system and a series of rankings and leaderboards, intended to encourage and motivate students. Students were now awarded 1,000 points for answering a question correctly on the first try, 325 points for a correct answer on the second or third tries, and no points after that. There was no penalty for incorrect answers; thus students could increase their score by getting more questions correct on the first, second or third tries, or by attempting more questions.

The second aspect of the change is that students could access a series of tabs on their homepage that showed the names of the top three scorers (based on cumulative points, starting at zero on the day of the change) in their class, school and among all users of that course. These leaderboards were updated in real time, and were separately available for the past week, month and all-time. The third change is that students could also see their own personal rank (again, in their class, school and among all users, and for the past week, month and all-time). However, below we argue that information on rank itself is unlikely to have an independent effect on our results. The system was otherwise completely unchanged during the period of our analysis.

II. B. Empirical Strategy

The point system and leaderboard were introduced at the same time to all users of the system. We therefore have no cross-student variation in exposure that can be used to identify the effects of the change. Instead, we start by just examining how performance changed upon the introduction of the system on September 20, 2011. Since we have data on the same students over time, we also include student fixed effects. Thus, we estimate:

$$Y_{i,t} = \beta_0 + \beta_1 Post_t + \alpha_i + \epsilon_{i,t}$$

where $Y_{i,t}$ is the number of questions answered correctly by student *i* on day *t*, *Post* is an indicator for before vs. after the policy changes, and α_i is the student fixed effect. We trim the post period to one month to match the approximately one month available prior to introduction.[12] The identifying assumption is that had it not been for the change in the system, there would have been no change in student performance around this date.[13] Since this is a strong assumption, below we conduct placebo tests using all other dates in place of September 20, 2011.

[12] Appendix Tables A.I and A.II show that the estimates are similar if we use one or two week intervals.
[13] All results are robust to adding linear time trends and trends interacted with the *Post* dummy to account for general changes in performance over the course of the school year); see Appendix Table A.III.

Under this assumption, this strategy identifies the net effect of introducing the points and leaderboard system, which itself may have implications for school policy. In order to explore the possible role of peer sanctions, we exploit the fact that the potential for being newly exposed to such sanctions will be greater for some students than others. Because students had been using the system for more than a month before the change, they would have had an estimate of their own performance (feedback could also come from exams, exercises or directly from teachers). This would likely have included some signal of relative performance as well. For example, in the extreme cases, a student who was getting most answers correct will likely perceive a higher leaderboard risk than a student who almost never got any answers correct. Therefore, students would have had an approximate sense of whether, if they continued their performance unchanged, they were likely to be among the top performers in the class and have this information revealed to others through the leaderboard. These are the students we predict will be the most likely to reduce effort if a fear of peer sanctions is operative.

Even though there were no points or leaderboards before the change, the company still captured data on all questions answered by each student. We therefore construct a measure of pre-change performance and leaderboard risk by examining the number of correct answers a student had in the month prior to the change, and look for differential responses by estimating separate regressions for each quartile of this distribution (computed within each classroom).[14] Beyond providing evidence for the potential role of peer sanctions, comparing across quartiles of pre-change performance in this way also allows us to weaken the identifying assumption by netting out changes in other factors that affect all students equally.

II.C. Data

We have data for the universe of questions answered. We restrict the sample to the most widely used courses, remedial English and Math.[15] We also exclude classrooms with fewer than 5 students (since then a top 3 leaderboard is not very informative for the class, and/or quartiles can't

[14] Though leaderboards were also available for the school and all users, the classroom is likely to be the most relevant set of peers that students care about, and the one students are most likely to check. And to be a top performer in the school or among all users, they must also be a top performer in their class.

[15] Other courses were more specialized and had very few students enrolled per school.

be computed) and those not using the system before the change. This leaves us with a sample of over 5,000 students across more than 100 schools.[16]

Each student is uniquely identified by an ID code. However, we have no other individual level data other than their first and last name. We can assign a likely gender to each student by matching to the Social Security Administration's database of gender frequencies by first name. Using a five year interval around the birth year for students in our sample, we assign a student as male or female if at least 80 percent of children with the same name born in those years was of that sex. This yields a likely gender assignment for 95 percent of students.

Because the data are click-based, if a student does not attempt to answer any questions on a particular day, they simply have no data for that day. However, we do not want to recode all such days as having gotten zero correct; for example, teachers do not use the system every day, and there are days when class is not held (e.g., school-wide assemblies, or school holidays or closings). However, we do not have variables that identify such days in our data set, and thus we don't know on which days a student chose not to answer any questions (in which case they should be recorded as a zero) vs. a day on which the system was just not used in their class. Therefore, in analyzing the data we recode a student as having answered zero questions correctly if they did not attempt any question but at least two other students in their class did.[17,18] The results are very similar if we use other thresholds, or if we don't recode any observations to zero, since most students attempt at least one question when others in their class also do.[19]

---

[16] Teachers and company employees occasionally created accounts to test the system. These cases create extreme outliers, e.g., answering over 300 questions correctly in a single day (the median, conditional on answering any, is 6), and then answering no questions on any other day. Since there is no variable to identify these cases, we trim the top one percent of observations (though the results are robust to including them).

[17] We use two students since there may be cases where a single student logs on accidentally, and since use by a single student is less indicative of the likelihood that all students were expected to work on that day.

[18] One problem is that this treats individual absences as getting no questions correct. This will not induce bias if absences are uncorrelated with the leaderboard timing. But absences may vary over the course of the year, which would cause differential recoding for our pre and post periods. We believe this to be unlikely, both since our placebo results show no similar changes around other dates, and since analysis across quartiles should account for any such effects that are common to all students.

[19] If all students stop answering questions when the leaderboard is in place, this approach would not count them as zeroes and we would underestimate the effects of the change. However, if our hypothesis is incorrect and the leaderboard does motivate students such that before, on some days not a single student would attempt even one question, but now they do, we will overstate the effect of the change. However, the share of days on which at least one student uses the system does not change at all after the leaderboard.

II.D. Main Effects

We first provide visual evidence of the effects in Figure I. The figure plots the average number of correct answers per day, separately for each of the four quartiles of the within-classroom distribution of performance in the month before the system was introduced. We also fit linear trends for the pre- and post-change periods separately, along with 95 percent confidence bands. For previously high performing students (quartiles 3 and 4), performance declines sharply on the day that the leaderboard is introduced (this first-day drop is statistically significant in regressions for quartile 4 if we include just the day before and day of the change, or if we compare to the same day of the previous week). The number of correct answers then stays lower for the remainder of the period (this persistence is confirmed by regressions that exclude the first day or first few days after the change). By contrast, there is no decline for students in quartiles 1 and 2. There is a decline on the day of the change for those in quartile 2, but the decline is not persistent. For students in quartile 1, the effect is an increase, which again even shows up on the very first day.

Table I shows regression results that confirm the visual evidence (robust standard errors in brackets).[20] Column (1) shows that the effect of the program across all students was negative. After the system is introduced, on average students answer 0.63 fewer questions correctly per day (significant at the 1 percent level). This is a 23 percent decline from the baseline of 2.7.

Columns (2) − (5) provide the results for each quartile of pre-change performance separately, which again allows us to both explore how the response to the leaderboard varies by leaderboard risk and helps net out other changes common to all students (results are similar if we pool the samples and use interactions). For students in the top quartile in column (5), the change was associated with answering 1.93 fewer questions correctly per day (significant at the 1 percent level). This represents a 38 percent decline from the pre-change baseline of 5.1.

As we move down from the top quartile in the pre-change performance distribution the effects on performance become less negative, and eventually for the bottom quartile turn positive

[20] The imbalance in observations across quartiles arises because most class sizes are not perfectly divisible by 4. Given the numbers 1, 2, 3, 4 and 5, most software defines the quartile cutoffs as 2, 3 and 4, using the higher of the two potential cutoffs for the first quartile. So in a classroom of size 5 the first quartile will have 2 observations and the other three will each have 1. More generally, one extra observation will be added to the first quartile for any class size yielding a remainder of 1 when divided by 4. With 6 observations (or a remainder of 2 when divided by 4), the first cutoff is the higher of the two candidate values and the second is the average of the two candidate values (for 1, 2, 3, 4, 5, and 6, the cutoffs are 2, 3.5 and 5). With 7 observations, the higher cutoffs for all three quartiles are used (cutoffs of 2, 4 and 6 for values of 1, 2, 3, 4, 5, 6 and 7). On net, more observations get assigned to quartile 1 and fewer to quartile 4, as observed here.

(in all cases, the results are significant at the one percent level). Again, these results are suggestive of a role for social sanctions, since it is the students who likely perceive the greatest risk of being in the leaderboard, and thus having their high performance publicly revealed, who cut back most.[21]

In Appendix Table A.IV, we find no gender differences in response to the change, either for the full sample or any quartile; the interactions between gender and the *Post* dummy are all small and not statistically significant (though we cannot rule out that some students' gender is misclassified, which would bias the coefficient towards zero). In separate results, we find no significant gender-by-subject interaction effects (e.g., girls reducing more in math classes).

So far, our analysis looks at a simple before vs. after comparisons around the specific date of the change. To explore the plausibility that the introduction of the points and leaderboard system caused these changes, we can consider the uniqueness of these results. Thus, in a series of placebo tests, we run the same regressions as above, but assign the "change date" to every alternate date starting one month after the true change date,[22] and ending one month before the end of the school year. For quartile 4 on its own, none of the 218 available placebo dates yield a greater decline in performance than the −1.93 found for the true change date. None yields a point estimate greater than 1.93 either; so even in a two-sided test, no other date in our sample yields as large a change in performance for this previously top-performing group as the day of the leaderboard introduction. Appendix Figure A.I provides a histogram with the distribution of placebo treatment effects, and shows that the estimated decline around the true date is an outlier in terms of sustained changes in student performance around any specific date. Such large and sustained increases or decreases in performance never occur for any other quartile either. Running the placebo tests for the other quartiles, no alternate date yields an estimated increase or decrease in performance of 1.93.

II. E. Alternative Mechanisms

One alternative mechanism to consider is whether performance declined due to the pressure of competition created by the leaderboard. We believe this is unlikely to explain our results. First,

---

[21] Of course, if all students equally feared peer social sanctions, we might expect them all to cut back to zero. However, with heterogeneity in disutility from peer sanctions (for example, for our field experiment below we show differences in effects by self-rated importance of popularity) then we will shift from a situation where top performers are the highest ability or most motivated to one where they are instead those who care the least about peer stigma (or, who are less likely to actually face such stigma).

[22] We start one month after the true date because our regressions use a one month interval around the change date, so any date less than one month after the true date would capture the effects of the true change.

using regressions like those above, column (1) of Table II shows that for the full sample, the percent of questions attempted that are answered correctly actually increases slightly after the leaderboard is introduced.[23] Thus, the overall decline in questions answered correctly found above is not due to a decline in performance on questions attempted, as might be expected if the pressure of competition were driving our results, but instead due entirely to attempting fewer questions (consistent with an effort to avoid the leaderboard).[24] For the top quartile, there is a 2 point decline in percent correct (from a base of 64 percent); still, over 90 percent of the total decline in questions answered correctly is attributable to attempting fewer questions.[25]

We can also explore use of the system outside of class.[26] Students working outside of school can access the system whenever they want, attempt as many questions as they like and take as long as they want to answer. They can also seek assistance from other people or resources or even collaborate with friends. Further, when working from home they are not facing off against other students in real time. Thus, the system affords them a great deal of opportunity to improve their performance in a less competitive environment if they choose to (but points earned at home do count towards the leaderboard). If competitive pressure is adversely affecting performance, we might expect less of a decline for outside of school use.[27] Table III shows that if we confine our analysis to use before 7:00AM or after 4:00PM[28] or on weekends, we find very similar patterns as above. Beyond suggesting competitive pressure is not likely at play, these declines in discretionary use at home again suggest an active choice to answer fewer questions, particularly by students previously at the top of the distribution, in order to avoid being on the leaderboard.

A second alternative to consider is whether information on rank affected performance.[29] For example, some students may not have known before how well they were doing relative to

[23] These regressions have fewer observations because percent correct is only defined on days where a student attempts at least one question.

[24] Though students may choose to avoid competition by not answering questions, or may be able to answer fewer questions under pressure, so we cannot completely dismiss that competition plays a role.

[25] Further, answering questions incorrectly could be an intentional effort to avoid the leaderboard, and is not necessarily entirely attributable to competitive pressure, and thus this is perhaps a lower bound.

[26] Not every student has equal access to the internet outside of school, but that should not vary much around the cutoff date, much less differentially by quartile (and for example it is unlikely that those in the top quartile suddenly had less access afterwards).

[27] Though again, some students may be so competition-averse that they choose to just stop using the system.

[28] The results are robust to alternate time cutoffs for in vs. out of school.

[29] Some studies have documented effects of rank. Tran and Zeckhauser (2012) find that providing private information on rank improved the performance of Vietnamese undergraduates. Barankay (2012) finds that eliminating information on rank improves performance among furniture salespersons, with additional

classmates, and may then decide that they don't need to work as hard (they could also feel that their hard work is paying off and decide to try harder). Students at the bottom of the distribution might not have known how far behind they were, and may try harder (or become demoralized and stop trying). However, we believe this is unlikely to explain our results. First, since the change took place over a month into the school year, students would have already had some information on their performance from exams, assignments or feedback from teachers. Second, as noted, the decline for previously top-performing students occurred on the very first day of the change. Since all students started at zero, during that first session there would not have been enough time for students to answer a lot of questions correctly, infer from this brief performance that they are (persistently) at the top of the class (beyond what they already knew prior to the change), and still have time to cut back enough (including overcoming their strong performance at the start) that we would see a large net decline on that very first day. Similarly, it is unlikely that rank gave teachers enough new information on performance to allow them to redirect attention or resources away from students performing well and towards those needing more help on that very first day. Similar arguments would hold for why information on rank is unlikely to explain the increases for the low performing students on the first day.

A third alternative to consider is whether the leaderboard reduced intrinsic motivation (Camerer and Hogarth 1999 and Bénabou and Tirole 2003, 2006). Though such loss of motivation could affect all students, it might affect those previously at the top more since they are more likely to be "rewarded" for their efforts by appearing on the leaderboard.[30] Though we cannot rule this out completely, we believe it is unlikely to explain our results. First, there is very little actual reward associated with the leaderboard. Most studies of intrinsic motivation crowd out have focused on financial incentives. It seems unlikely that students will believe their effort is "cheapened" by the possibility that by doing well, they might appear on the leaderboard. It also seems likely that students performing well in a course designed to help them pass the high school

evidence suggesting a demoralization effect whereby those receiving a lower than expected rank reduce performance. Ashraf, Bandiera and Lee (2014) find that providing information on rank decreases performance, particularly for those at the lower end of the distribution. Further, this effect holds even before rank is revealed, potentially consistent with “self-handicapping” (Benabou and Tirole 2002).

[30] Though most discussions of intrinsic motivation would suggest the opposite pattern than we observe. If the leaderboard crowds out prior intrinsic motivation, low performing students might suffer the biggest declines because they will not have that motivation replaced by extrinsic motivation, since they know they are unlikely to be rewarded; see Leuven, Oosterbeek and van der Klaauw (2010).

exit exam and graduate were not motivated much by intrinsic motivation, or, to the extent that they were, that this small recognition would reduce that motivation. Second, most studies in academic settings have found that rather than reduce motivation, financial awards either improve or do not affect performance. Further, using direct self-reports, Kremer, Miguel and Thornton (2009) find no evidence that academic financial rewards reduced intrinsic motivation for girls in Kenya.

A fourth alternative to consider is whether the changes may have altered student behavior or interactions in the classroom in other ways. For example, students may repeatedly check the leaderboard for updates during time allocated to answering questions, distracting themselves or others and reducing time available for answering questions. However, as noted above, we find nearly identical patterns when examining questions answered outside of class hours, when students are less likely to be with others and when so few are simultaneously using the system that the leaderboard does not change rapidly in real time. Further, students have much more time at their disposal when working outside of class so any distraction should have less of an impact.[31]

We can also rule out alternative explanations such as the decline in performance simply being due to the sudden change or newness of the system. First, the decline was not common to all students, and in fact was found only for previously high performing students, while worse performing students actually improved. Second, the effects appear to persist beyond the first day's decline (as evidenced by Figure I, and by the fact that the regression results are robust to excluding the first day or first few days after the change), whereas over time students should become more familiar with the system and thereby improve.

As a final alternative, we consider mean reversion (top performing students do worse, low performing students do better). However, there is no explicit design that would lead to such reversion; questions are drawn from the database at random, so question difficulty is not a function of previous success. We also believe that statistical or incidental mean reversion is unlikely to explain our results. First, the pre-change quartile is based on over a month of performance, so any randomness or luck is likely to have balanced out. Second, as noted, the biggest change is in the number of questions attempted, not the percent correct; this likely reflects a conscious choice of effort, whereas a student simply on a lucky (unlucky) streak would likely experience a decline (increase) in the percent answered correctly. Finally, in terms of mean reversion driven by students

[31] Similarly, declines in use out of school, plus the fact that the declines are not uniform across quartiles, also suggest that the changes were not due to changes in how teachers interact with students in the class.

putting in a fixed amount of effort, but just varying the timing (e.g., some students work hard upfront and tail off, while others wait to start working hard) the fact that we find no other changes this large in our placebo test around any other date suggests that this is unlikely to be the case.[32]

II. F. Summary

The results from this natural experiment suggest that students actively reduce effort and performance in order to avoid appearing on the leaderboard. However, we cannot completely rule out other potential channels, nor can we isolate peer pressure as the reason for leaderboard aversion (though we do note that to the extent that we cannot rule out that factors such as competition or reduced intrinsic motivation explain our results, we note that most other policies or practices that recognize top performers (such as awards or honors) will also create these other effects as well, so the results still carry important implications for school practices). Further, there remains a question of whether such effects are found more widely, or among non-remedial students. To address these questions, we next turn to our field experiment.

III. FIELD EXPERIMENT

III.A. Experimental Design

We conducted our experiment in the four largest public high schools in a disadvantaged area of south Los Angeles.[33] We visited each school once, between December 2013 and April 2014. The sample was confined to 11thgrade students, since this is when many students begin preparing for the SAT. The four schools each have around 3,000 students. In addition to being larger on average, these schools have a higher share of students eligible for free and reduced price meals (84% vs. 68%) and students of Hispanic ethnicity (96% vs. 69%) compared to the average school in the Los Angeles Unified School District (LAUSD). The median income in the ZIP codes around these four schools is also lower than for all schools in the district ($39,533 vs. $48,898).[34]

[32] Further, the material changes over the course of the year, so answering a lot of questions early on would not prepare a student for all of the material in the class or on the exit exam they are preparing for.
[33] We focused on the largest high schools for logistical and budgetary reasons. To prevent communication among students that could contaminate the experiment, we needed to conduct our experiment simultaneously in one period across different classrooms, or in two class periods immediately following each other, with no overlap of enrolled students. Achieving a sufficiently large sample with a limited budget therefore required visiting large schools with many classes running simultaneously each period.
[34] Source: California Department of Education (http://www.cde.ca.gov/ds/dd/), for academic year 2012-3.

We would therefore not want to generalize our results to other schools. However, we do note that these schools account for approximately 7 percent of all high school enrollment in the LAUSD. Further, from a policy perspective, low performing schools such as these are the ones where it is perhaps most important to understand the barriers to educational investments and performance. Finally, we note that despite these differences, the fraction of seniors in these four schools who take the SAT is the same as for LAUSD as a whole (51%).

Within each school, our visits were coordinated with principals and counselors to choose on what day we could visit and during which period(s). These considerations were typically about scheduling logistics for the schools and our research team. During the selected periods, we visited honors and non-honors classrooms, across a range of subjects. Overall, we visited 26 classrooms across the four schools, with a total 825 students (all of whom participated in the study; we did not contact absent students). Neither students nor teachers were informed about the subject of our visit or that there would be an intervention related to the SAT (principals were informed in advance, but agreed not to communicate the purpose of our visit ahead of time).

Students in the selected classrooms were offered the opportunity to sign up for free access to a commercial, online SAT preparation course. The course was created by a well-known test prep company that students in these schools are familiar with. The course includes practice exams, a library of pre-recorded videos and instructional content, live online class sessions, analysis of individual performance with areas requiring additional focus and test taking strategy.

Prior to our study, no students in these schools were using the course. The company does not currently offer this software to individuals, instead selling subscriptions to schools, who then make it available to individual students (at a cost of about $200 per student). None of the schools in which we conducted our study had purchased this software prior to our intervention. In a separate follow up survey at one of our schools (conducted after the intervention), we asked students to estimate the cost of the software; on average, they estimated the value at $260. Thus, especially for these low income students, this is a valuable offer (perceived and actual) that they would be forgoing if they chose not to sign up (confirmed by the fact that signup rates are very high when the decision is private). Finding that observability alone is sufficient to deter signup would be an indication that these peer social concerns can be quite powerful.[35]

[35] Though of course not all students plan to take the SAT, and it would be of little value to such students (unless they gave away or sold their online access to someone else, which we did not explicitly preclude).

After a brief introduction, students were given a form offering them the opportunity to sign up for the course (the forms are shown in Appendix Figure A.II). In particular, after asking students their name, sex and favorite subject in school, the form contained the following statement:

> "[Company Name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like."

The forms then had one of the following two options:

> "Your decision to sign up for the course will be kept completely private from everyone, except the other students in the room."

which we refer to as the "public" signup condition, or:

> "Your decision to sign up for the course will be kept completely private from everyone, including the other students in the room."

which we refer to as the "private" signup condition.

Thus, the sole difference between the two forms was a single word, "except" or "including" (in practice, we did not reveal any signup decisions). We also note that the difference in expected privacy is for classmates, as opposed to teachers, school administrators or parents.[36]

Students were not given any additional information, and were told that all questions should be held until after all forms had been collected. When all students had completed the first form, the research team collected them and handed out a second form with additional questions (discussed below; this form can be found in Appendix Figure A.III).[37] When students had completed the second form, the research team collected it and handed out assent and consent forms for authorization to access GPA information. The entire intervention took less than 10 minutes.

The forms with the different privacy assurances were pre-sorted in an alternating pattern and handed out to students consecutively in their seats.[38] By randomizing at the level of the student within the classroom, we ensure that students in the public and private signup groups were

---

[36] Though we cannot rule out that students also inferred that privacy related to these others as well.

[37] In the fourth school, we included additional questions at the end of the second form (see Figure A.III).

[38] In some classrooms, students are seated alphabetically, while in others they choose where to sit. Thus, we cannot rule out that students sitting near each other have more connection to each other than students chosen at random. However, this should not affect our estimates of the public vs. private treatments.

otherwise treated exactly the same in every other way. So for example there are no differences in how the experimenters or teachers treated students with different privacy statements, no differences in encouragement to enroll or in classroom environment or characteristics. We also did not allow students to communicate with each other until all forms were returned, so there would be no contamination across groups and so students would not realize that they were being given different terms of privacy (and even if students looked at each other's desks, the forms looked identical at a glance because they only differed by one word; see Appendix Figure A.II).

Because the difference between the two forms was just a single word, the treatment was very small and subtle. This makes it less likely that students would respond to the difference, and we will therefore likely underestimate the effects of peer pressure. We chose not to implement treatments that made signup even more explicitly public, such as by asking students to raise their hands in the class, come to the front of the room or put their name on a public signup sheet in the room. First, doing so would have required a much greater number of classrooms and schools, and thus significantly higher cost, in order to have reasonable statistical power, since treatments like this could only be implemented at the classroom level. Related, introducing variation at the classroom level could introduce more possible random variation in student, classroom or teacher attributes (or implementation of the treatment) across treatment groups that could separately influence signup. A second reason is that the method of signing up (i.e., having the public treatment involve raising a hand or staying after class to sign up and the private treatment involve signing up on an individual sheet of paper) could itself affect signup rates. Having all students signup through the same exact process but varying only a single word for the two groups makes it clearer that it was just the public vs. private nature of the decision that explains any difference in signup rates. Finally, having a more public treatment such as through raising hands or coming to the front of the room to sign up could have allowed for the other kinds of peer effects that we want to exclude, such as social learning or coordination.

As noted above, our priors (aided and confirmed by initial pilot testing) were that the social acceptability of undertaking effort or an investment could vary across settings, particularly with respect to academic performance or baseline levels of effort or investment. Therefore, we explicitly chose both honors and non-honors classes for the experiment, yielding 560 students in non-honors classes and 265 in honors classes.

Table IV presents tests of covariate balance. As expected given that randomization was among students within classrooms, the two groups are very well balanced on all measured dimensions, including sex, age, ethnicity, number of honors classes and grade point average (the first three are measured directly in our survey, the latter two are drawn from matching our data to administrative records provided by the schools).[39]

III. B. Testing the Peer Pressure Mechanism

As noted above, any differences in the response to whether the signup decision is public or private across students in honors and non-honors classes could arise for reasons other than differences in norms. For example, honors and non-honors students may differ along many social, economic and demographic dimensions, or may have different aspirations or expectations, some of which could affect how they respond to differences in whether information is private.

In order to reduce this heterogeneity and create a more comparable set of students in honors and non-honors classes, which will allow us to estimate more cleanly the effect of changing just the composition of peers to whom the signup decision is potentially revealed, we can exploit the fact that many students do not take exclusively honors or non-honors classes. In the schools in our sample, students are allowed to choose whether to take an honors or non-honors version of every subject for which both are offered. Per school policy, they cannot be denied entry into any honors class that they want to take (even if they have poor grades), nor can they be forced to take an honors class they do not want to take. Many students therefore choose to take just a few honors classes, for example choosing a subject that they are particularly interested in or a class with a teacher they like or heard good things about.[40]

We can therefore examine students taking just some, but not all, honors classes, and exploit variation in the timing of those courses relative to the timing of when our research team arrived to conduct the experiment. For any given some-honors student, whether the period when we arrived and conducted our study corresponded to one of their honors classes or one of their non-honors

[39] We were able to get information on honors classes and GPA for 94 percent of our sample. The remainder had moved to different classrooms or schools by the time we entered our data and asked for records; school counselors were unable to assist us in matching these students. Missing information does not significantly correlate with treatment. Also, accessing individual GPA data requires both child assent and parental consent, which we did not receive from 16 percent of students. Therefore, we can only provide GPA data at the group level, and cannot use it in our regressions. However, we also asked students to self-report grades on the survey handed out after the signup form was collected.

[40] Taking only a few honors classes also helps students manage their workload or keep up their GPA.

classes should be exogenous with respect to their attributes. The effects of making signup public or private in honors vs. non-honors classes for this group of students therefore more cleanly isolates how signup varies when essentially at random we offer it to them when they are sitting in the room with other honors students or other non-honors students.

In practice, this strategy involves restricting our analysis to students taking between one and three honors classes.[41] We note that in implementing this strategy, we must condition on the number of honors classes being taken so that we only compare students taking the same number. To see this, note that the full set of some-honors students we find in non-honors classes will include a greater share of students taking just one honors class relative to students taking three honors classes (setting aside differences in the size of these two groups), since the former are much more likely to be in a non-honors class during any given period. By contrast, the set of some-honors students we find in honors classes will contain a larger share of three-honors students relative to one-honors students. Since one- and three-honors students likely differ from each other in many ways, our empirical strategy relies on comparing only among those taking the same number of honors classes, who should therefore be similar, just exploiting variation in whether we happened to arrive when they were sitting in one of their honors classes or one of their non-honors classes.

One potential concern is class scheduling. For example, suppose in the extreme case we visited only one school and that honors classes for the various subjects are offered uniquely across periods, i.e., period 1 offers honors only in English, and honors English is only offered period 1. In this setting, if we arrived 1st period, the some-honors students found in an honors class will all be taking honors English, while those found in a non-honors class will be taking honors only in other subjects. If students taking honors in different subjects differ from each other, particularly in ways that affect how they respond to whether their decisions are public or private (independently of peer pressure), then we will not rule out selection. Though we have no strong priors that such students would respond differently, we believe that in practice this is not a concern for our analysis. First, because these are large schools, there are multiple honors and non-honors sections for each subject, offered during different periods throughout the day. So visiting during one particular period will not necessarily skew the some-honors students we find in an honors class towards a particular honors subject relative to some-honors students we find in a non-honors class. Further,

[41] In our sample, taking four honors classes is effectively taking all honors (only 9 students take 5 honors classes). Consistent with this, we find no four-honors students in any of the non-honors classes we visited.

we visited each school during two separate periods. Finally, we visited different schools, each of which has different schedules (and we visited different schools during different periods).

This strategy assumes that information is to an extent localized. But a some-honors student sitting in an honors class when offered the course under the public regime may worry that peers in their non-honors classes will learn that they signed up (especially since other students in the class are also taking a mix of honors and non-honors classes). This would work against finding differences based on whether these students are offered the course when with their honors or non-honors peers, and suggests we may underestimate peer pressure.[42]

Overall, there are 343 some-honors students (42% of our sample). Appendix Table A.V provides means for key covariates. We note that in columns (3) and (4) of Panel A, covariates are not balanced between those we surveyed in honors and non-honors classes (though they are balanced across the public and private treatments). As discussed above, those we find in an honors class are taking more honors classes (and have a higher average GPA) than those we find in a non-honors class. Thus, as a demonstration, Panel B shows means for the set of students taking exactly two honors classes (we focus on two-honors students because in practice, with our smaller sample sizes we find very few one-honors students in an honors class and very few three-honors students in a non-honors class, leading to small cell sizes and noisier estimates).[43] Overall, the two groups are now very similar in terms of attributes (and covariates are again balanced across public and private treatments).[44] They are also well-balanced across honors subjects; of the two total honors classes they are taking, the groups differ only by 0.08, 0.02 and 0.12 in terms of the number of math/sciences, social sciences and humanities honors classes. Though none of the differences are statistically significant, to absorb any residual variation, in separate regressions below we also add controls for attributes and honors subjects taken (this does not change the estimates appreciably).

[42] We cannot assess whether information flows across classes (or whether students believe it does). It is possible that students don't talk much about these kinds of efforts, and it is only when it is directly observed that it is relevant. There may also be a practice among some-honors students that "what happens in honors class, stays in honors class." For example, some-honors students may want to work hard and succeed in their honors classes, and may then worry that if they tell their non-honors peers what another some-honors student did in an honors class, that other student could in turn do the same to them.

[43] We find only 13 one-honors students in an honors class, both because they don't take many honors classes and because we visited fewer honors classes, and only 9 three-honors students in a non-honors class, both because they don't take as many non-honors classes and because there are fewer three-honors students.

[44] One concern is that honors classes may be smaller than non-honors classes, and peer pressure may differ by class size. However, in our sample, the difference is very small, and not statistically significant.

III. C. Empirical Strategy

We begin by regressing an indicator for whether individual *i* in classroom *c* chose to sign up for the prep course (*Signup*) on an indicator for whether they were offered the public or private treatment (*Public*), an indicator for whether the class they are in at the time of the offer was an honors or non-honors class (*Honors*) and the interaction between *Public* and *Honors*:[45]

$$Signup_{i,c} = \beta_0 + \beta_1 Honors_c + \beta_2 Public_{i,c} + \beta_3 Honors_c * Public_{i,c} + \varepsilon_{i,c}$$

where β_2 and β_3 are the coefficients of interest, namely the estimated effects of making the signup decision public in non-honors classes and the differential impact of public signup in honors relative to non-honors classes, respectively. In additional specifications, we add other covariates (age and dummies for sex and Hispanic) as well as surveyor and classroom fixed effects; the latter further isolate the within-classroom variation in the public vs. private condition across students. These results will capture the overall effects of making signup public rather than private in the two types of classes, which can carry implications for school policies and practices.

To then more cleanly test the isolated peer pressure mechanism, we estimate the same regressions while limiting the sample to some-honors students, adding dummies for the number of honors classes to ensure that we only compare students taking the same number of such classes (and in additional specifications, we add controls for attributes and honors subjects).

III. D. Main Results

We begin by providing the raw signup rates across public and private conditions, in both honors and non-honors classes. Figure II displays the findings. In non-honors classes, the private signup rate is 72%, while the public rate is 61%. The difference is significant at the 1 percent level. In honors classes, private and public signup rates are very high overall, and very similar: 92% of students sign up under the private treatment, while 93% sign up under the public one (p=0.631). These high signup rates suggest that students indeed valued the course being offered, consistent with their beliefs about the cost of the course.[46] Further, the fact that signup is not affected by privacy in the honors class shows that there is no general effect of privacy itself (such as students

[45] We present separate regressions for honors and non-honors classes in Appendix Table A.VI.

[46] Since private signup rates are already close to 100% in honors classes, it will be difficult to find any large positive effect of public signup due to data censoring.

always having a strong preference for greater privacy); though it is possible that the value placed on privacy differs between the kinds of students who are in honors and non-honors classes or that the demand for (or value of) the course is so much higher in honors classes (since more students want to go to college) that these students are willing to accept the loss of privacy in exchange for the course. We will separate out this possibility below.

In Table V, we present the results in regression format. In column (1), we present the results without controls (which replicate the sign-up rates from Figure II); in column (2) we add individual covariates and in column (3) we further add classroom and surveyor fixed effects. The results are very similar across specifications, suggesting that randomization was successful. We again conclude that making signup public rather than private reduces signup rates in non-honors classes, by a statistically significant 11−12 percentage points. But there is again no effect in honors classes. We believe these results are valuable in themselves, aside from testing for peer pressure as the driving mechanism, with important implications for school policy and practices by showing a large, negative effect of observability on investment choices in non-honors classes.

This first set of results indicates that there is not a universally negative effect of making the signup decision public. Nevertheless, this is not yet sufficient to establish the existence of different social norms in honors vs. non-honors classes, nor that students are responding to those differences. We therefore turn to our analysis of some-honors students. Having established above that there are no significant differences between such students that were offered the SAT course in an honors or a non-honors class (once we condition on the number of honors classes being taken), we can show that, by contrast, their classmates in those classes are very different. In non-honors classes, the private signup rate among no-honors classmates is 65%, while in honors classes the rate among all-honors students is 100% (the *p*-value of the difference is 0.000). There are also dramatic differences in peers' GPA (2.03 in non-honors vs. 3.54 in honors, with p=0.000). Some-honors students fall between the two, with 86% private signup rates and a 2.67 GPA.

These results establish that the peer groups are indeed very different in honors vs. non-honors classes, and in a way that helps us formulate our hypotheses on the direction of social pressure effects for students taking some-honors classes. If peer pressure pushes students towards conforming to the locally prevailing norm within the classroom, we expect public signup to be lower than private sign up in non-honors classes, and higher in non-honors classes. In Table VI, we estimate regressions using the full sample of some-honors students. For ease of presentation, the table shows results from separate regressions for honors and non-honors classes, so the results can be read from the public dummy, rather than several different interaction terms.[47] In non-honors classes, the effect of the public treatment is to reduce signup rates by a statistically significant 15−17 percentage points. In honors classes, the public treatment increases signup rates by 7−9 percentage points, with statistical significance at the 10 percent level in three of the four specifications. Viewed in a different way, when choices are public, signup rates are over 20 percentage points greater when (otherwise identical) students make them in one of their honors classes rather than one of their non-honors classes.

Of course, we cannot generalize the results for these some-honors students to all students (though the full sample results showing improved signup by making it private in non-honors classes still holds). However, it is still valuable to document a set of students for whom the localized influence of peers can have such a dramatic effect. Further, the set of some-honors students represents about 42 percent of the sample. Finally, these some-honors students may be the most relevant "marginal students"; those taking all honors classes are already making high levels of effort and investment, whereas those not taking any honors classes may require deeper interventions, or altogether different policies, in order to increase their effort.

III. E. Heterogeneity

Our main underlying hypothesis for why peer observability may affect choices is that students worry about what their peers will think of them. On a second form handed out to students after they had turned in the signup form, we asked students how important they thought it was to be popular in their school, on an increasing scale of 1 to 5.[48] Though these are of course just subjective, self-reports, they can provide suggestive corroborating evidence of our proposed mechanism. If the effects that we observe are driven by fear of social sanctions, or seeking social approval, we would expect students who are more concerned with popularity to be more responsive to whether signup is public or private. To assess this hypothesis, we split our sample as close as

[47] Appendix Figure A.IV shows the raw signup rates for two-honors students (again, for any presentation of means we must compare for a specific number of honors classes, and cell sizes for one- and three-honors students are small so the means are noisy). The figure reveals that the public treatment decreases signup in non-honors classes dramatically, while increasing it in honors classes.

[48] The question was: "On a scale 1-5, how important do think it is to be popular in your school? (1: not important…5: very important)."

possible to half, according to the importance attributed by students to being popular (answers 1 and 2 (not important) vs. 3, 4 or 5 (important)). Figures III and IV present the results for the raw signup rates.[49] Figure III shows that for students in non-honors classes who say that it is important to be popular, the signup rate is 20 percentage points lower in the public condition than in the private condition (p=0.002). For those who care less about popularity, the effect of a public decision is small (4 percentage points) and not statistically significant (p=0.427). In Figure IV, we observe the opposite pattern for honors classes, although on a smaller scale (since the private take up rates were already very close to 100%): a positive effect of public sign up for those who care more about popularity, and no difference for those who care less. Table VII presents the results in regression format, which confirm these results. Thus, we find that students who believe it is important to be popular move in the direction of locally prevailing norms (in both honors and non-honors classes) when signup is public rather than private, while those who do not think it is important are unaffected by whether signup is public or private.

Appendix Table A.VII explores heterogeneity by gender for the full sample. Male students are less likely to sign up when the decision is private than female students are (significantly so in non-honors classes), and the interaction of the public condition with the male indicator is always negative (although never statistically significantly so). These results suggest that boys may be somewhat more concerned about publicly displaying effort in school, but we look at these findings with caution, given the small size of the effects and the lack of statistical significance.

III. F. Account Login Data

Our main objective is to test for peer pressure, for which the signup decision is the relevant outcome. However, we also obtained data on whether students actually logged into the online system later to activate their accounts (data on intensity of usage are not available). It is worth emphasizing that in analyzing this outcome, we lose experimental control since students in the public and private treatments are likely to have communicated or coordinated with each other after our team left the classroom. In doing so, they may have changed their beliefs about whether others would learn about their decision. Such communication also provides scope for other forms of peer

[49] The results and those below return to the full sample, since stratifying by popularity leads to extremely small cell sizes (popularity*honors/non-honors*public/private) for the subset of some-honors students. The results for the some-honors students do show the same qualitative pattern, but are less precisely estimated.

effects beyond peer pressure, such as social learning or consumption externalities. So the estimates from this analysis are not as useful for testing our hypothesis. Further, our analysis was designed to detect effects on signup rates, and we may therefore be underpowered to detect subsequent account login rates. However, activating the account is a useful policy outcome, indicating how much you can actually change adoption of an investment just by varying whether it is public or private. Examining this outcome can also help establish that signing up for the course was not just cheap talk, i.e., whether students at signup actually intended to follow through and use the course.

Overall, 81 percent of students who signed up for the course logged in to activate their account, which is a fairly high "follow-through" rate and again confirms that students indeed valued the course. Overall, the unconditional mean take-up (login, conditional on being offered the course) is 61 percent. This is broadly similar to the 51 percent of students in our sample schools who take the SAT.[50] Students in honors classes had a slightly higher follow-through rate (78% vs. 84%), though the difference is not statistically significant.

The results are shown in Appendix Figure A.V (the conclusions from regressions are similar). For the full sample of students, we find that in non-honors classes, making the course public reduces the rate of logging in to use the system by 8.2 percentage points (from a base of 57 percent when signup is private; *p*-value=0.051). In honors classes, as with the signup decision, there is no difference in login rates between public and private treatments (77% for private, 78% for public).[51] We also note that the follow-through rates did not differ across any of the (honors/non-honors)×(public/private) groups.[52]

III. G. Other Concerns and Interpretations

One concern for external validity is that students may not have valued the course greatly (e.g., they believed the course was not very good, or they were already taking another course),[53]

[50]The rate here is slightly higher, but there may be students in 11th grade who still think they would like to go to college, but who ultimately do not (because of performance, finances or other factors).

[51] For two-honors students (Panel B), the results are similar to those for signup, but less precisely estimated.

[52] The follow-through rate for some-honors students in honors classes is 81% under the public treatment and 82% under the private treatment. Thus, the positive peer pressure effect observed above (increases in signup rates under the public treatment for some-honors students in honors classes) is unlikely to be just cheap talk, since they are just as likely to follow-through and actually login and activate their account.

[53] A related possibility is that students may have thought that they would have another chance to sign up later. However, we believe this is unlikely to account for our results. First, even if students believed they would have another chance, they would have to further believe that the later opportunity would differ on

and perhaps in settings with higher stakes, students are less affected by peer pressure. However, we note that signup is extremely high when privacy from classmates was ensured. Further, as noted above, follow-through rates for activation were very high. Finally, students estimated the cost at $260, which is a high cost for these low-income students. Though of course it remains possible that for many students, the true value of this course was low.

With some investments that students may make in school, there is also the possibility that undertaking such efforts reveals low ability, such as the need for extra help or assistance. Of course, this is just one possible form of peer social concerns or pressure, or a micro foundation for such behavior, and thus does not challenge our results. However, we believe that such effects are unlikely to underlie our results. SAT preparation, whether through books or classes, is very common, and not often associated with representing low ability. In our survey, students reported that they believed that on average about 43% (64% in honors classes) of their classmates were taking some other course to prepare for the SAT. Further, honors students in our sample had very high signup rates (over 90%), suggesting that this is not a course only for the worst students.

Alternatively, students may not want to undertake efforts if final outcomes are also observable, such as due to a "fear of failure": students who believe they have a high likelihood of failure on some observable outcome (such as getting into a good college, or any college at all), may choose not to undertake effort (or even actively signal that they are not putting in effort) so that if they fail, others will believe it was because they did not try, rather than that they tried and still failed. Again, we believe the asymmetric response to the public treatment makes this alternative less likely, since we would then need the effects to go different ways in different classes (i.e., some-honors students have a fear of failure in their non-honors classes, but the reverse of the fear of failure when in their honors classes).[54]

A final issue to consider is whether the effects are due to consumption externalities. Having more peers take the course (as might be expected in honors classes) may make the course more

privacy. Second, since we concluded the study, no students who had not signed up communicated to our team (students took away forms with our contact information) or their teachers that they were interested in the course. Finally, we asked students from the last school we visited (after signup was complete) whether they believed they would have another chance to sign up, and 85% said no. This may even overstate the extent of such beliefs, since the act of asking the question may suggest or elicit that belief.

[54]Though fear of failure effects could differ across settings. For example, students may fear failure more around non-honors peers, who might mock them for even trying. On the other hand, fewer of their non-honors peers will be going to good colleges, or to college at all, so failing is not as stark a contrast as it might be compared to their honors peers.

valuable because students can study together or learn from each other. The reverse would hold in non-honors classes, where fewer peers are likely to take it. Though we cannot completely rule out this possibility, we believe it is unlikely to drive our results. Consider the some-honors students. If they believed that students in all classes would also be offered the prep course, then the full set of their friends who will be offered and take up the course, and thus the expected consumption externalities, should not differ based on whether they are sitting with their honors or non-honors peers when they are offered the course.[55] If these students instead believed that the course was only being offered to those in the class with them at that time, then under the private condition we should expect higher signup rates for those sitting in an honors class than for those sitting in a non-honors class (since they should expect more of their honors class peers to take it). However, as noted above, these private signup rates do not differ significantly. Thus, though there may be consumption externalities, students do not appear to act as though there are when they make their private signup decisions. In addition, we note that though consumption externalities on their own could explain a difference in signup rates in honors and non-honors classes, it is less clear that it should affect differential signup within each class based on whether signup is public or private. However, we cannot rule out that beliefs about consumption externalities could differ within each class based on whether a student was in the public or private signup regime. This could arise if students themselves share our hypothesis; in other words, students given the public signup sheet in an honors class believe more of their classmates will signup than students given the private signup sheet (and the reverse in non-honors classes).[56]

IV. CONCLUSION

We find evidence that student effort and investments are highly responsive to concerns about peer observability using both a natural experiment that introduced a leaderboard that revealed top performers, and a field experiment varying whether the decision to sign up for an

[55] Though beliefs could differ by class type; some-honors students in honors classes may believe the course is only being offered to honors classes, while those in non-honors classes may believe it is being offered to all students, or only those in non-honors classes.

[56] Though this will again depend on beliefs about whether the course was offered to all classes. A some-honors student in an honors class who gets the public signup may believe that more of their peers will sign up; but they may also think that same condition will reduce the number of peers that will sign up in their non-honors classes (though they may be more likely to study with friends in their honors classes). So beliefs about the difference in the number of friends that will take the course may be ambiguous.

SAT prep course would be revealed to classmates. We also find evidence suggesting that the results are specifically driven by concerns over popularity and the possibility of facing social sanctions or gaining social approval depending on effort or investments, or at least, a desire to conform to prevailing social norms among peers in the classroom. The results have important implications for understanding the nature and impact of peer interactions in the classroom more generally.

Though we are unable with our data to link these changes in behavior to longer run education or labor market outcomes, the fact that we find similar results in two different settings suggests that such effects may be more widespread.

REFERENCES

Akerlof, George A. (1997). "Social Distance and Social Decisions," *Econometrica*, 65(5), p. 1005-1027.

Ashraf, Nava, Oriana Bandiera and Scott Lee (2014). "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior and Organization*, 100, p. 44-63.

Austen-Smith, David and Roland Fryer (2005). "An Economic Analysis of 'Acting White,'" *Quarterly Journal of Economics*, 120(2), p. 551-583.

Ausubel, David (1977). Maori Youth: A Psychoethnological Study of Cultural Deprivation. Massachusetts: The Christopher Publishing House.

Bandiera, Oriana, Iwan Barankay and Imran Rasul (2010). "Social Incentives in the Workplace," *Review of Economic Studies*, 77 (2), p. 417 - 458.

Barankay, Iwan (2012). "Rank Incentives: Evidence from a Randomized Workplace Experiment," mimeo, University of Pennsylvania.

Becker, Gary, and Kevin M. Murphy (2000). "Social Forces, Preferences and Complementarity," in Social Markets and the Social Economy, Cambridge: Harvard University Press.

Bénabou, Roland, and Jean Tirole (2002). "Self-confidence and Personal Motivation," *Quarterly Journal of Economics*, 117(3), p. 871–915.

- - and - - (2003). "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70(3), p. 489–520.

- - and - - (2006). "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), p. 1652-1678.

Berman, Eli (2000). "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, 115(3), p. 905-953.

Brown, B. Bradford (1982). "The Extent and Effects of Peer Pressure among High School Students: A Retrospective Analysis," *Journal of Youth and Adolescence*, 11, p. 121–133.

- -, Donna Rae Clasen, and Sue Ann Eicher (1986). "Perceptions of Peer Pressure, Peer Conformity Dispositions, and Self-reported Behavior among Adolescents," *Developmental Psychology*, 22, p. 521–530.

- - (2004). "Adolescents' Relationships with Peers," in Richard M. Lerner and Laurence Steinberg, eds., Handbook of Adolescent Psychology, p. 363–394. New York: Wiley.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman (2014). "Understanding Mechanisms Underlying Peer Effects," *Econometrica*, 82(4), 1273-1301.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2012). "Social Networks and the Decision to Insure," University of Michigan Working Paper.

Camerer, Colin F. and Robin M. Hogarth (1999). "The Effects of Financial Incentives in Experiments: A Review and Capital-labor-production Framework," *Journal of Risk and Uncertainty*, 19, p. 7–42.

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009). "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27 (3), 439-464.

- -, Bruce I. Sacerdote, and James E. West (2013). "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects," *Econometrica*, 81(3), p. 855-882.

Coleman, James (1961). The Adolescent Society: The Social Life of the Teenager and Its Impact on Education. Glencoe, Illinois: Free Press.

Dobbie, Will and Roland G. Fryer, Jr. (2011). "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics,* 3(3), p. 158-87.

Duflo, Esther, Pascaline Dupas and Michael Kremer (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101 (5), 1739–1774.

Durlauf, Steven N. and Yannis M. Ioannides (2010). "*Social Interactions*," *Annual Review of Economics*, 2(1), p. 451-478.

Epple, Dennis, and Richard Romano (2011). "Peer Effects in Education: A Survey of the Theory and Evidence," in Jess Benhabib, Alberto Bisin and Matthew O. Jackson, eds., Handbook of Social Economics, Vol. 1B, Elsevier/North-Holland: Amsterdam.

Fordham, Signithia and John U. Ogbu (1986). "Black Students' School Success: Coping with the 'Burden of Acting White'," *The Urban Review*, 18, p. 176-206.

Fryer, Roland (2007). "A Model of Social Interactions and Endogenous Poverty Traps," *Rationality and Society*, 19(3), p. 335-366.

- - (2011). "The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap," in Jess Benhabib, Alberto Bisin and Matthew O. Jackson, eds., Handbook of Social Economics, Vol. 1B, Elsevier/North-Holland: Amsterdam.

- - and Lawrence F. Katz (2013). "Achieving Escape Velocity: Neighborhood and School Interventions to Reduce Persistent Inequality," *American Economic Review*, 103(3), p. 232-237.

- - and Paul Torelli (2010). "An Empirical Analysis of 'Acting White,'" *Journal of Public Economics,* 94(5-6), p. 380-96.

Gans, Herbert (1962). The Urban Villagers: Group and Class in the Life of Italian-Americans. New York: The Free Press.

Jacob, Brian A. (2004). "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago," *American Economic Review*, 94(1), p. 233-58.

Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz (2007). "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1), p. 83–119.

Kremer, Michael, Edward Miguel and Rebecca Thornton (2009). "Incentives to Learn," *Review of Economics and Statistics,* 91, p. 437-456.

Lee, Kenneth W. and William G. Warren (1991). "Alternative Education: Lessons from Gypsy Thought and Practice," *British Journal of Educational Studies*, 39 (3), p. 311-324.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. (2010). "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," *Journal of the European Economic Association*, 8(6), p. 1243–1265.

Ligon, Ethan (1998). "Risk-Sharing and Information in Village Economies," *Review of Economic Studies* 65(4), p. 847-864.

Magruder, Jeremy (2010). "Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa," *American Economic Journal: Applied Economics*, 2(1), p. 62-85.

Mas, Alexandre and Enrico Moretti (2009). "Peers at Work," *American Economic Review*, 99(1), 112–145.

Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, 118(2), p. 549-597.

- - (2014). "Community Networks and the Process of Development," *Journal of Economic Perspectives*, 28(4), p. 49-76.

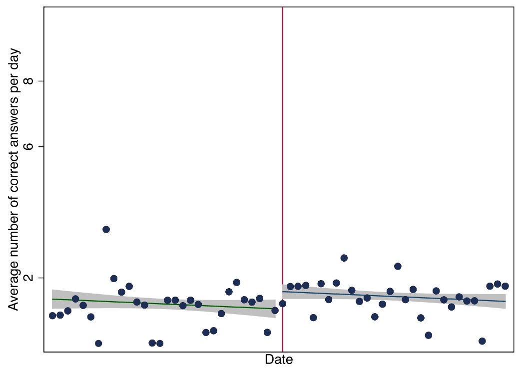
Oreopoulos, Philip (2003). "The Long-Run Consequences of Living in a Poor Neighborhood," *Quarterly Journal of Economics*, 118(4), p. 1533-75.

Postlewaite, Andrew (2011). "Social Norms and Social Assets," *Annual Review of Economics*, 3, p. 239-

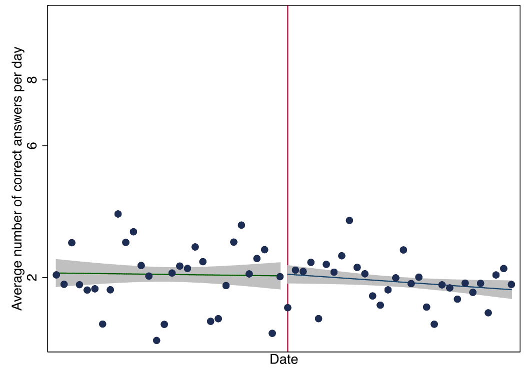
259.
Sacerdote, Bruce (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” *Quarterly Journal of Economics*, 116 (2), 681–704.
Santor, Darcy A., Deanna Messervey and Vivek Kusumakar (2000). “Measuring Peer Pressure, Popularity, and Conformity in Adolescent Boys and Girls: Predicting School Performance, Sexual Attitudes, and Substance Abuse,” *Journal of Youth and Adolescence*, 29, 163–182.
Topa, Giorgio (2001). "Social Interactions, Local Spillovers and Unemployment,” *Review of Economic Studies*, 68 (2), p. 261-295.
Townsend, Robert M. (1994). “Risk and Insurance in Village India.” *Econometrica*, 62(3), p. 539-591.
Tran, Ahn and Richard Zeckhauser (2012). “Rank as an Inherent Incentive: Evidence from a Field Experiment,” *Journal of Public Economics*, 96, p. 645–650.
Wang, Shing-Yi (2013). “Marriage Networks, Nepotism and Labor Market Outcomes in China,” *American Economic Journal: Applied Economics*, 5(3): 91-112.
Zimmerman, David J. (2003). “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment,” *Review of Economics and Statistics*, 85 (1), 9–23.

FIGURE I: AVERAGE NUMBER OF CORRECT ANSWERS PER DAY

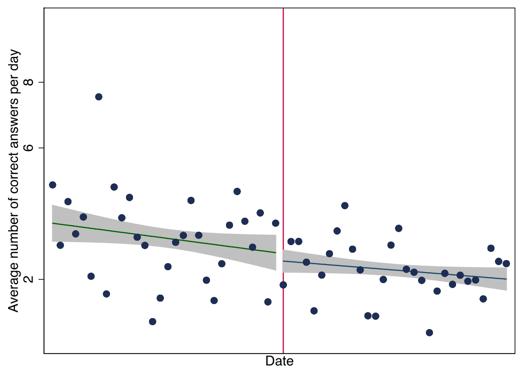
QUARTILE 1



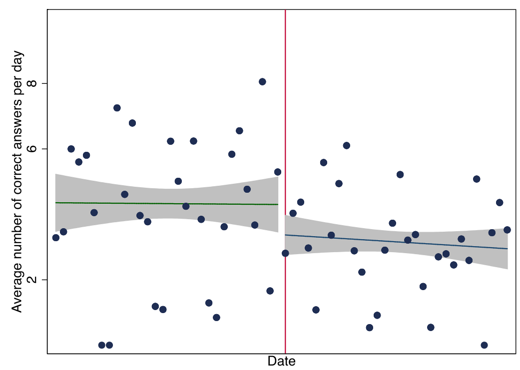
QUARTILE 2



QUARTILE 3

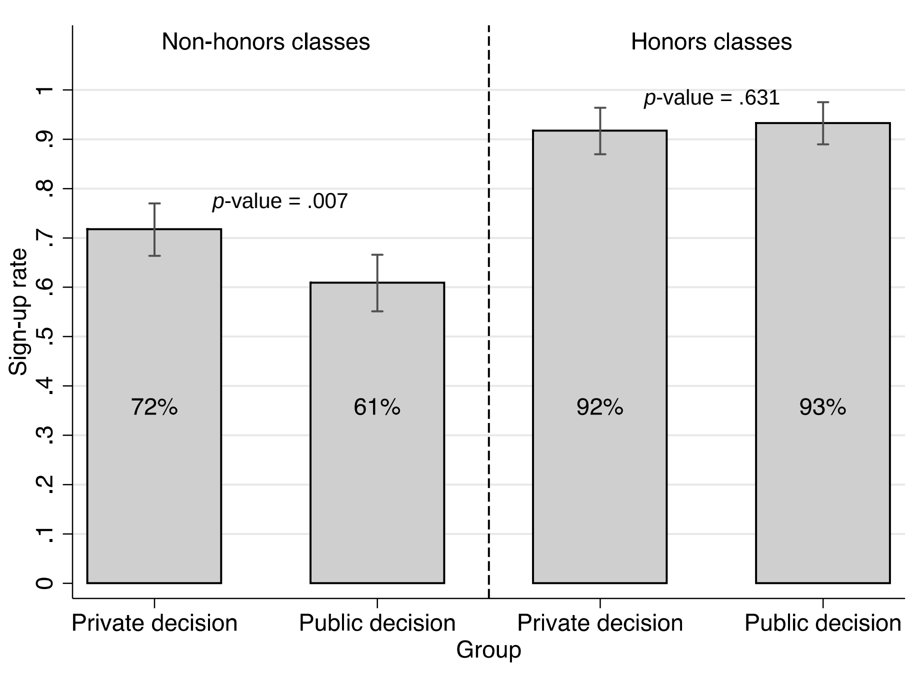


QUARTILE 4



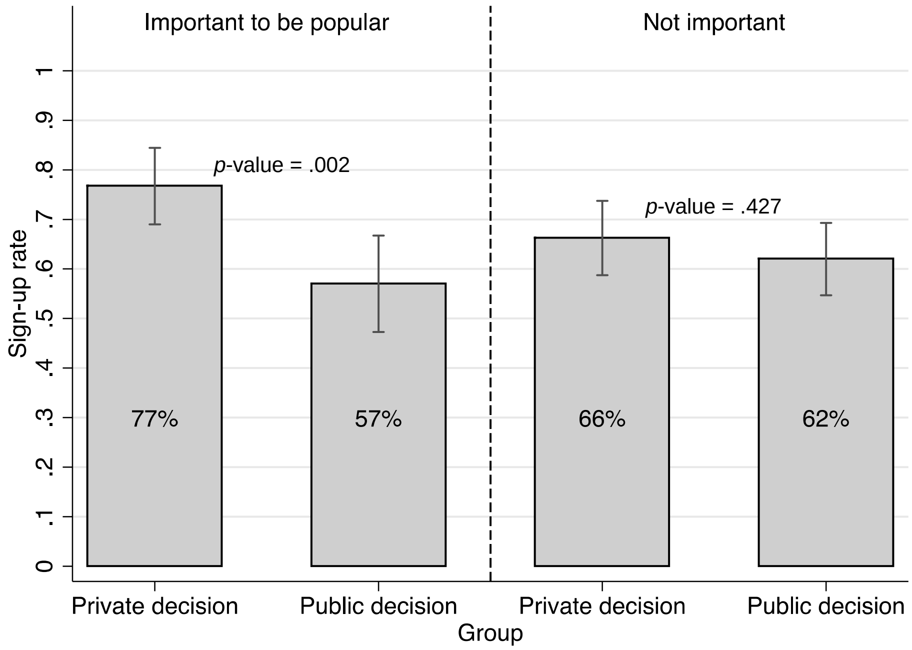
Notes: These figures plot, for each day in the period from 30 days before to 30 days after the introduction of the new system, the average number of correct answers per day. Each figure plots a different quartile of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. There are 60 observations plotted per quartile. The figures also fit linear trends separately before and after the introduction of the new system, and the 95% confidence interval associated with the trends. The vertical line corresponds to the day of the introduction of the new system, September 20, 2011.

FIGURE II: SIGNUP RATES FOR PRIVATE VS. PUBLIC DECISIONS, NON-HONORS VS. HONORS CLASSES



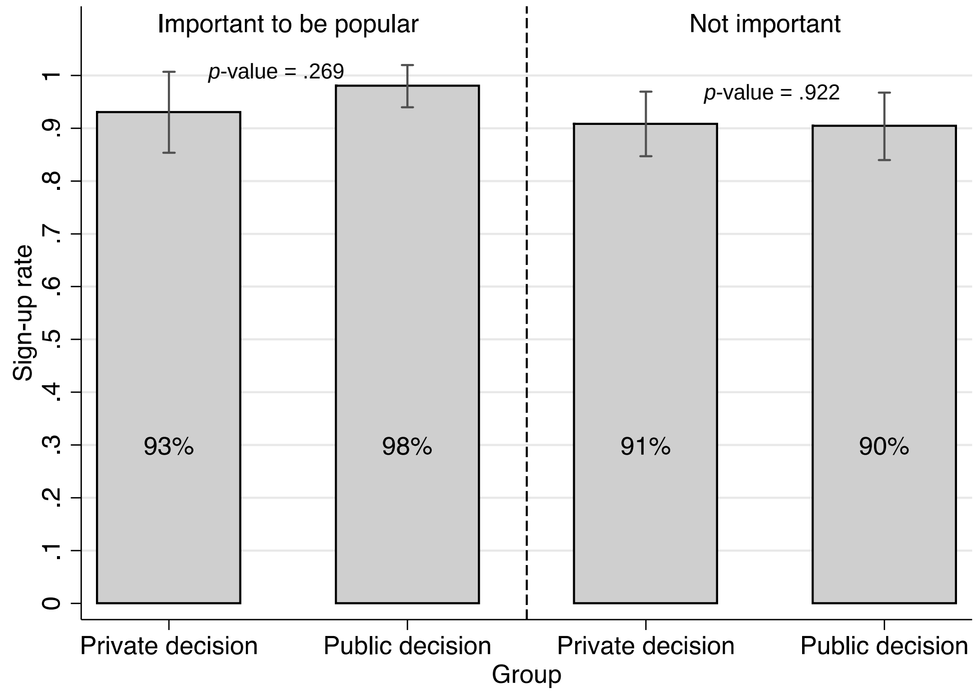
Notes: This figure presents the means and 95% confidence intervals of the signup rates for students in the private and public conditions, separately for honors and non-honors classes. There are 560 observations for non-honors classes and 265 for honors classes.

FIGURE III: SIGNUP RATES FOR PRIVATE VS. PUBLIC DECISIONS: IMPORTANCE OF BEING POPULAR (NON-HONORS CLASSES)



Notes: This figure presents the means and 95% confidence intervals of the signup rates for students in the private and public conditions in non-honors classes, separately for students who consider important to be popular in their school and those who do not. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, "How important is it to be popular in your school?" from a 1-5 scale to a dummy variable (answers 3-5 were coded as considering it important, 1-2 as not important). There are 216 observations in the “important to be popular” panel and classes and 325 in the “not important” panel.

FIGURE IV: SIGNUP RATES: PRIVATE VS. PUBLIC DECISIONS: IMPORTANCE OF BEING POPULAR (HONORS CLASSES)



Notes: This figure presents the means and 95% confidence intervals of the signup rates for students in the private and public conditions in honors classes, separately for students who consider important to be popular in their school and those who do not. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, "How important is it to be popular in your school?" from a 1-5 scale to a dummy variable (answers 3-5 were coded as considering it important, 1-2 as not important). There are 92 observations in the “important to be popular” panel and classes and 170 in the “not important” panel.

TABLE I: EFFECTS OF THE POINTS AND LEADERBOARD SYSTEM

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.6266***	0.1717***	-0.5391***	-1.2486***	-1.9340***
	[0.035]	[0.037]	[0.066]	[0.086]	[0.128]
Constant	2.7076***	1.3440***	2.5072***	3.6633***	5.1408***
	[0.028]	[0.026]	[0.052]	[0.070]	[0.106]
Observations	95,342	37,171	22,978	20,427	14,766
R-squared	0.185	0.161	0.159	0.173	0.174
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. All columns restrict the analysis to the time window between one month before the introduction and one month after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE II: EFFECTS OF POINTS AND LEADERBOARD SYSTEM:
PERCENT OF ANSWERS CORRECT PER DAY

Dependent variable	% of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	0.0114***	0.0609***	0.0034	-0.0220***	-0.0213***
	[0.003]	[0.006]	[0.007]	[0.006]	[0.007]
Constant	0.5132***	0.4132***	0.5070***	0.5652***	0.6373***
	[0.002]	[0.004]	[0.005]	[0.004]	[0.005]
Observations	30,186	10,383	7,167	7,216	5,420
R-squared	0.423	0.395	0.385	0.407	0.409
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the percentage of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. All columns restrict the analysis to the time window between one month before the introduction and one month after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE III: EFFECTS OF POINTS AND LEADERBOARD SYSTEM:
QUESTIONS ANSWERED OUTSIDE OF SCHOOL HOURS

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.6608***	0.2938	-0.6713*	-0.8853*	-1.6600***
	[0.200]	[0.218]	[0.385]	[0.455]	[0.532]
Constant	6.7561***	4.5241***	6.4438***	7.5475***	9.8540***
	[0.179]	[0.199]	[0.344]	[0.407]	[0.471]
Observations	19,135	6,630	4,477	4,463	3,565
R-squared	0.458	0.493	0.431	0.421	0.425
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. In all columns, we drop questions answered on weekdays between 7am and 4pm. All columns restrict the analysis to the time window between one month before the introduction and one month after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE IV: BALANCE OF COVARIATES

	Private condition	Public condition	*p*-value
	(1)	(2)	(3)
Male	0.506	0.518	0.704
	[0.501]	[0.500]	
Age	16.74	16.75	0.851
	[0.535]	[0.489]	
Hispanic	0.96	0.959	0.899
	[0.196]	[0.2]	
# of honors classes	1.351	1.367	0.88
	[1.486]	[1.477]	
GPA	2.52	2.48	0.546
	[0.894]	[0.856]	
Observations	411	414	

Notes: Columns 1 and 2 report the mean level of each variable, with standard deviations in brackets, for the private and public conditions. Column 3 reports the *p*-value for the test that the means are equal in the two conditions.

TABLE V: EFFECT OF PUBLIC TREATMENT ON SIGNUP DECISION

Dependent variable:	Dummy: Student signed up for the SAT prep course		
	(1)	(2)	(3)
Public treatment	-0.1083***	-0.1194***	-0.1229***
	[0.040]	[0.040]	[0.040]
Honors dummy	0.1998***	0.1718***	
	[0.036]	[0.037]	
Public*Honors	0.1240**	0.1334***	0.1363***
	[0.051]	[0.052]	[0.051]
Mean of private sign-up in non-honors classes		0.717	
Includes individual covariates	No	Yes	Yes
Includes classroom and surveyor FE	No	No	Yes
Observations	825	789	789
R-squared	0.090	0.117	0.180

Notes: Column 1 presents OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign up dummy, an honors class dummy and the interaction of the two. Column 2 replicates column 1 adding individual covariates (age and dummies for male and Hispanic). Column 3 replicates column 2 adding surveyor and classroom fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE VI: EFFECT OF PUBLIC TREATMENT ON SIGNUP DECISION:
HONORS AND NON-HONORS CLASSES SEPARATELY FOR STUDENTS TAKING 1-3 HONORS CLASSES

Dependent variable:	Dummy: Student signed up for the SAT prep course							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public treatment	-0.1673***	-0.1486**	-0.1465**	-0.1467**	0.0850*	0.0729	0.0834*	0.0887*
	[0.061]	[0.061]	[0.063]	[0.064]	[0.047]	[0.046]	[0.045]	[0.048]
Mean of private take-up	0.85				0.87			
Includes individual covariates	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Includes classroom and surveyor FE	No	No	Yes	Yes	No	No	Yes	Yes
Includes honors subjects variables	No	No	No	Yes	No	No	No	Yes
Observations	184	176	176	176	159	155	155	155
R-squared	0.074	0.149	0.269	0.269	0.046	0.086	0.201	0.205
Sample:	Non-honors classes				Honors classes			

Notes: This table restricts the sample to students taking one, two or three honors classes. Columns 1 to 4 restrict the sample to non-honors classes, and columns 5 to 8 restrict to honors classes. Columns 1 and 5 present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign up dummy, controlling for dummies on the number of honors classes taken by the student. Columns 2 and 6 replicate columns 1 and 4 adding individual covariates (age and dummies for male and Hispanic). Columns 3 and 7 replicate columns 2 and 5 adding surveyor and classroom fixed effects. Robust standard errors in brackets. Columns 4 and 8 replicate column 3 and 7 adding controls for the number of honors classes taken by subject categories (math/sciences and social sciences - the omitted category is humanities). *** p<0.01, ** p<0.05, * p<0.1.

TABLE VII: EFFECT OF PUBLIC TREATMENT ON SIGNUP DECISION: BY IMPORTANCE OF POPULARITY

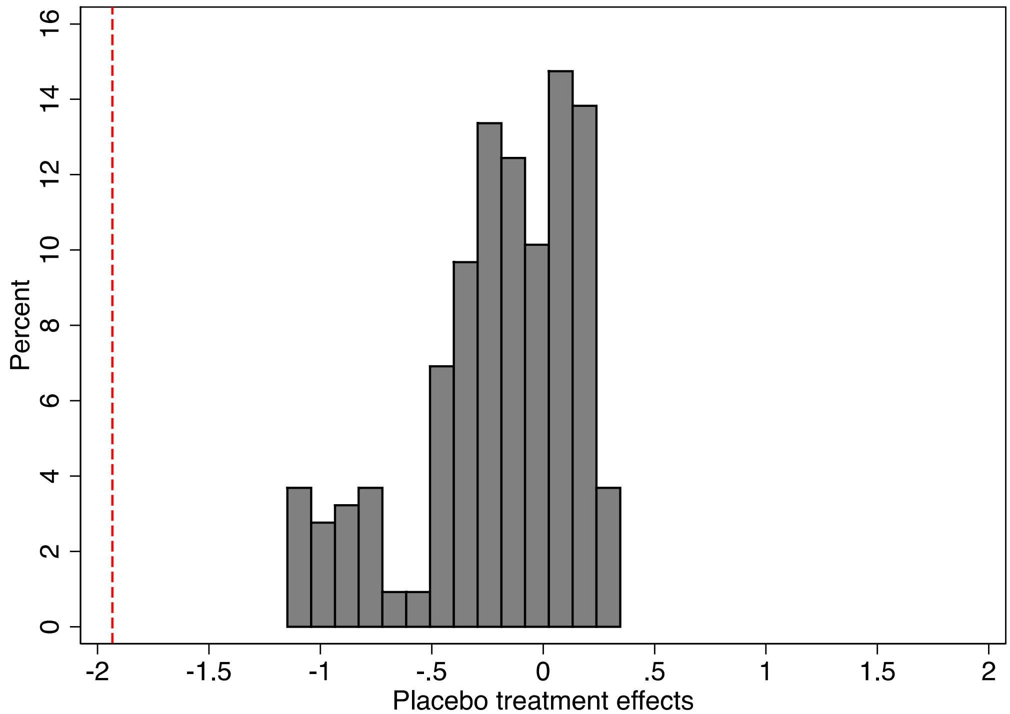
Dependent variable:	Dummy: Student signed up for the SAT prep course					
	(1)	(2)	(3)	(4)	(5)	(6)
Public treatment	-0.0425	-0.0518	-0.0483	-0.0044	-0.0220	-0.0215
	[0.053]	[0.054]	[0.054]	[0.045]	[0.043]	[0.043]
Important to be popular dummy	0.1049*	0.1347**	0.1480***	0.0222	0.0113	0.0084
	[0.055]	[0.055]	[0.055]	[0.050]	[0.053]	[0.051]
Public*Important to be popular	-0.1548*	-0.1487*	-0.1672**	0.0538	0.0828	0.0820
	[0.083]	[0.083]	[0.083]	[0.063]	[0.063]	[0.066]
Mean of private sign-up for students who do not find it important to be popular		0.662			0.908	
Includes individual covariates	No	Yes	Yes	No	Yes	Yes
Includes classroom and surveyor FE	No	No	Yes	No	No	Yes
Observations	541	521	521	262	256	256
R-squared	0.020	0.053	0.118	0.011	0.051	0.152
SAMPLE		Non-honors classes			Honors classes	

Notes: The first three columns of this table restrict the sample to non-honors classes, while the last three focus on honors classes. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, “How important is it to be popular in your school?” from a 1-5 scale to a dummy variable (answers 3-5 were coded as considering it important, 1-2 as not important). Columns 1 and 4 present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign up dummy, a dummy on whether the student consider it important to be popular in his/her school and the interaction of the two. Columns 2 and 5 replicate columns 1 and 4 adding individual covariates (age and dummies for male and Hispanic). Columns 3 and 6 replicate columns 2 and 5 adding surveyor and classroom fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

ONLINE APPENDIX

NOT TO BE PUBLISHED

FIGURE A.I:
DISTRIBUTION OF PLACEBO TREATMENT EFFECTS FOR QUARTILE 4



Notes: This histogram displays the distribution of placebo treatment effects estimated for quartile 4 of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. We run the same regressions as in our main specification, but assign the introduction of the point and leaderboard system to every other date, starting one month after the true date of the change, and ending one month before the end of the school year; there are 218 such days plotted here (each of these regressions has 14,766 observations). The dashed line represents our estimated treatment effect for quartile 4 (–1.93).

FIGURE A.II SIGNUP SHEETS

A. "Public" Signup Sheet

Student Questionnaire

First name:______________________________

Last name:______________________________

Gender (please circle one): Female / Male

What is your favorite subject in school? (Please circle one)

a. Math b. English Language Arts c. History/Social Studies d. PE/Elective

[Company name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like.

Your decision to sign up for the course will be kept completely private from everyone, except the other students in the room.

Would you like to sign up for the free [Company name] course? (Please pick one option)

Yes / No

If yes, please provide the following contact information:

Email address: ________________________________

Phone number: (_____)__________________________

TURN OVER FORM AND WAIT PATIENTLY

Form A337

B. "Private" Signup Sheet

Student Questionnaire

First name:______________________________

Last name:______________________________

Gender (please circle one): Female / Male

What is your favorite subject in school? (Please circle one)

a. Math b. English Language Arts c. History/Social Studies d. PE/Elective

[Company name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like.

Your decision to sign up for the course will be kept completely private from everyone, including the other students in the room.

Would you like to sign up for the free [Company name] course? (Please pick one option)

Yes / No

If yes, please provide the following contact information:

Email address: ________________________________

Phone number: (_____)__________________________

TURN OVER FORM AND WAIT PATIENTLY

Form A338

FIGURE A.III SECOND FORM

Student Questionnaire (2)

First name:______________________________

Last name:______________________________

Gender (please circle one): Female / Male

Ethnicity (please circle one):

a. White b. Black c. Hispanic d. Asian e. Other

Do you plan to attend college after high school? (Please choose one option)

a. Yes, four-year college
b. Yes, two-year college/community college
c. No
d. Don't know

In general, how are your grades? (Please choose one option)

a. Mostly A's
b. Mostly A's and B's
c. Mostly B's and C's
d. Mostly C's and D's
e. Mostly D's and F's

On a scale 1-5, how important do think it is to be popular in your school?
(1: not important … 5: very important)
1 2 3 4 5

On a scale 1-5, how popular would you say you are in your school?
(1: not popular … 5: very popular)
1 2 3 4 5

Hypothetically, which would you prefer? (Please circle one)

a. 50 dollars now
b. 75 dollars in six months

On a scale 1-5, how often do you think about your life when you are 40 years old?
(1: never … 5: very often)
1 2 3 4 5

Do you ever skip/ditch school with your friends?

a. Sometimes
b. Never

Do most of your closest friends plan to graduate and go to a good college?

a. Yes
b. No

Which of the following defines you the best?

a. I do what my friends do
b. I do things my own way

(FIGURE CONTINUES ON NEXT PAGE)

FIGURE A.III SECOND FORM (CONTINUED) (ONLY USED IN THE FOURTH SCHOOL)

How much do you think is the regular price of the SAT prep course that was just offered to you free of charge? ____________ dollars.

When you made your choice on whether to sign up for the SAT prep course, did you expect you might have another chance to sign up in the future? (Please pick one option)

a. Yes
b. No

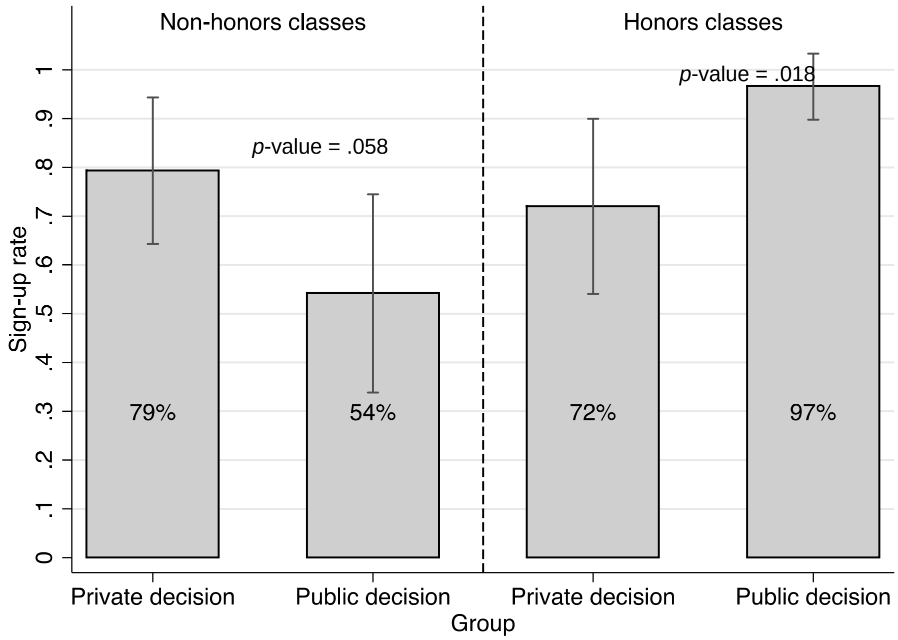
What % of your classmates do you think have already taken or plan to take an SAT prep course other than the one we offered today? ______%

Have you been listed as a Gifted/Talented student in your school? (Please pick one option)

a. Yes
b. No
c. Don't know

TURN OVER FORM AND WAIT PATIENTLY

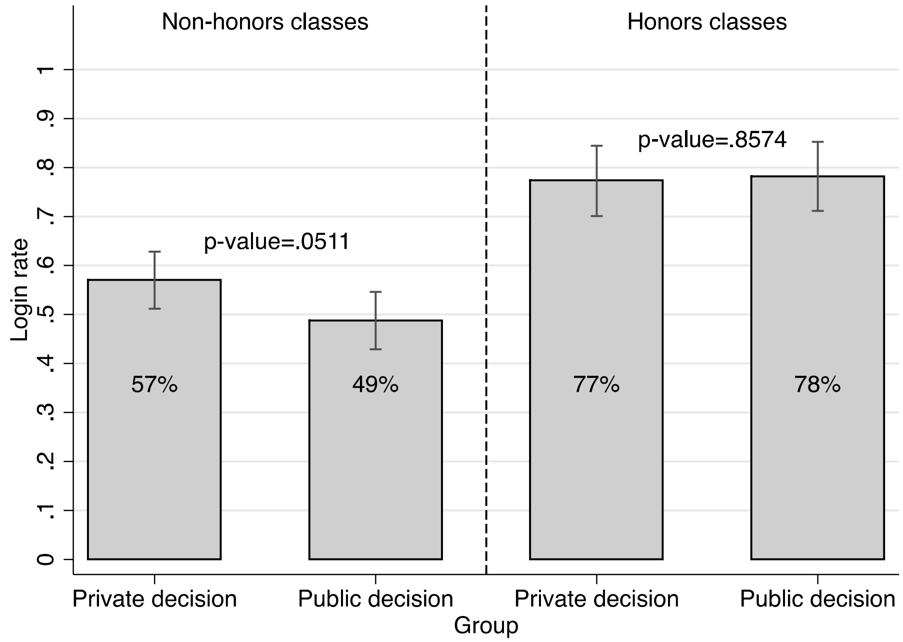
FIGURE A.IV: SIGNUP RATES FOR PRIVATE VS. PUBLIC DECISIONS, NON-HONORS VS. HONORS CLASSES: STUDENTS TAKING TWO HONORS CLASSES



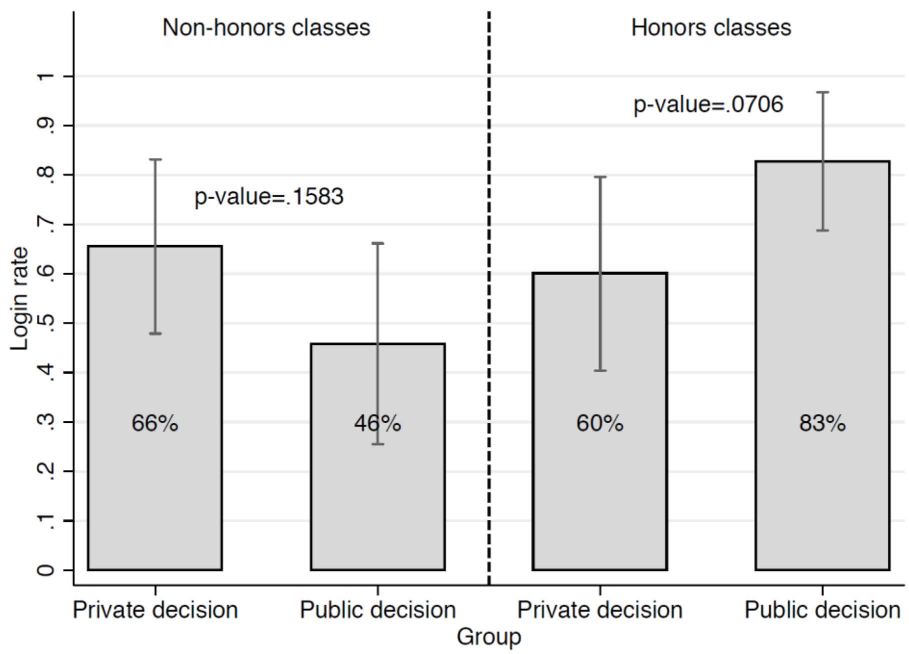
Notes: This figure presents the means and 95% confidence intervals of the signup rates for students in the private and public conditions, separately for honors and non-honors classes. In Panel A, there are 184 observations for non-honors classes and 159 observations for honors classes. For Panel B, there are 53 observations for non-honors classes and 54 for honors classes.

FIGURE A.V: LOGIN RATES

A. FULL SAMPLE



B. TWO-HONORS STUDENTS



Notes: These figures present means and 95% confidence intervals of the login rates under the private and public conditions, separately for honors and non-honors classes for the full sample (A) and the sample of 2-honors students (B). In Panel A, there are 560 observations for non-honors classes and 265 for honors classes. In Panel B, there are 53 observations for non-honors classes and 54 for honors classes.

TABLE A.I: EFFECTS OF POINTS AND LEADERBOARD SYSTEM: ONE WEEK BEFORE VS. AFTER

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.5316***	0.2393***	-0.5036***	-1.0601***	-1.7972***
	[0.060]	[0.069]	[0.116]	[0.146]	[0.212]
Constant	2.7182***	1.3279***	2.5262***	3.6637***	5.2128***
	[0.043]	[0.041]	[0.085]	[0.108]	[0.157]
Observations	30,296	11,779	7,350	6,497	4,670
R-squared	0.273	0.244	0.214	0.241	0.295
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. All columns restrict the analysis to the time window between one week before the introduction and one week after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.II: EFFECTS OF POINTS AND LEADERBOARD SYSTEM: TWO WEEKS BEFORE VS. AFTER

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.5172***	0.2497***	-0.4082***	-1.0904***	-1.8144***
	[0.044]	[0.049]	[0.084]	[0.107]	[0.158]
Constant	2.6858***	1.3032***	2.5090***	3.6427***	5.0832***
	[0.032]	[0.030]	[0.060]	[0.080]	[0.120]
Observations	55,911	21,694	13,515	11,996	8,706
R-squared	0.216	0.193	0.182	0.201	0.202
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. All columns restrict the analysis to the time window between two weeks before the introduction and two weeks after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.III: EFFECTS OF POINTS AND LEADERBOARD SYSTEM: WITH TIME TRENDS

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.7344***	-0.0413	-0.5413**	-1.3048***	-2.1681***
	[0.138]	[0.164]	[0.262]	[0.334]	[0.497]
Constant	3.2607***	1.7509***	3.0376***	4.3618***	6.0285***
	[0.084]	[0.082]	[0.158]	[0.216]	[0.332]
Observations	95,342	37,171	22,978	20,427	14,766
R-squared	0.186	0.162	0.160	0.174	0.174
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system, a linear time trend, and the interaction of the time trend with the post-system change dummy. All columns restrict the analysis to the time window between one month before the introduction and one month after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.IV: EFFECTS OF POINTS AND LEADERBOARD SYSTEM: GENDER HETEROGENEITY

Dependent variable	Number of correct answers per day				
	(1)	(2)	(3)	(4)	(5)
Post-system change dummy	-0.5906***	0.2135***	-0.5390***	-1.1607***	-1.8871***
	[0.053]	[0.059]	[0.100]	[0.126]	[0.198]
Post-system*male	-0.0447	-0.0864	0.0012	-0.1110	0.0122
	[0.073]	[0.080]	[0.138]	[0.178]	[0.271]
Constant	2.7120***	1.3593***	2.5024***	3.6480***	5.1549***
	[0.029]	[0.028]	[0.055]	[0.073]	[0.111]
Observations	86,270	33,546	20,780	18,663	13,281
R-squared	0.186	0.162	0.156	0.178	0.172
Sample:	FULL	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4

Notes: This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system, and the interaction of the post-system change dummy with a male student dummy. All columns restrict the analysis to the window between one month before the introduction and one month after it. Column 1 presents the results for the entire sample. Columns 2-5 present results by quartile of the distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.V: BALANCE OF COVARIATES FOR SOME-HONORS STUDENTS

	Private condition [1]	Public condition [2]	*p*-value [1]=[2] [3]	Non-honors [4]	Honors [5]	*p*-value [4]=[5] [6]
			PANEL A. STUDENTS TAKING 1-3 HONORS CLASSES			
Male	0.412	0.491	0.14	0.5	0.396	0.054
	[0.494]	[0.501]		[0.501]	[0.491]	
Age	16.694	16.708	0.745	16.748	16.647	0.022
	[0.395]	[0.425]		[0.45]	[0.351]	
Hispanic	0.964	0.977	0.505	0.961	0.981	0.268
	[0.186]	[0.152]		[0.194]	[0.136]	
GPA	2.676	2.666	0.904	2.461	2.914	0.00
	[0.764]	[0.783]		[0.673]	[0.811]	
# math/sciences honors	0.359	0.422	0.302	0.239	0.566	0.00
	[0.539]	[0.592]		[0.499]	[0.590]	
# social sciences honors	0.853	0.789	0.136	0.788	0.849	0.216
	[0.417]	[0.48]		[0.423]	[0.48]	
# of humanities honors	0.694	0.682	0.856	0.342	1.088	0.00
	[0.653]	[0.568]		[0.509]	[0.455]	
Observations	170	173		184	159	
			PANEL B. STUDENTS TAKING 2 HONORS CLASSES			
Male	0.333	0.434	0.289	0.415	0.352	0.506
	[0.476]	[0.50]		[0.498]	[0.482]	
Age	16.648	16.703	0.519	16.731	16.617	0.177
	[0.423]	[0.44]		[0.45]	[0.406]	
Hispanic	0.944	0.981	0.327	0.942	0.981	0.300
	[0.231]	[0.139]		[0.234]	[0.136]	
GPA	2.756	2.582	0.212	2.765	2.576	0.1725
	[0.687]	[0.744]		[0.55]	[0.846]	
# math/sciences honors	0.278	0.283	0.955	0.321	0.241	0.384
	[0.452]	[0.5]		[0.510]	[0.432]	
# social sciences honors	0.926	0.906	0.828	0.906	0.926	0.827
	[0.47]	[0.491]		[0.30]	[0.61]	
# of humanities honors	0.815	0.774	0.665	0.736	0.852	0.224
	[0.517]	[0.466]		[0.56]	[0.408]	
Observations	54	53		53	54	

Notes: Panel A. restricts the sample to students taking between 1 and 3 honors classes, Panel B. restricts to those taking 2 honors classes. Columns 1 and 2 report the mean level of each variable, with standard deviations in brackets, for the private and public conditions; column 3 reports the *p*-value of a test that the means are the same in both conditions. Columns 4 and 5 report the mean level of each variable, with standard errors in brackets, for non-honors and honors classes; column 6 reports the *p*-value of a test that the means are the same in both types of classes.

TABLE A.VI: EFFECT OF PUBLIC TREATMENT ON SIGNUP DECISION:
HONORS AND NON-HONORS CLASSES SEPARATELY

Dependent variable:	Dummy: Student signed up for the SAT prep course					
	(1)	(2)	(3)	(4)	(5)	(6)
Public treatment	-0.1083***	-0.1195***	-0.1231***	0.0157	0.0095	0.0092
	[0.040]	[0.040]	[0.040]	[0.033]	[0.032]	[0.031]
Mean of private take-up		0.717			0.917	
Includes individual covariates	No	Yes	Yes	No	Yes	Yes
Includes classroom and surveyor FE	No	No	Yes	No	No	Yes
Observations	560	531	531	265	258	258
R-squared	0.013	0.042	0.104	0.001	0.035	0.139
Sample:		Non-honors classes			Honors classes	

Notes: The first three columns of this table restrict the sample to non-honors classes, while the last three focus on honors classes. Columns 1 and 4 present OLS regressions of a dummy variable on whether the student signed up for the SAT prep course on a public sign up dummy. Columns 2 and 5 replicate columns 1and 4 adding individual covariates (age and dummies for male and Hispanic). Columns 3 and 6 replicate columns 2 and 5 adding surveyor and classroom fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.VII: EFFECT OF PUBLIC TREATMENT ON SIGNUP DECISION: HETEROGENEITY BY GENDER

Dependent variable:	Dummy: Student signed up for the SAT prep course					
	(1)	(2)	(3)	(4)	(5)	(6)
Public treatment	-0.0836	-0.1053*	-0.1088*	0.0232	0.0249	0.0360
	[0.056]	[0.057]	[0.059]	[0.031]	[0.032]	[0.034]
Male student dummy	-0.0887*	-0.1119**	-0.0951*	-0.0814	-0.0555	-0.0333
	[0.054]	[0.054]	[0.054]	[0.053]	[0.051]	[0.049]
Public*Male	-0.0454	-0.0256	-0.0257	-0.0078	-0.0366	-0.0640
	[0.079]	[0.081]	[0.081]	[0.070]	[0.069]	[0.069]
Mean of private sign-up for female students		0.766			0.95	
Includes individual covariates	No	Yes	Yes	No	Yes	Yes
Includes classroom and surveyor FE	No	No	Yes	No	No	Yes
Observations	560	531	531	265	258	258
R-squared	0.027	0.042	0.104	0.026	0.036	0.142
SAMPLE		Non-honors classes			Honors classes	

Notes: The first three columns of this table restrict the sample to non-honors classes, while the last three focus on honors classes. Columns 1 and 4 present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign up dummy, a male dummy and the interaction of the two. Columns 2 and 5 replicate columns 1 and 4 adding individual covariates (age and male and Hispanic dummy). Columns 3 and 6 replicate columns 2 and 5 adding surveyor and classroom fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.