

Appendices for “Should College Be “Free”?” (for online publication)

Douglas N. Harris and Jonathan Mills

Appendix A. Theory and Hypothesis Details

Appendix B. Additional Information on Milwaukee

Appendix C. Randomization Details

Appendix D. Program Communications and Their Possible Effects

Appendix E. Bootstrap Standard Errors

Appendix F. Local Average Treatment Effects

Appendix G. Additional Tests Related to Merit Requirements

Appendix H. Effect Heterogeneity Results

Appendix I. Cost-Benefit Analysis Details

Appendix J. Pre-Specification of Analysis Plan

A. Theory and Hypothesis Details

The main text summarized hypotheses about the effects of financial aid. This section elaborates on the origin of those hypotheses using a formal model.

A.1 Standard Economic Theory

We assume that students (indexed by i) are rational and maximize utility. Specifically, they choose whether to obtain a college degree and therefore receive an earnings payoff w_i , which accounts for the opportunity cost of lost earnings while in college, varies across individuals, and follows a well-defined distribution with properties outlined below. We normalize earnings to zero in the event students do not attend college.

For simplicity, the price of college p is fixed for all students and equals the economic cost of college production. The government may offer financial aid in the form of a fixed grant g to every student who attends college. When $p > g$, students take out loans. For simplicity, we assume the interest rate is zero, as this is not a key factor in the model.

For now, we assume that students have perfect information and make their dichotomous decision about whether to obtain a college degree d_i to maximize indirect utility:

$$V(d_i) = d_i \cdot (w_i - p + g), \quad d_i \in \{0,1\}. \quad (\text{a1})$$

Given the above assumptions, when students choose not to attend college ($d_i = 0$), indirect utility is zero. Students will choose to attend college if and only if $w_i - p + g > 0$ and the variation in w_i ensures that some students attend college and others do not. This leads to our first hypothesis:

Hypothesis 1: The average treatment effect of aid on college attainment is positive. This follows directly from equation (1) because as g rises, the net price declines and the share of students with $V(d_i = 1) > 0$ rises. This hypothesis also maintains when we relax the simplifying assumptions.

Student High School Effort. Student effort e_i during high school may also affect college degree choices. For example, it may be that high school effort increases the chances of gaining admission to college. For this reason, we now alter the decision from “receive a degree” to “attend and/or matriculate” m_i , so the students choose (m_i, e_i) , which in turn determine whether

students receive a degree through the function $d_i(m_i, e_i)$.¹ We also assume that effort comes at a cost $c(e_i)$, which is increasing in e_i and strictly concave, and this modifies the optimization problem to:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - p + g) - c(e_i), \quad m_i \in \{0,1\}. \quad (\text{a2})$$

In this case, optimal effort e_i^* occurs where $\frac{\partial V}{\partial e_i} = (\partial d / \partial e_i)(w_i - p + g) - c'(e_i) = 0$. This implies that optimal effort is increasing in g with $\frac{\partial^2 V}{\partial e_i \partial g_i} = \frac{\partial d}{\partial e_i} > 0$.

Hypothesis 2a: Optimal high school effort e_i^ is increasing in g .*

Hypothesis 2b: Optimal high school effort e_i^ is decreasing in $c'(e_i)$.*

To this point, we have assumed that students are perfectly informed. But prior research suggests that students, especially from low-income families, are imperfectly informed about college financial aid (Avery & Kane, 2004; Hoxby & Turner, 2015; Dynarski et al., 2021). One implication is that, if some students underestimate g , then communicating the true g earlier will increase high school effort.

Hypothesis 2c: Early commitment aid increases optimal high school effort e_i^ more than late commitment aid.*

Alternatively, rather than underestimating g , students might just be uncertain about it and hold rational expectations. Further, if students are risk-averse, then it might seem that we still arrive at *Hypothesis 2c* even without underestimating g .² But whether risk-aversion increases the benefit of early aid commitments depends on the return to effort. If $E \left[\frac{\partial d}{\partial e_i} \right]$ and/or $E[w_i - p + g]$ are large, then additional effort is a kind of insurance, and risk-aversion leads to increased effort; however, if these expected values are small, then increasing effort is more like gambling and reduces effort of risk-averse agents relative to the certainty/risk-neutrality case (McGuire, Pratt, & Zeckhauser, 1991).³ For this reason, uncertainty and risk-aversion alone are insufficient to

¹ High school effort could also affect the probability of completing a college degree conditional on college entry if, for example, the degree production function $d_i(m_i, e_i)$ took a form such as $d_i = m_i e_i^\gamma$. We do not focus on these complementarities because this over-complicates the model and distracts from the core testable hypotheses.

² This is especially plausible in this case because young people whose parents have lower levels of education, as in Milwaukee, are more risk-averse than other young people (Hryshko, Luengo-Prado, & Sorensen, 2011). Risk-aversion also increases with age (e.g., Harbaugh, Krause, & Vesterlund, 2002), so teenagers may be less risk-averse than adults.

³ McGuire, Pratt, and Zeckhauser (1991) explore general forms of uncertainty theoretically and in a way unrelated to college decisions. Also, see Skaperdas and Gan (1995) for a similar theoretical analysis and conclusions in a tournament setting.

obtain *Hypothesis 2c*.⁴

Merit Requirements. Suppose next that the government establishes student eligibility requirements based on academic performance, i.e., merit aid, so that student academic performance at the end of high school a_i has to be above some threshold A (e.g., a minimum grade point average or test score) in order to receive g . Students can exert effort e_i during high school to meet these requirements. This effort increases academic outcomes according to the education production function $h(e_i)$, which is increasing in e_i and strictly concave. The extent of effort required to meet the requirement depends on student academic endowment α_i , such that academic outcome $a_i = \alpha_i + h(e_i)$. Whether students meet the merit requirement and have the opportunity to use the grant can be represented by the following indicator function: $1[\alpha_i + h(e_i) \geq A]$.

This yields the more complex indirect utility function:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - p + (1[\alpha_i + h(e_i) \geq A] \cdot g) - c(e_i) \quad (\text{a3})$$

Students again choose (m_i, e_i) and, if they meet the merit requirements, then receive g as a price reduction.

Equation (3) implies two main ways that the addition of merit requirements might influence the percentage of students attending college. On the one hand, it reduces the probability that students receive g and, by *Hypothesis 1*, reduces the share of students choosing college. This is only false when: (a) $\alpha_i > A \forall i$ (in which case the merit requirements are never binding); (b) some students have $\alpha_i < A$, but $h'(e_i) \gg c'(e_i)$ (i.e., there is a strong incentive effect); or (c) the merit requirements direct aid to students who are most responsive to it (i.e., there is a strong targeting effect). The potential targeting effect is not evident from (3) as specified above and requires modifying the degree attainment function to be $d_i(\alpha_i, m_i)$ with $\frac{\partial d_i}{\partial \alpha_i} > 0$, so that college degree attainment is increasing in the academic endowment even apart from the grant and associated merit requirements. This discussion of the incentive and targeting

⁴ This might seem to contradict Dynarski and Scott-Clayton (2006) who write: “Uncertainty about aid similarly blunts its impact on behavior: high school students most sensitive to cost are unlikely to start down the path to college [i.e., put forth effort] if they do not know it is affordable. For those on the margin of college entry, concrete information about aid simply arrives too late.” Our model adds precision to this statement by clarifying that “uncertainty” per se is insufficient, but imperfect information in the form of under-estimation of the net benefits of college is sufficient. Also, Dynarski and Scott-Clayton (2006) focus on students “on the margin” for whom additional effort is likely to constitute more of a gamble than insurance, in which case uncertainty does indeed “blunt [financial aid’s] impact.”

effects yields two hypotheses:

Hypothesis 3a: Merit requirements reduce the intent-to-treat (ITT) effect on college outcomes in proportion to the share of students with $\alpha_i < A$, unless the incentive effect is large. The ITT effect pertains to all students who are offered the merit-based grant. So, when many students have $\alpha_i < A$, fewer students will receive the grant, unless this is fully offset by a strong incentive to increase effort.

Hypothesis 3b: Merit requirements increase the treatment-on-treated (TOT) effect on college outcomes if the targeting effect is positive. The TOT is the effect of the grant offer on those who meet the merit requirements and obtain the grant funds. The reduction in the number of students receiving the grant is irrelevant in the TOT, where being treated means meeting the merit requirement and being eligible to receive the funds. So, any targeting effect increases the TOT for aid with merit requirements relative to the TOT for aid *without* merit requirements.

Merit Aid Effects on High School Effort. Next, we show that the optimal effort level e_i^* is a non-linear function of the academic endowment α_i . When students have perfect information and $\alpha_i \geq A$, they need not exert any additional effort to meet the requirements (i.e., their baseline level of effort is sufficient to obtain the grant). Students with $\alpha_i < A$, in contrast, will either: (a) exert exactly enough additional effort to obtain $\alpha_i = A$; or, if α_i is sufficiently low, the disutility of effort exceeds the utility from w_i , in which case $e_i^* = 0$.

Figure A1 illustrates this with the high school effort response function. High school effort is on the vertical axis and the academic endowment effort is on the horizontal axis; the vertical dashed line indicates the academic merit threshold A . To focus attention on the potential additional effort induced by the merit requirements, we normalize the baseline level of effort without the requirements to zero.⁵ The solid black line is the effort response function when students are certain about all the parameters.⁶

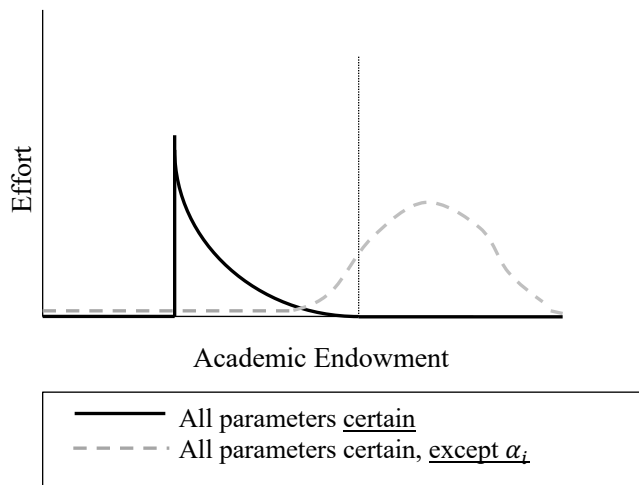
Now, suppose that students are risk-averse and that α_i for each student comes from a random

⁵ Note that the response function in the certainty case is curved when $\alpha_i \in [A - \varepsilon, A]$ because of the concavity of $h(e_i)$. With a declining return to effort, it takes more and more effort to move the same distance closer to A .

⁶ We assume there is a continuum of students in terms of α_i , which ensures that the effort function exists and that we can observe the effort level associated with each endowment.

draw from a well-defined distribution (all other parameters remain known).⁷ (This is especially plausible if the merit threshold is based on a measure like a college entrance exam where the score is not known until the end of high school.) Further, suppose that rational expectations so that $E[\alpha_i] = \alpha_i$; that is, students expect to draw their true α_i . The effort response function in this uncertainty case is reflected in the dashed grey line in Figure 1A.

Figure A1: High School Effort Response by Academic Endowment in Response to Merit Requirements (Standard Theory)



⁷ For simplicity, we assume that w_i and α_i are independent of one another. While this is unrealistic, relaxing this assumption does not change the main conclusions here.

At the extremes, effort response is the same in the certainty and uncertainty cases; students with very low and very high α_i do not increase their effort in response to the performance incentives. Unlike the certainty case, however, most of the effort with uncertain α_i comes from students with $\alpha_i > A$ because, with unknown α_i , they wish to insure against the possibility that their draw of α_i is lower than expected, while students with $\alpha_i < A$ would not exert additional effort because of the low likelihood of reaching A means they are essentially gambling with a low probability of payoff (McGuire, Pratt, & Zeckhauser, 1991). Given this, the peak of the effort response function for the risk-averse group will be to the right of A for all plausible forms of risk-aversion.⁸

A further implication of Figure A1 is that when certainty about α_i is high, we can expect additional effort to peak to the left of the threshold. As certainty about α_i declines, the peak of the effort response function shifts to the right. Also, there is no sharp change in effort at the threshold under any of these scenarios. With the most realistic case, where students are uncertain, students will increase effort on both sides of the threshold.

In using the above analysis to develop hypotheses for our evaluation of TDP, two factors seem important. On the one hand, TDP involves a fair amount of certainty about α_i because students are given regular updates about their GPAs and attendance by both their schools and the program administrator. On the other hand, as low-income teenagers, Milwaukee high school students may be highly risk averse. Putting these two facts together implies that the actual effort response function is somewhere in between the black and grey lines in Figure A1.

Hypothesis 4: The high school effort response function has an inverse-U shape with a peak near the merit threshold. This hypothesis is vague because the point of the peak depends on the level of certainty about the endowment and the nature of risk-aversion, which are somewhat speculative.

So far, for simplicity, we have assumed that there is only one type of effort, but it is easy to see how there might be at least two forms: effort directed at meeting the merit requirements (e_i^A) and other effort that may increase future earnings in other ways (e_i^o). Assuming that e_i^A also contributes positively to increasing future wages (e.g., effort that increases GPA also increases

⁸ Some students with $E[\alpha_i] < A$ might still exert some effort depending on the nature of the risk aversion. Also, note that the effort response becomes less clear when we add uncertainty over other parameters and allow for the possibility that they are all correlated.

productivity), e_i^A increases indirect utility in two ways, while e_i^o contributes in only one way. Further, if the marginal product and marginal cost functions are otherwise the same for both forms of effort, then this means $\partial V/\partial e_i^A > \partial V/\partial e_i^o$. This leads to the following hypothesis that is also testable in the analysis that follows:

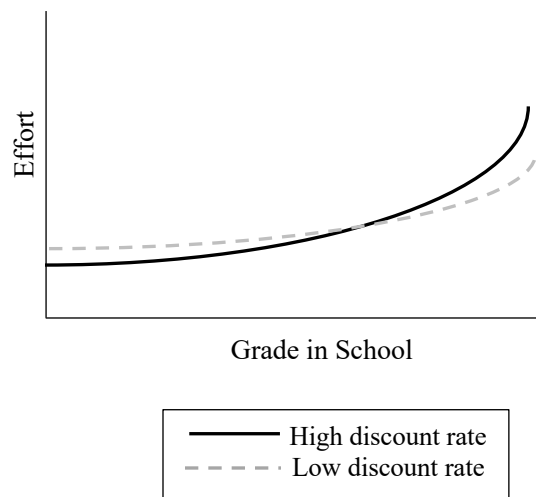
Hypothesis 5: When aid includes merit requirements, the effect of aid on effort is especially positive for high school behaviors associated with the requirements, ceteris paribus.

Timing of High School Effort Effects. Another dimension of the effort optimization problem is that students can exert effort in multiple high school time periods, not in the single high school period assumed above. We therefore allow effort levels and effort costs to vary by grade r , so that $e_i = \sum_r e_{ir}$ and $c(e_i) = \sum_r (1 - \delta)^t c(e_{ir})$. For all discount rates $\delta > 0$, the standard model predicts some delay in student effort until the end of high school and this delay is increasing in the discount rate. However, if the effort cost function has $c''(e_{ir}) < 0$ within each grade r , then this concavity partially counteracts the delaying effect from the discount rate and makes it more desirable to smooth effort across grades. The effort response timing under these two scenarios are shown in Figure A2.

Hypothesis 6: Students increase their effort more as they near the end of high school.

Since Hypothesis 6 is driven by the discount rate, note that, like risk-aversion, the discount rate declines with age, income, and education (e.g., Lawrence, 1991; Samwick, 1998). This implies that the delays in effort will be largest for low-income teenagers from families with lower levels of education who are the subject of this study. For this reason, Figure 1B displays two lines, one for low- and high-discount-rate students.

Figure A2: High School Effort by Age/Grade (Standard Theory)



A.2. The Behavioral Economics of Financial Aid

Status quo bias refers to the idea that people tend to stay with their current path unless there is a clearly superior alternative (Samuelson and Zeckhauser, 1988). This bias seems especially likely where there is decision complexity (Dynarski & Scott-Clayton, 2006), which increases the cognitive load of decisions. Relatedly, complex decisions can also involve complex administrative processes that increase the chances that students will miss a step, so that efforts will keep students from pursuing their intended course of action (Avery and Kane, 2004; Bettinger et al., 2009).

Loss aversion refers to the idea that the loss of utility from giving something up that we already have may be larger in absolute value than the utility gain from receiving that same item (Kahneman & Tversky, 1979). This is related to status quo bias in the sense that a “loss” can only occur when people already have, or perceive that have they, something at the time the decision is being made. Loss aversion may be related to the notion of loan aversion in the sense that people worry that, in borrowing funds, that debt poses a risk of losing what they already have (especially if they do not finish college or are unable to find a job after college). There is some evidence of loan aversion among high school students (e.g., Boatman et al., 2017) and we provide more evidence of this later.

Anchoring is the idea that people set reference points as rules of thumb to help guide their decisions (Tversky & Kahneman, 1974). Often there is an almost infinite number of possible options and these anchors, or reference points, reduce the number of options to something requiring lower cognitive load. In the context of financial aid, when student eligibility requirements involve merit thresholds, these thresholds might alter students’ anchors and change their decisions, apart from the effort incentives in the standard model.

Salience means that information has its greatest impact when it is most prominent or available (Chetty, Looney, and Kroft, 2009). This may be because people are prone to forget information when it is received well before (or after) a decision is made, e.g., students have forgotten about student eligibility requirements in past studies of financial aid (Angrist et. al., 2009), reducing its effects.⁹

⁹ Dynarski and Scott-Clayton (2006) also discuss “identity salience.” For example, when people fill out application forms for government aid, the wording may trigger negative feelings and guilt, due to stereotypes associated with one’s own identity (Bertrand, Mullainathan, and Shafir, 2004). Similarly, Dynarski and Scott-Clayton (2006) point

We see no reason why these behavioral economic theories would substantially alter *Hypotheses 1, 2a, 2b, 3a, 3b* or *5*. But this is not the case with *Hypotheses 2c, 4* and *6*. We therefore re-state these latter three assumptions in their original forms and then explain how they require modification based on behavioral economics.

Hypothesis 2c (Original): Early commitment aid increases optimal high school effort e_i^ more than late commitment aid.* This original hypothesis was rather tenuous and depended either on students' underestimation of the net returns to college education or particular assumptions about uncertainty and risk-aversion (i.e., the insurance case). But the hypothesized effect turns clearly positive when we introduce status quo bias. For students who have low initial expectations of college entry (e.g., first-generation college students) (Dynarski & Scott-Clayton, 2006), a specific early commitment could change the default by sending the message to students that they are "college material."

Hypothesis 4 (Original): The high school effort response function has an inverse-U shape with a peak near the performance threshold. By the same reasoning as in *Hypothesis 2c* above, students with $\alpha_i \ll A$ might increase their effort, despite the high cost because of a shift in their anchor or reference point. In contrast, for many of the high- α_i students, the announcement of A might reduce their anchors and therefore their effort (e.g., a student with a 3.5 GPA might reduce effort if it is announced that grants require only a 2.5 GPA). These behavioral economic concepts suggest a negative relationship between e_i^* and α_i , not an inverted-U.

Hypothesis 6 (Original): Students increase their effort more as they near the end of high school. Salience reinforces the effects of discounting, which led to the prediction that effort will put off until twelfth grade because this is when the (initial) decision to attend college is arises. In addition, however, salience means that students are also likely to respond with additional effort when they first learn about the grant. In the case of TDP, this was in the fall of ninth grade.

In addition to modifying some of the original hypotheses, two additional hypotheses emerge from the above behavioral economic theories:

Hypothesis 7 (New): Free college creates a discontinuous increase in the effect of g at the point where $p = g$. To see why, suppose we add an additional psychological cost $\zeta > 0$ if and only if $p - g > 0$, reflecting that students have to take out loans and are

out that some FAFSA questions ask about criminal behavior and drug use and argue that this reinforces negative stereotypes for low-income and minority students.

loan averse ($\zeta = 0$ otherwise). Indirect utility becomes:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot ((w_i - p + g) - (\zeta[p - g > 0])) - c(e_i) \quad (4)$$

In this situation, free college removes the psychological cost and shifts the return to a college degree upwards, above and beyond the marginal effect of increasing g when $p > g$ (We would also frame ζ as a cost of decision complexity, which is also suddenly eliminated when $p = g$.¹⁰)

Hypothesis 8 (New): The ITT effects of aid are largest when combining free college, early commitment, and no merit requirements. Above, we outlined hypotheses for each of these policies as if they are independent of one another (*Hypotheses 2b, 3a, 3b, and 7*), but they may also be complementary. To see why, suppose we combine equations (3) and (4) to include all the various factors that could be involved in the college decision, including those rooted in behavioral economics. This yields:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - p + (1[\alpha_i + h(e_i) \geq A] \cdot g - \zeta[p - g > 0])) - c(e_i) \quad (5)$$

This equation illustrates a key point: that free college does little to reduce decision complexity or loan aversion in the presence of student eligibility requirements. In contrast, a policy of free college without such requirements simplifies the problem to maximizing just: $V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - c(e_i))$. Thus, the degree of simplification depends on how the policy elements are combined.

The above models and hypotheses are useful in several ways. First, they help show us how both the standard theory and behavioral economics may help explain students' financial aid responses. Second, the models yield specific testable hypotheses about student behavior that inform and guide the empirical analysis of TDP and our interpretation of the results that follow.

¹⁰ The decision entailed in equation (4) would be even more complex if we added in real-world factors such as the variability of p across colleges and time, and that g is a function of p through the complex aid formula.

Appendix B. Additional Information About Milwaukee

Table B1 provides detailed descriptive statistics for the entire TDP cohort. Below, we also provide additional information about the high schools involved.

Milwaukee is home to the most ambitious school choice experiment in the nation; school-age students are eligible to attend an array of charter, inter-district choice, and private schools with public funding.¹ As shown in Table B2, 20 of the 36 TDP eligible schools are some form of traditional public schools and the others are charter schools, which operate semi-autonomously from MPS and have smaller enrollments.²

Given the high school graduation merit requirement, it is worth noting the academic requirements that the state and MPS have to qualify for high school graduation (MPS, n.d., p.2). Under the district's standards, MPS ninth graders have to complete 4.0 units of English/language arts and 3.0 units each of mathematics (only courses that include or go beyond Algebra I; remedial courses do not count toward this total), science, and social studies, among other requirements involving physical education, service learning, and standardized test scores. General Educational Development certificates (GEDs) do not qualify.

In 2008-09, just before implementing a districtwide requirement to take the ACT, MPS's composite ACT score was 17.3, which is below the urban district average (18.3) and the national average (21.1). More recently, the district has mandated ACT participation as part of its college-going efforts, which has reduced the ACT average, but in ways that make it less comparable with other cities during the reported time period.

In addition to common programs such as the federal TRIO college access programs and other community-based efforts, MPS has attempted to emulate nearby Chicago Public Schools and created a variety of district-wide programs to address these low college enrollments, including the creation of two college access centers located outside the schools, but accessible to all students attending publicly funded high schools. These efforts were apparently successful as the percentage of students who met the TDP requirements (control and treatment) who went on

¹ Approximately 20,000 students receive private school vouchers to attend in elementary and middle schools and many switch back to MPS schools when they enter high school. The Chapter 220 program funds students in Milwaukee to attend suburban districts, though it is much smaller with approximately 2,700 total students transferring either into or out of MPS across all grades. One implication of this is that we have some missing MPS data on some pre-treatment (eighth grade) data in the TDP sample.

² The choice of variables in Table A2 is driven by the need for school-level confidentiality. Publicly available variables such as race and free or reduced lunch eligibility would give away school identities.

to college increased by roughly 10 percentage points over the course of the project.³ This is informative regarding the counterfactual.

³ As noted earlier, 64 percent of students at baseline who met the TDP requirements went on to college. In contrast, of the 469 (437) students in the TDP control (treatment) group who met the requirements, 350 (335) students went to college. These translate into 74.6 and 76.7 percent, respectively.

Table B1: Descriptive statistics for dependent variables

Variable	Full Sample					Senior Exit Survey Sample	
	N	Mean	s.d.	Min	Max	Mean	p-value diff
PANEL A: Dependent Variables							
<i>High School Completion</i>							
Attendance pct	5033	0.808	0.188	0.023	1.000	0.917	0.000
Passed 90% Threshold	5033	0.452	0.498	0.000	1.000	0.723	0.000
GPA	4948	1.760	0.961	0.000	4.000	2.375	0.000
Passed 2.5 GPA Threshold	4948	0.246	0.430	0.000	1.000	0.420	0.000
Meets Both DP requirements	4948	0.212	0.409	0.000	1.000	0.377	0.000
Math MAP	4761	0.006	0.950	-3.500	3.414	0.368	0.000
Read MAP	4753	0.001	0.917	-3.876	2.910	0.338	0.000
Transferred schools	5038	0.433	0.496	0.000	1.000	0.187	0.000
Student missing in 2014-15 data	5038	0.263	0.440	0.000	1.000	0.000	0.000
Student Grad. On Time, Reg. Diploma	5038	0.498	0.500	0.000	1.000	0.964	0.000
Student Grad. Any Credential by 2016	5038	0.550	0.498	0.000	1.000	0.978	0.000
<i>Steps to College</i>							
# college support programs participated in	2146	0.550	0.902	0.000	14.000		
Planning to go to college	2146	0.681	0.466	0.000	1.000		
Planning on 4-year college	2146	0.449	0.497	0.000	1.000		
Planning on 2-year college	2146	0.232	0.422	0.000	1.000		
Planning on full-time college	2144	0.569	0.495	0.000	1.000		
# colleges applied to	2146	2.405	2.542	0.000	9.000		
First choice college is highest prestige	1920	0.112	0.315	0.000	1.000		
Completed FAFSA	2129	0.519	0.500	0.000	1.000		
Applied for scholarships	1461	0.617	0.486	0.000	1.000		
Awarded scholarships	808	0.420	0.494	0.000	1.000		
Intends to use scholarships	1461	0.617	0.486	0.000	1.000		
Biggest roadblock to college is cost	2135	0.555	0.497	0.000	1.000		
Planned on attending college, did not enroll	2146	0.522	0.500	0.000	1.000		
Planned on 4-year college, enrolled in 2-year	2146	0.139	0.346	0.000	1.000		

Variable	Full Sample					Senior Exit Survey Sample	
	N	Mean	s.d.	Min	Max	Mean	p-value diff
<i>College Enrollment</i>							
Any college attendance	5038	0.325	0.468	0.000	1.000	0.631	0.000
2-year college enrollment	5038	0.199	0.400	0.000	1.000	0.365	0.000
4-year college enrollment	5038	0.176	0.381	0.000	1.000	0.373	0.000
Competitiveness	935	0.895	0.657	0.000	2.000	0.923	0.045
Part-time status	5038	0.207	0.405	0.000	1.000	0.404	0.000
Full-time status	5038	0.219	0.414	0.000	1.000	0.454	0.000
In-state college	5038	0.283	0.451	0.000	1.000	0.553	0.000
Out-of-state college	5038	0.066	0.249	0.000	1.000	0.127	0.000
TDP eligible college	1594	0.829	0.377	0.000	1.000	0.838	0.114
College quality (grad. rate)	1566	0.428	0.190	0.100	0.970	0.445	0.000
Persistence to spring	5038	0.079	0.269	0.000	1.000	0.174	0.000
Ever graduated	2866	0.036	0.187	0.000	1.000	0.042	0.024
Ever graduated from 2-year institution	2866	0.023	0.151	0.000	1.000	0.027	0.024
Ever graduated from 4-year institution	2866	0.013	0.111	0.000	1.000	0.014	0.388
PANEL B: Independent Variables							
Female	4995	0.491	0.500	0.000	1.000	0.559	0.000
Age	4995	15.149	0.697	13.114	21.243	14.946	0.000
Black	4995	0.624	0.484	0.000	1.000	0.589	0.086
Asian	4995	0.059	0.236	0.000	1.000	0.098	0.003
Hispanic	4995	0.198	0.398	0.000	1.000	0.180	0.354
White	4995	0.110	0.313	0.000	1.000	0.128	0.034
Free/Red. Price Lunch	4995	0.831	0.374	0.000	1.000	0.781	0.003
Special Education	4995	0.210	0.408	0.000	1.000	0.117	0.000
English Language Learner	4995	0.109	0.312	0.000	1.000	0.108	0.931
English at Home	4995	0.857	0.350	0.000	1.000	0.841	0.229
Spanish at Home	4995	0.097	0.296	0.000	1.000	0.090	0.604
Hmong at Home	4995	0.025	0.156	0.000	1.000	0.044	0.017
Other Lang. at Home	4995	0.022	0.145	0.000	1.000	0.025	0.160
Student College-Going Probability	4818	0.280	0.248	0.005	0.917	0.404	0.000
School College-Going Culture	4218	3.144	0.319	2.333	3.788	3.227	0.001
School Climate	4218	2.736	0.380	1.167	3.685	2.787	0.038

Notes: High school outcomes and student characteristics are from MPS administrative data (with outcomes cumulative across grades). Math and reading scores are averaged across all available tests (a maximum of three per year). College outcomes are from the NSC, collected one year after on-time high school graduation. The “p-value diff” is the p-value of the difference between the survey sample mean and the full sample mean. The means do not match those reported in the main text tables because the latter focuses on the control group mean and those here are full sample means, focused on the post-treatment years.

Table B2: School characteristics by school and pair

Pair/School	# Eligible TDP Recipients	Closure Year	School Type	Attendance	8th Grade Adjusted GPA
1T	58		Charter	0.96	3.42
1C	74		Charter	0.96	2.95
2T	294		Citywide/Specialty	0.96	3.19
2C	371		Citywide/Specialty	0.96	3.10
3T	410		Citywide/Specialty	0.94	2.62
3C	249		Citywide/Specialty	0.94	2.54
4T	59	2012	Charter	0.89	1.96
4C	133		Citywide/Specialty	0.95	2.50
5T	71		Citywide/Specialty	0.92	2.47
5C	79	2012	Charter	0.93	2.32
6T	450		Traditional	0.89	2.00
6C	306		Traditional	0.89	1.75
7T	269		Citywide/Specialty	0.87	1.75
7C	350		Traditional	0.89	1.81
8T	11		Partnership	0.75	1.39
8C	23	2013	Charter	0.93	2.24
9T	296		Charter	0.87	1.77
9C	333		Traditional	0.87	1.87
10T	23		Charter	0.76	1.12
10C	168		Citywide/Specialty	0.93	2.37
11T	69		Charter	0.89	1.74
11C	23	2012	Charter	0.90	1.42
12T	162		Charter	0.86	1.57
12C	100		Charter	0.84	1.59
13T	278		Traditional	0.86	1.86
13C	151		Traditional	0.86	1.64
14T	65	2014	Charter	0.92	1.98
14C	26		Charter	0.89	2.39
15T	35	2012	Charter	0.82	1.54
15C	13	2014	Charter	0.72	0.76
16T	7		Alternative	0.73	0.84
16C	7		Alternative	--	1.07
17T	16		Partnership	0.75	1.10
17C	39		Alternative	0.72	1.20
18T	15		Partnership	0.85	2.17
18C	19	2012	Charter	0.81	1.57

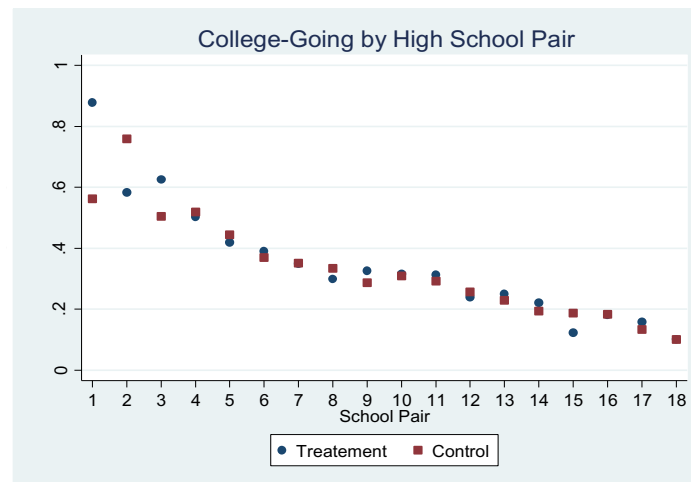
Notes: Pair/School indicates the pair number and which of the schools is the treatment school (T) and which is control (C). School Type categories include: Traditional, Charter, Citywide/Specialty (non-charter MPS schools without attendance zones), Alternative schools are those serving students with special needs. Partnership schools are similar to Alternative schools, except they are operated by a private provider under MPS contract. Attendance and GPA refer to the eighth-grade information for the TDP cohort. Other information, such as demographics, are omitted to avoid identifying the specific schools.

Appendix C. Randomization Details

The main text discusses the use of pair randomization where schools were paired based on the pre-treatment college-going rate. Specifically, we averaged the college attendance rate from the 2008-09 and 2009-2010 graduating classes (where available) to reduce random error. This is illustrated in Figure C1, which shows the baseline college-going rate for each school within each of the 18 pairs.

In one case, only the 2009-10 actual rate was available and we used that instead of the two-year average. Six of the 36 schools were too new to have any actual college attendance rate therefore MPS staff estimated a model of college entry using data from the other 30 schools and used this model to predict college entry in the other six schools. In the pairing process, we stratified according to whether the school had actual versus predicted attendance rate, which explains why the control-treatment differential is larger in a few cases in Figure C1.

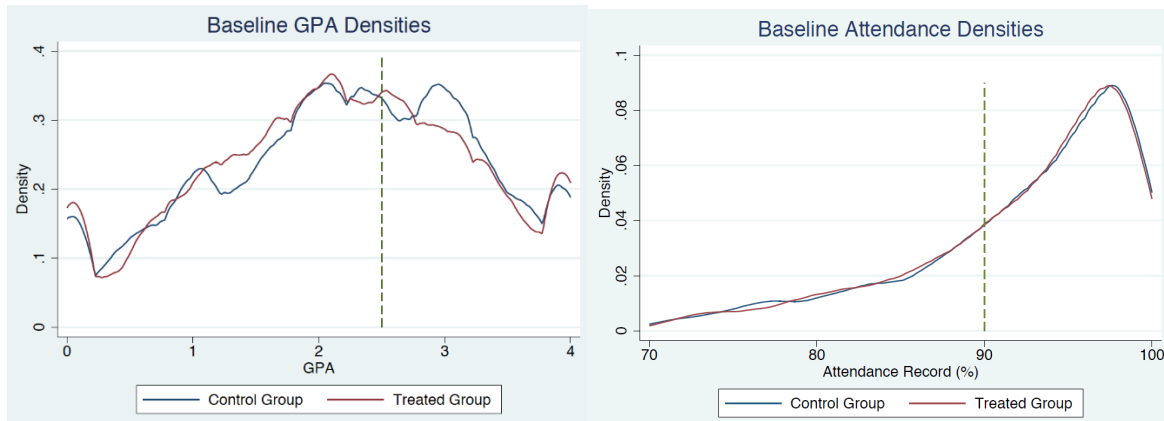
Figure C1: Baseline Equivalence



Notes: In Figure C1, each circle (square) is the pre-treatment on-time college entry rate for each treatment (control) *school*. These are shown for each pair of schools within which randomization occurred. In the pairing process, we stratified according to whether the school had actual versus predicted attendance rate, which explains why the control-treatment differential is larger in the first three pairs.

In Table 1 of the main text, we showed that this approach led to some small differences in the baseline control and treatment group means. In addition, we show the distribution of GPA and attendance in the control and treated schools in Figure C2.

Figure C2: Baseline Equivalence Distributions for GPA and Attendance



Notes: Figure C2 shows kernel density plots for baseline GPA and attendance for the control and treatment groups. These are the main two academic requirements that can be tested (high school graduation is the main additional requirement and has no baseline value).

Appendix D. Program Communications and Their Possible Effects

This section provides some additional description of the communications between the program administrator, students, and counselors. One copy of the announcement letter was hand-delivered to students at their schools on the announcement day and the other copy of the letter was sent home to parents the same day. Most schools also held assemblies with the students on the day of the announcement, one of which was attended by the lead author.

Schools were directed to return letters to the district office if the students were no longer in the schools. Of the 2,587 sent, only 84 were returned. Evidence presented later in the paper about student awareness reinforces the fact that most treatment group students received at least one of the letters.

Figure D1 provides an example of the letter provided to students. One key observation is that the letter says the \$12,000 “may be enough to cover the entire cost” of a degree from Milwaukee Area Technical College. The letter also provides information about the program requirements.

Figure D1: Sample Letter to Students



Would you like to see <Student First Name> go to college?

The Degree Project[®] is a special opportunity for your student to receive a college scholarship of up to \$12,000.

Whether they go to a two-year college like Milwaukee Area Technical College, where the scholarship may be enough to cover the entire cost, or a four-year college like the University of Wisconsin-Milwaukee—\$12,000 can go a long way toward paying for college.

How Does Your Student Earn \$12,000*

- 1 Earn a cumulative high school grade point average (GPA) of 2.5 or higher
- 2 Achieve a cumulative attendance record of 90% or higher
- 3 Graduate by August 30, 2015, from an eligible MPS school

Preparing for College

We want to help your student pay for college and be prepared for success when they get there.

To help, we developed these pieces to show you how they are progressing toward meeting the GPA and attendance requirements and provide recommendations on the classes needed to do well in college.

While we offered no specific hypotheses regarding the role of communication, financial experiments are not necessarily about money alone. The incentive to respond might depend on the nature of the communications students receive about the money (e.g., Bloom, Hill, & Riccio, 2003; Benhassine et al., 2013), including communication frequency and personalization.¹ One prior experiment found that college students generally forgot about their scholarship eligibility (Angrist et al., 2010), reinforcing the potential role of communication frequency, and that problem is likely to be worse with the typical low-income ninth grader who has somewhat lower college expectations and for whom college is far in the future. Others have found effects from providing information about existing aid (Oreopoulos & Dunn, 2013; Dynarski et al., 2021). The source of communication may also matter; hearing something from an unknown entity, such as Great Lakes, is likely to generate a different response than a message from a specific person (Valant & Newark, 2020) and this might be especially true with messages delivered in-person by a trusted adult, such as a school counselor. TDP provided both types of communications.

Student self-reported an average of three communications total, but with a wide range (0-19 communications).² This variation in communications is also associated with the TDP effects. We estimated a version of equation (5), adding an interaction term between treatment and the number of communications students reported receiving about TDP (see Appendix E). Consistent with some prior research, the effects of TDP are more positive when accompanied by more communications. The number of communications each student receives is clearly endogenous (e.g., counselors communicated with students who are more likely to go to college and/or respond to the scholarship offer), but these results are still noteworthy even as a descriptive exercise because they highlight the large differences in how students experienced what was seemingly a standardized program. The program administrators sent all students the same letters and communicated with all high school counselors the same way.

At least two factors seem to explain this low mean and wide range of communications. First, while communications were supposed to be standardized for students in school, some

¹ Others have also emphasized the specific messages participants receive (Bloom, Hill, & Riccio, 2003). Research by Oreopoulos and Dunn (2013) reinforces the potential importance of information in taking the steps toward college entry.

² The surveys asked students each year about: (a) whether they received a letter that year about a \$12,000 scholarship; and (b) how many times they heard about TDP from adults within their school. These measures were used to create a count of the number of communications. For reference purposes, if a student stayed in school all four years and reported receiving the letter every year and hearing from the school counselor twice per year, yielding three communications annually, then this would yield $3 \times 4 = 12$ communications in total.

students did not receive those communications due to attrition (dropping out or leaving MPS schools). Second, interviews with school counselors suggest that some of them consciously avoided communicating with specific students who were below the merit thresholds; in fact, they even came to view the students themselves in a more negative light because of their failure to respond academically to meet the merit requirements (Rifelj & Kuttner, 2020). This is consistent with recent evidence that school counselors vary considerably in their ability to help student get to college (Mulhern, 2019). More broadly, the implementation of aid programs, including the behavior and effectiveness of school counselors in helping students utilize free college programs, may be as important as the policy design.

Table D1 provides estimates of the interaction between treatment and the number of communications. The table shows only the interaction term. As explained in the main text, the strong positive interactions may mainly reflect the way in which counselors communicated with students.

Table D1: Average treatment effect by number of communications received

	N	Control Mean	Trt	Comm X Trt
Any College Enrollment	3523	0.422	-0.071* (0.026)	0.029** (0.008)
2-year college	3523	0.246	-0.007 (0.021)	0.008 (0.007)
4-year college	3523	0.245	-0.068** (0.023)	0.026*** (0.004)
Competitiveness	843	0.944	-0.064 (0.063)	0.016 (0.011)
Part-time status	3523	0.275	-0.050 (0.025)	0.014* (0.006)
Full-time status	3523	0.289	-0.061** (0.022)	0.029*** (0.006)
In-state college	3523	0.365	-0.061** (0.019)	0.028** (0.008)
Out-of-state college	3523	0.098	-0.022 (0.020)	0.002 (0.003)
TDP eligible college	1413	0.832	-0.001 (0.042)	0.005 (0.006)
TDP ineligible college	3523	0.069	-0.010 (0.018)	0.001 (0.003)
College quality (grad. rate)	1390	0.447	-0.036* (0.017)	0.010** (0.003)
Persistence (Fall 2015-Spring 2019)	3523	0.114	-0.026 (0.013)	0.010** (0.003)
Undermatched college entry	735	0.071	0.018 (0.031)	-0.003 (0.005)

Notes: Estimates based on Model 4 with non-linear functional form for lagged covariates (see prior tables for details). Students self-reported number of communications about TDP, and the number of times an adult in the school spoke with them about the project, which we summed across the two measures and across four years. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix E. Bootstrap Standard Errors

The main text explains that the usual clustered standard errors may not satisfy the usual asymptotic assumptions with only 36 clusters. As an alternative, we also estimated standard errors via a modified bootstrap method that accounts for the paired randomization design. The point estimates for Tables E1-E4 are identical to the main text (Tables 3-6), except for the alternative standard errors. We note that the bootstrap routine would not converge with Model 4.

These tables also include the R^2 for each estimate, which apply to both these tables and Tables 3-6.

Table E1: Average Treatment Effects on High School Academic Outcomes
(Bootstrap Standard Errors)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Attendance pct	5038	0.813	0.004 (0.011) {0.0007}	-0.003 (0.005) {0.5982}	-0.004 (0.006) {0.1826}	n/a n/a n/a
90% or above	5038	0.469	0.002 (0.024) {0.0011}	-0.011 (0.013) {0.4023}	-0.010 (0.019) {0.1833}	n/a n/a n/a
GPA	5038	1.801	-0.006 (0.047) {0.0018}	-0.011 (0.051) {0.2444}	-0.009 (0.064) {0.1467}	n/a n/a n/a
2.5 or above	5038	0.263	-0.009 (0.015) {0.0016}	-0.010 (0.017) {0.1890}	-0.008 (0.017) {0.0608}	n/a n/a n/a
Meets Both DP reqs.	5038	0.228	-0.008 (0.016) {0.0014}	-0.014 (0.017) {0.2529}	-0.006 (0.037) {0.0188}	n/a n/a n/a
Math MAP	5038	-0.057	0.085* (0.034) {0.0000}	0.039 (0.024) {0.7425}	0.002 (0.032) {0.1953}	n/a n/a n/a
Read MAP	5038	-0.021	0.026 (0.054) {0.0008}	0.012 (0.032) {0.6694}	-0.014 (0.052) {0.1948}	n/a n/a n/a
Transferred schools	5038	0.407	0.057 (0.068) {0.0026}	0.058 (0.064) {0.0985}	0.108 (0.108) {0.0795}	n/a n/a n/a
Missing in 2014-15 data	5038	0.261	0.001 (0.029) {0.0000}	0.001 (0.026) {0.0900}	-0.004 (0.070) {0.0543}	n/a n/a n/a
Grad On Time w/ Reg Diploma	5038	0.505	0.006 (0.025) {0.0002}	-0.001 (0.021) {0.1974}	-0.018 (0.041) {0.1283}	n/a n/a n/a
Grad w/ Any Credential, Anytime	5038	0.550	0.019 (0.023) {0.0000}	0.013 (0.018) {0.1774}	-0.001 (0.048) {0.1203}	n/a n/a n/a
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: See notes for Table 3. The main difference between Table E1 and Table 3 is the switch to school cluster robust standard errors (in parentheses) for inference. The other difference is that we report the R^2 in {brackets}. The results from models 1-3 are presented. Model 4 would not converge with bootstrapped standard errors.

Table E2: Average Treatment Effects on College Expectations and Non-Merit Effort
(Bootstrap Stand Errors)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
College expectations						
Planning to attend college	5038	0.680	0.041 (0.022) {0.0000}	0.041 (0.023) {0.0335}	0.023 (0.074) {0.0196}	n/a n/a n/a
Planning on 4-year college	5038	0.465	0.025 (0.033) {0.0010}	0.022 (0.031) {0.1047}	0.008 (0.076) {0.0480}	n/a n/a n/a
Planning on 2-year college	5038	0.216	0.016 (0.025) {0.0015}	0.019 (0.025) {0.0437}	0.015 (0.067) {0.0037}	n/a n/a n/a
Planning full-time college	5038	0.565	0.055 (0.028) {0.0001}	0.054 (0.029) {0.0666}	0.037 (0.075) {0.0089}	n/a n/a n/a
Steps to College						
# college support prog participated in	5038	0.463	0.149 (0.089) {0.0096}	0.149 (0.091) {0.0155}	0.056 (0.187) {0.0044}	n/a n/a n/a
# colleges applied to	5038	2.354	0.309* (0.136) {0.0004}	0.304* (0.138) {0.0411}	0.262 (0.294) {0.0217}	n/a n/a n/a
First choice college is highest prestige	5038	0.140	-0.029 (0.026) {0.0083}	-0.029 (0.025) {0.0859}	-0.027 (0.042) {0.0537}	n/a n/a n/a
Financial aid and college cost (senior year)						
Completed FAFSA	5038	0.519	0.044 (0.023) {0.0000}	0.043 (0.023) {0.0518}	0.029 (0.071) {0.0298}	n/a n/a n/a
Applied for scholarships	5038	0.596	0.079* (0.039) {0.0018}	0.074* (0.037) {0.0606}	0.046 (0.218) {0.0107}	n/a n/a n/a
Awarded scholarships	5038	0.393	0.066* (0.028) {0.0027}	0.061* (0.024) {0.0474}	0.077 (0.108) {0.0005}	n/a n/a n/a
Intends to use scholarships	5038	0.589	0.095*** (0.020) {0.0034}	0.086*** (0.017) {0.0676}	0.081 (0.095) {0.0165}	n/a n/a n/a

Appendix E: Bootstrap

Biggest roadblock to college is cost	5038	0.557	-0.001 (0.021) {0.0000}	-0.002 (0.020) {0.0236}	-0.006 (0.079) {0.0012}	n/a n/a n/a
Summer Melt						
Planned on college, but did not enroll	5038	0.528	0.047 (0.027) {0.0002}	0.046 (0.029) {0.0905}	0.047 (0.050) {0.0084}	n/a n/a n/a
Planned on 4-year college, but enrolled in 2-year college	5038	0.135	0.021 (0.023) {0.0001}	0.021 (0.023) {0.0080}	0.044 (0.034) {0.0007}	n/a n/a n/a
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: See notes for Tables 3 and 4. The main difference between Table E2 and Table 4 is the switch to school cluster robust standard errors (in parentheses) for inference. The other difference is that we report the R^2 in {brackets}. The results from models 1-3 are presented. Model 4 would not converge with bootstrapped standard errors.

Table E3: Average Treatment Effects on Initial College Enrollment
(Bootstrap Standard Errors)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Any College Enrollment	5038	0.334	0.005 (0.022) {0.0004}	-0.004 (0.019) {0.2121}	-0.016 (0.038) {0.1353}	n/a n/a n/a
2-year college	5038	0.198	0.012 (0.014) {0.0000}	0.007 (0.011) {0.0543}	0.001 (0.029) {0.0280}	n/a n/a n/a
4-year college	5038	0.188	-0.004 (0.024) {0.0008}	-0.009 (0.023) {0.2260}	-0.017 (0.048) {0.1335}	n/a n/a n/a
Part-time status	5038	0.214	-0.002 (0.023) {0.0003}	-0.008 (0.021) {0.1075}	-0.020 (0.028) {0.0535}	n/a n/a n/a
Full-time Status	5038	0.224	0.009 (0.022) {0.0001}	0.003 (0.020) {0.2163}	-0.011 (0.040) {0.1355}	n/a n/a n/a
In-state college	5038	0.290	0.008 (0.017) {0.0002}	0.001 (0.011) {0.1897}	-0.014 (0.026) {0.1264}	n/a n/a n/a
Out-of-state college	5038	0.074	-0.012 (0.020) {0.0009}	-0.013 (0.020) {0.0207}	-0.013 (0.029) {0.0003}	n/a n/a n/a
TDP eligible college	5038	0.828	0.018 (0.042) {0.0000}	0.011 (0.044) {0.0474}	-0.032 (0.106) {0.0249}	n/a n/a n/a
Coll. Competitiveness	5038	0.929	0.015 (0.085) {0.0029}	0.004 (0.061) {0.2191}	-0.025 (0.296) {0.1298}	n/a n/a n/a
Coll. grad. Rate	5038	0.440	-0.011 (0.017) {0.0040}	-0.014 (0.018) {0.2053}	-0.010 (0.040) {0.0885}	n/a n/a n/a
Undermatched college entry	5038	0.079	0.017 (0.029) {0.0001}	0.024 (0.020) {0.2183}	-0.001 (0.046) {0.0271}	n/a n/a n/a
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: See notes for Tables 3 and 5. The main difference between Table E3 and Table 5 is the switch to cluster robust standard errors (in parentheses) for inference. The other difference is that we report the R^2 in {brackets}. The results from models 1-3 are presented. Model 4 would not converge with bootstrapped standard errors.

Table E4: Average Treatment Effects on College Graduation and Life Outcomes
(Bootstrap Standard Errors)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
College graduation						
	5038	0.020	0.005 (0.004) {0.0001}	0.004 (0.003) {0.0095}	0.003 (0.009) {0.0114}	n/a n/a n/a
Ever graduated from college						
	5038	0.012	0.005 (0.003) {0.0002}	0.005 (0.003) {0.0070}	0.003 (0.005) {0.0071}	n/a n/a n/a
Ever graduated from 2-year college						
	5038	0.007	0.000 (0.004) {0.0000}	-0.001 (0.004) {0.0024}	0.000 (0.006) {0.0062}	n/a n/a n/a
Ever graduated from 4-year college						
Employment outcomes						
	5037	0.602	0.011 (0.020) {0.0002}	0.007 (0.021) {0.0346}	0.015 (0.040) {0.0000}	n/a n/a n/a
Ever employed						
	5037	0.719	0.007 (0.013) {0.0000}	0.001 (0.014) {0.0998}	0.004 (0.034) {0.0251}	n/a n/a n/a
Ever employed or enrolled in college						
	5037	0.639	0.005 (0.014) {0.0000}	-0.001 (0.015) {0.0766}	0.002 (0.035) {0.0136}	n/a n/a n/a
Employed or enrolled in college in 2017						
Earnings						
	5037	5.061	0.140 (0.207) {0.0004}	0.093 (0.209) {0.0364}	0.162 (0.352) {0.0005}	n/a n/a n/a
Full sample						
	5037	5.122	0.111 (0.159) {0.0003}	0.074 (0.159) {0.0434}	0.161 (0.343) {0.0002}	n/a n/a n/a
Excluding college enrollees in 2017						
	5037	0.033	0.005 (0.004) {0.0157}	0.007 (0.004) {0.0395}	0.014*** (0.003) {0.0155}	n/a n/a n/a
Ever incarcerated						
	5037	0.018	-0.002 (0.005) {0.0000}	-0.002 (0.005) {0.0130}	0.003 (0.016) {0.0036}	n/a n/a n/a
Ever had child during High School						
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: See notes for Tables 3 and 6. The main difference between Table E3 and Table 6 is the switch to cluster robust standard errors (in parentheses) for inference. The other difference is that we report the R^2 in {brackets}. The results from models 1-3 are presented. Model 4 would not converge with bootstrapped standard errors.

Appendix F. Local Average Treatment Effects

This section presents two different types of Local Average Treatment Effects (LATE) focused on the effects of receiving TDP funds. The first is the IV-TOT and the second is the RD. See the main text for discussion of the assumptions.

Table F1 focuses on key college and life outcomes from Table 3C and Table 3D from the main text. (The IV-TOT requires that students receive the money, so this method cannot be applied to the high school outcomes.) The estimated effects for two-year college graduation remain statistically significant. Also, by construction, the other coefficients cannot be significant because they were insignificant in the ITT.

Table F1: Treatment on Treated (TOT) Effects on College Outcomes

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Panel A. College outcomes						
College enrollment						
Ever enrolled in college	3601 - 4948	0.841	0.029 (0.090)	-0.031 (0.081)	-0.083 (0.068)	0.003 (0.075)
Ever enrolled in 2-year college	3601 - 4948	0.304	0.071 (0.054)	0.029 (0.047)	0.012 (0.059)	0.043 (0.048)
Ever enrolled in 4-year college	3601 - 4948	0.688	-0.022 (0.093)	-0.042 (0.103)	-0.096 (0.076)	0.078 (0.096)
College graduation						
Ever graduated from college	3601 - 4948	0.063	0.027 (0.015)	0.028 (0.019)	0.019 (0.017)	0.021 (0.019)
Ever graduated from 2-year college	3601 - 4948	0.037	0.030*** (0.008)	0.035*** (0.010)	0.020** (0.007)	0.030** (0.011)
Ever graduated from 4-year college	3601 - 4948	0.026	-0.002 (0.015)	-0.007 (0.013)	-0.000 (0.013)	-0.009 (0.013)
Panel B. Life outcomes						
Employment outcomes						
Ever employed	3601 - 4948	0.567	0.071 (0.074)	0.043 (0.082)	0.092 (0.065)	0.043 (0.061)
Ever employed or enrolled in college	3601 - 4948	0.948	0.051 (0.052)	0.010 (0.058)	0.025 (0.060)	0.034 (0.035)
Employed or enrolled in college in 2017	3601 - 4948	0.866	0.038 (0.053)	-0.009 (0.053)	0.019 (0.074)	0.034 (0.038)
Earnings (full sample; IHS)	3601 - 4948	4.870	0.877 (0.705)	0.122 (0.757)	0.919 (0.564)	0.596 (0.720)
Ever incarcerated	3601 - 4948	0.000	-0.008 (0.012)	0.005 (0.012)	-0.017** (0.006)	0.016 (0.013)
Ever had child during High School	3601 - 4948	0.002	-0.011 (0.020)	-0.005 (0.012)	0.029 (0.021)	-0.013 (0.016)
Lagged Dependent Variable (i)				X		
Lagged Dependent Variable (j)					X	X
School Covariates (i)						X

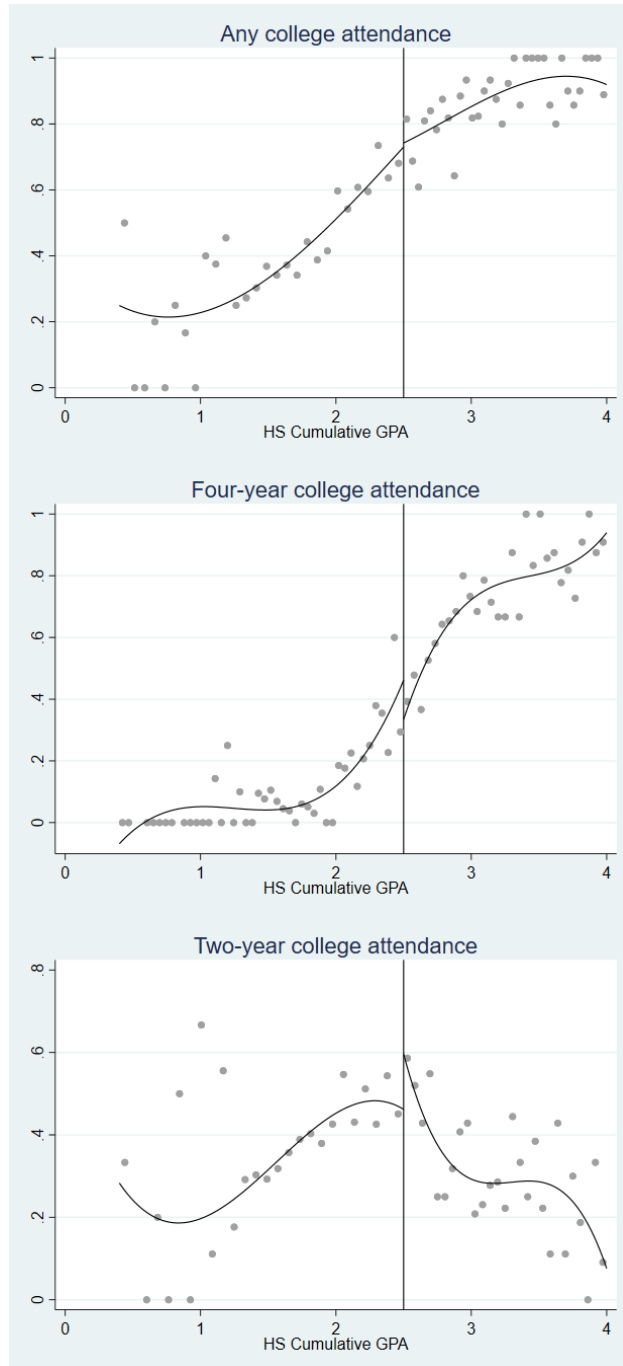
Notes: See notes for Table 3A and Table 3D. Treatment is redefined here as being sent a congratulatory letter at the end of high school, indicating that students met the requirements. We estimated the treated-on-treated (TOT) using assignment to treatment as an instrumental variable. See text discussion regarding possible violations of the exclusion restriction in this case. The baseline means differ from the prior tables and are calculated from the control group compliers (i.e., those who met the performance requirements). We note that the ever-enrolled in two- and four-year college numbers sum to a number greater than the total ever-enrolled number because of transfers across sectors.

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Randomization provides evidence of average marginal effects and the distribution of those effects across the GPA range. GPA can also be viewed as a forcing variable in an RD framework. While we do not see this as informative for the reasons outlined in the main text, Figure C1 shows the results of a basic RD analysis where the outcomes are various forms of college attendance, based on Quan and Harris (2020). These use the Model 4 covariates.

We focus on GPA as the forcing variable because this is much more likely than attendance to be a binding constraint (i.e., very few students met the GPA requirement but did not meet the attendance requirement or others). Since this is an RD, identification is from within-treatment group variation, which has the further implication that pair effects drop out; we replace these with school fixed effects. The forcing variable is *cumulative* GPA during high school, as opposed to the baseline GPA used in many of the other figures in this study. See Quan and Harris (2020) for McCrary test results and optimal bandwidth variations, which yield somewhat erratic results.

Figure F1: Test of Aid Targeting: Effect Heterogeneity by Baseline GPA on College Outcomes



Appendix G. Additional Tests Related to Merit Requirements

G.1. Targeting

Figure 1 in the main text show effects by GPA and provides additional evidence that treatment effects do not vary on this dimension. Tables G1-G3 reinforce these findings using all four models and interaction terms between treatment and GPA. Since we are mainly interested here in how the *receipt* of funding affected outcomes, we restricted the sample to students who were already above the performance thresholds at baseline, and then estimated the effects with interaction terms to test whether students with higher GPA and test scores saw larger effects. (This reduces the samples sizes, as noted at the bottom of each table.)

We focus on treatment interactions with GPA and test scores because these are the most commonly used merit metrics (and GPA is used with TDP). We report estimates using regression model 4, which includes pair effects, school-level covariates, and other controls.

These results reinforce the finding from the main text that merit requirements do not target funds to students who are more likely to respond to aid. They show a mix of positive and negative effects and are imprecisely estimated.

Table G1: Effect heterogeneity by baseline performance for initially qualified, high school outcomes

	TreatedXGPA	TreatedXMath	TreatedXReading
Attendance pct	0.030 (0.015)	0.012 (0.010)	0.012 (0.013)
90% or above	0.061 (0.061)	0.017 (0.024)	0.036 (0.023)
GPA	0.124 (0.113)	0.058 (0.068)	0.134* (0.069)
2.5 GPA or above	0.009 (0.057)	-0.001 (0.038)	0.032 (0.039)
Meets Both DP req.s	0.027 (0.073)	0.000 (0.039)	0.021 (0.046)
Math MAP	0.687 (0.386)	-0.206* (0.098)	-0.156 (0.143)
Read MAP	-0.139 (0.341)	-0.100 (0.121)	0.109 (0.107)
Transferred schools	-0.039 (0.067)	0.001 (0.034)	-0.002 (0.034)
Missing in 2014-15 data	-0.031 (0.036)	-0.012 (0.020)	0.015 (0.025)
Grad On Time w/ Reg Diploma	0.073* (0.034)	0.043* (0.022)	0.014 (0.028)
Grad w/ Any Credential, Anytime	0.061 (0.044)	0.049 (0.030)	0.034 (0.033)
N range	237 - 1060	237 - 1060	236 - 1054
Pair Indicators (j)	X	X	X
Baseline Performance (i)	X	X	X
Student Covariates (i)	X	X	X

Notes. See notes to Table 3A. Sample is conditioned to individuals with 8th grade GPAs and attendance rates at or above the TDP threshold (2.45 and 90%, respectively). Significance levels: * p<0.05, ** p<0.01, *** p<0.001

Table G2: Effect heterogeneity by baseline performance for initially qualified, college entrance outcomes

	TreatedXGPA	TreatedXMath	TreatedXReading
Any College Enrollment	-0.023 (0.039)	-0.006 (0.033)	-0.021 (0.047)
2-year college	0.007 (0.050)	0.020 (0.030)	0.015 (0.047)
4-year college	-0.051 (0.057)	-0.029 (0.031)	-0.035 (0.035)
Competitiveness	0.011 (0.121)	-0.027 (0.063)	-0.148 (0.080)
Part-time status	-0.030 (0.061)	-0.046 (0.037)	-0.073 (0.043)
Full-time status	-0.045 (0.070)	0.006 (0.040)	-0.001 (0.040)
In-state college	-0.026 (0.052)	0.015 (0.032)	0.001 (0.041)
Out-of-state college	-0.031 (0.030)	-0.003 (0.017)	-0.003 (0.029)
TDP eligible college	-0.056 (0.048)	0.050 (0.028)	0.064 (0.047)
College quality (grad. rate)	0.012 (0.025)	-0.030 (0.020)	-0.024 (0.025)
Undermatched college entry	-0.036 (0.087)	0.038 (0.029)	0.033 (0.037)
N range	409 - 1060	409 - 1060	407 - 1054
Pair Indicators (j)	X	X	X
Baseline Performance (i)	X	X	X
Student Covariates (i)	X	X	X

Notes. See notes to Table 3A. Sample is conditioned to individuals with 8th grade GPAs and attendance rates at or above the TDP threshold (2.45 and 90%, respectively). Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G3: Effect heterogeneity by baseline performance for initially qualified, long run outcomes

	TreatedXGPA	TreatedXMath	TreatedXReading
College graduation			
Ever graduated from college	0.010 (0.021)	0.018 (0.010)	0.007 (0.009)
Ever graduated from 2-year college	0.007 (0.017)	0.021 (0.009)	0.012 (0.007)
Ever graduated from 4-year college	0.003 (0.013)	-0.003 (0.006)	-0.005 (0.007)
Employment outcomes			
Ever employed	-0.001 (0.059)	0.002 (0.032)	0.010 (0.041)
Ever employed or enrolled in college	0.012 (0.026)	-0.005 (0.025)	-0.038 (0.024)
Employed or enrolled in college in 2017	-0.022 (0.044)	-0.029 (0.031)	-0.053 (0.031)
Earnings			
Full sample	0.094 (0.558)	-0.093 (0.393)	0.038 (0.424)
Excluding college enrollees in 2017	0.204 (0.833)	-0.091 (0.470)	-0.243 (0.416)
Ever incarcerated	-0.022 (0.015)	-0.018* (0.007)	-0.004 (0.010)
Ever had child during High School	-0.003 (0.013)	0.009 (0.005)	0.007 (0.005)
N Range	699 - 1060	699 - 1060	695 - 1054
Pair Indicators (j)	X	X	X
Baseline Performance (i)	X	X	X
Student Covariates (i)	X	X	X

Notes. See notes to Table 3A and Table 3D. Sample is conditioned to individuals with 8th grade GPAs and attendance rates at or above the TDP threshold (2.45 and 90%, respectively). Significance levels: * p<0.05, ** p<0.01, *** p<0.001

G.2. Disappointment Effects

One theory not mentioned above is that people might get disappointed because they do not meet some expectation or fall short of reaching a threshold that is tied to some form of compensation. In this case, the specific concern is that some initially eligible students might have gotten their hopes up only to fall below the merit threshold, become ineligible, and lose the expected aid.

Testing for disappointment effects is complicated by the fact we cannot directly observe people's expectations to reasonably identify who might be disappointed. The TDP experiment offers an excellent test because the control group were never really aware of the program, so we can reasonably assume they experienced no disappointment as a result of TDP. We test for disappointment in three steps: (1) identify students in both the control and treatment groups who were just above the merit thresholds at baseline; (2) among students in both groups, identify those who experienced declines in the range 0-0.3 GPA points immediately post-treatment; specifically, at the end of the freshman year; and (3) test for differences between the two groups in their GPA changes in the following (sophomore) year. Step (2) helps ensure that TDP did not induce some other response, other than disappointment, that makes the treatment group different from the control group. To be clear, the treatment group in this analysis initially received a letter saying they were eligible then received a second letter, in the summer between the freshman and sophomore years saying they were no longer eligible (at least temporarily since they had a chance in the sophomore year to raise the GPA back up again). We call this the "above-then-below" group.

Table A1 shows the results. It appears that the above-then-below treatment group experienced a larger decline in GPA in the second year (-0.1264) than the control group that was otherwise similar in their baseline GPAs and freshmen year GPA trajectories (see the first row), though this difference is imprecise. We also see no evidence of the same pattern when we look at control and treatment group students who were just below the threshold (see the second row), reinforcing that the pattern is driven by something happening just to the treatment group that is just above the threshold to start with.

Table G4: Test for Disappointment Effects

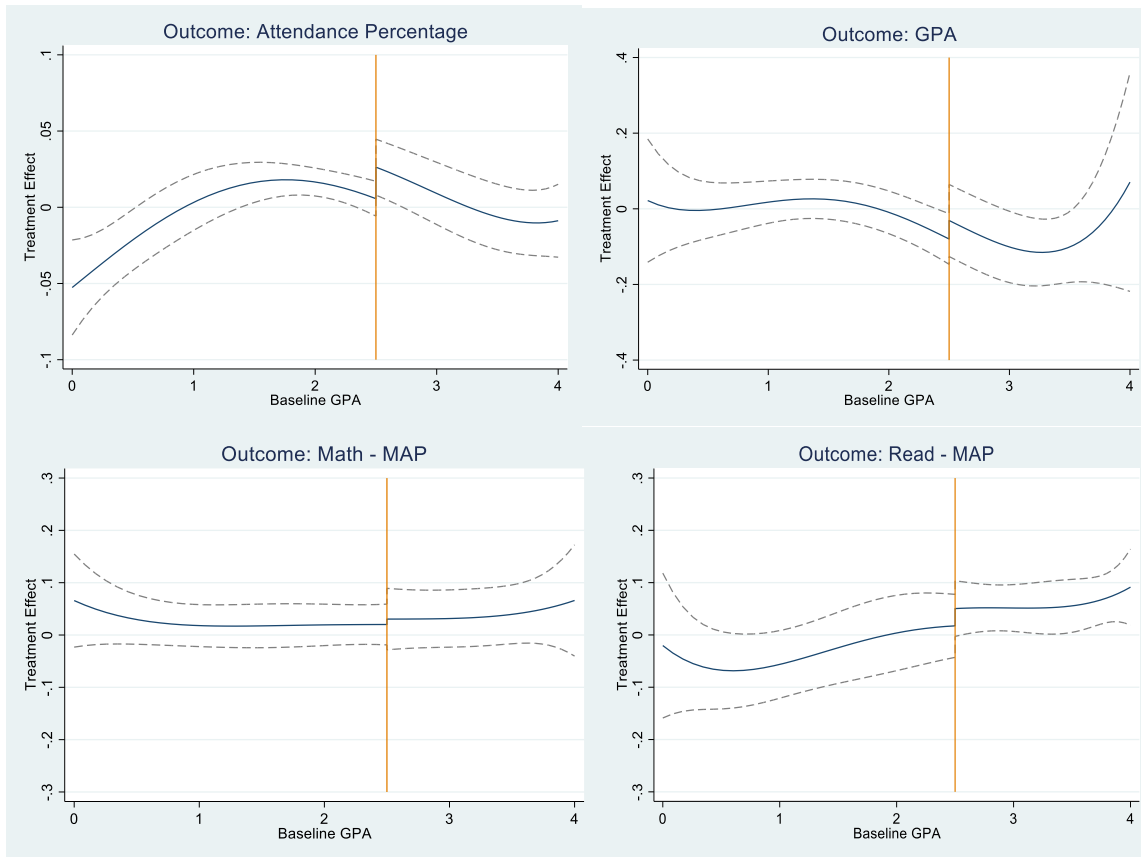
	Treatment	Control	Diff. (T-C)
Eligible	-0.1475	-0.0211	-0.1264 (0.0887)
Ineligible	0.0512	0.0558	-0.0046 (0.0289)
Diff (E-I)	-0.1987 (0.0670)	-0.0769 (0.0649)	-0.1218 (0.0933)

Notes: Robust standard errors in parentheses.

Another way to test for disappointment is to look within the treatment group and compare those who were initially barely eligible with those who were initially barely ineligible. Since the thresholds were arbitrary, students' could not have responded to them in advance. Looking down the first column of Table A1, we see again that the initially eligible group experienced a larger decline in the sophomore year than students who were initially ineligible; this difference of -0.1987 GPA points is also fairly precise. Moreover, we see a smaller drop in the control group (see the second column). This test is less convincing than the one above because it could be that students just below the threshold were motivated to work harder to get above the threshold, which is obviously unrelated to disappointment.

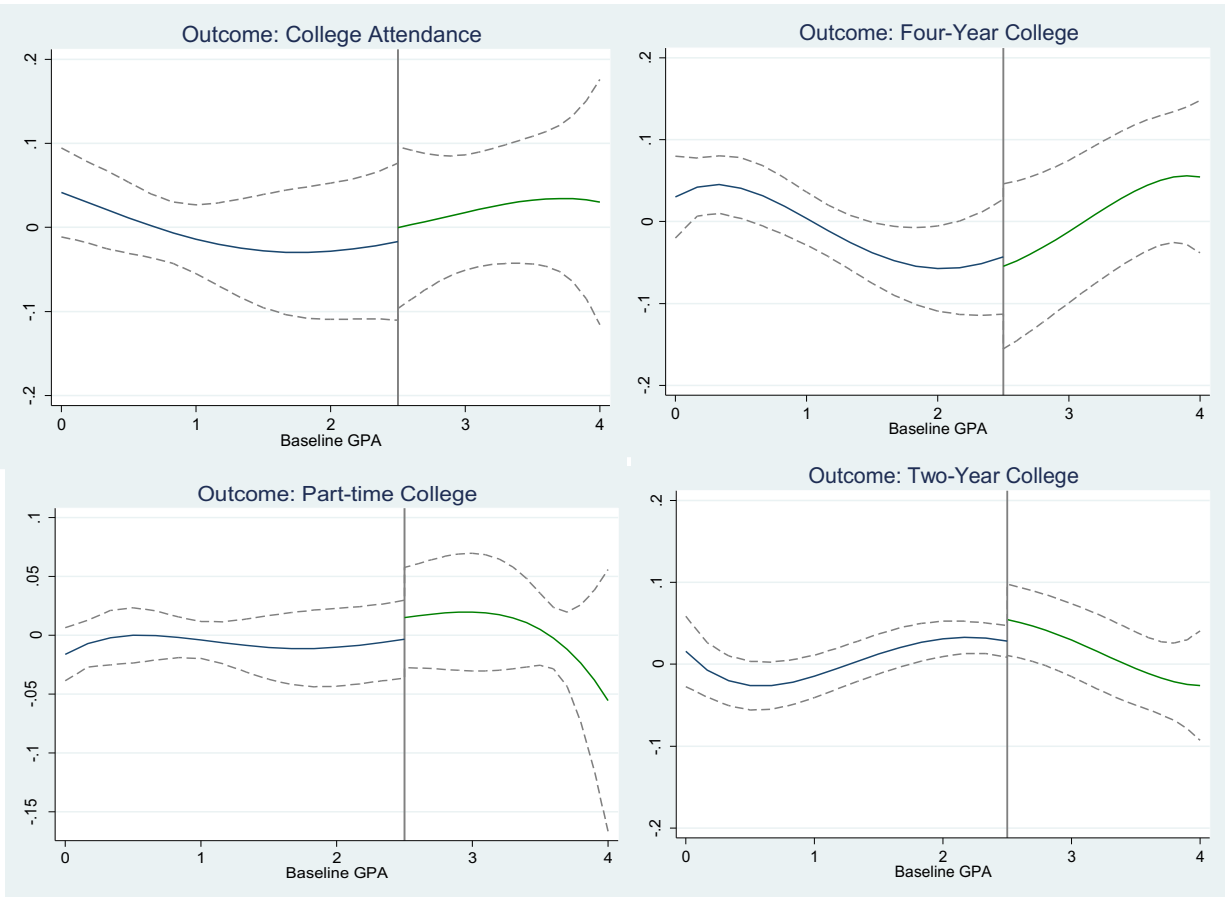
An alternative interpretation of the above patterns is that the merit requirements created an anchoring effect. This can occur when people use irrelevant information to guide decisions. In this case, students above the threshold might have reduced their effort because the 2.5 GPA threshold led them to believe they were doing more than necessary to prepare for college, so they reduced their effort. This is more difficult to test, however. The above test is specifically designed to capture disappointment because it is limited to those students who were above the bar and then fell below.

*Figures G1: Test of Merit-Based Incentive Effects:
Local Linear Regression Effects by Baseline GPA on High School Academic Outcomes*



Notes: The panels of Figure 1A test the standard theory that the effort response function peaks near the threshold and is null in the tails (the inverse-U shape show in Figure 1A and *Hypothesis 4*). The academic endowment in this case is baseline GPA. (While TDP entailed other merit requirements, GPA was most likely to be binding on students.) As in all the effect heterogeneity analyses, these are based on the Model 4 regression with the full set of covariates, while also allowing the regression line to differ on either side of the merit threshold. 95% confidence intervals are shown.

*Figure G2: Test of Merit-Based Aid Targeting:
Local Linear Regression Effects on College Outcomes by Baseline GPA*

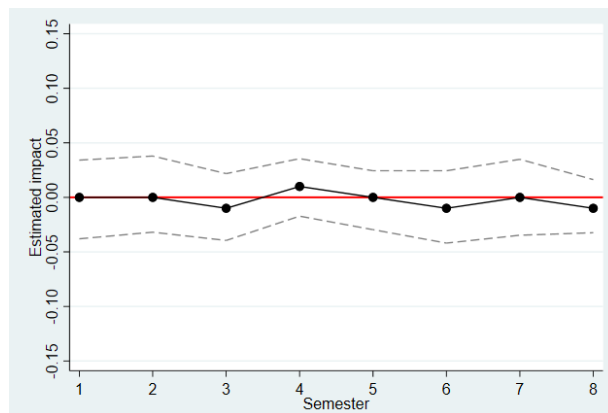


Notes: The panels of Figure 1B test the theory that the merit requirements target aid to students who respond to it most, which would arise if the effects were increasing in baseline GPA, especially to the right of the threshold. See additional notes in Figure 1A. 95% confidence intervals are shown.

Appendix H. Additional Effect Heterogeneity

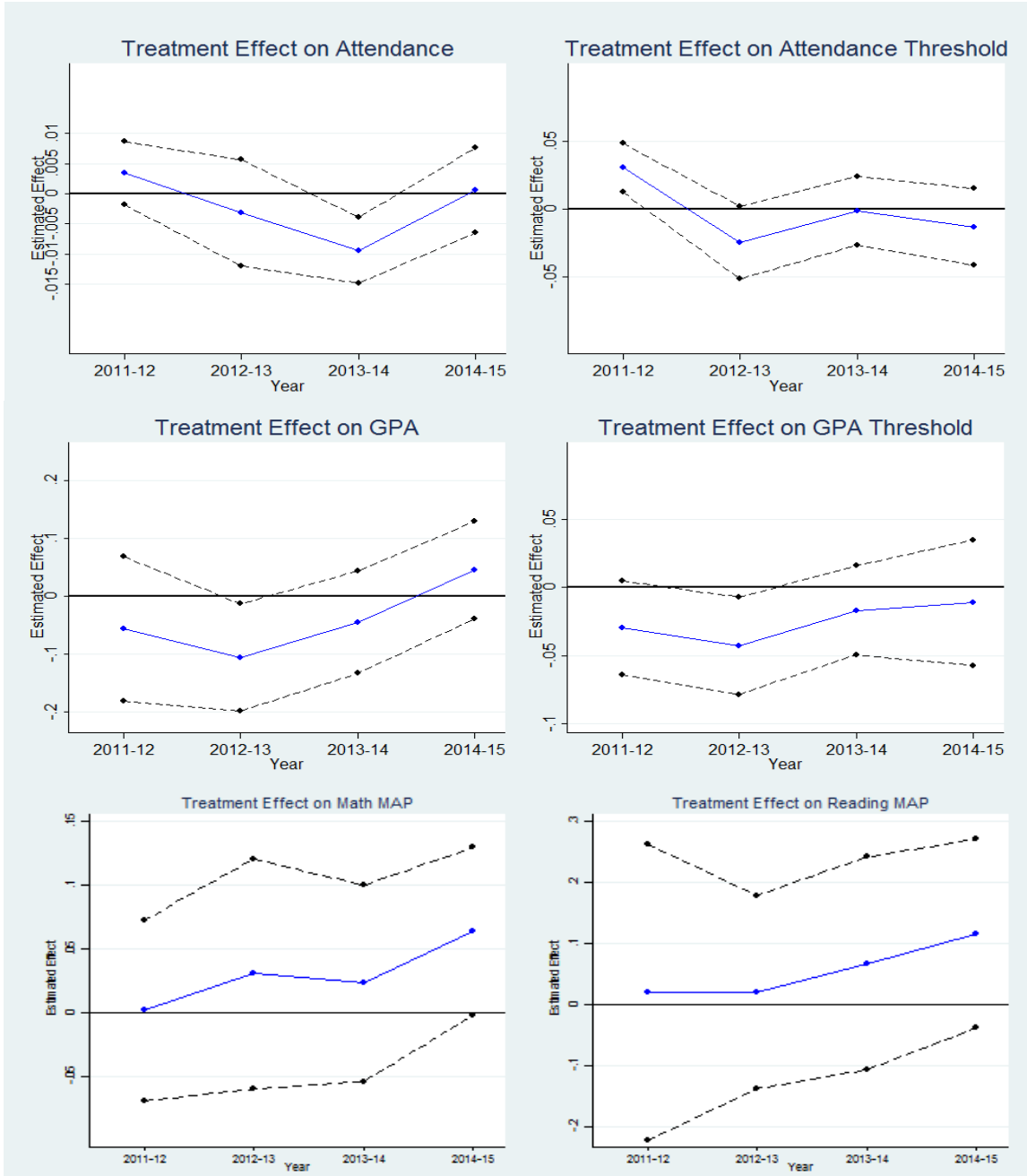
In this section, we provide additional evidence regarding effect heterogeneity. Figure H1 provides evidence about effects on college enrollment by semester. Figure H2 provides effects by grade during high school. Figure H3-H5 provides effects by baseline GPA and provide evidence about incentive and targeting effects. All results in this section are based on regression model 4 from the main text.

Figure H1. Enrollment Effects by Semester, Post-High School



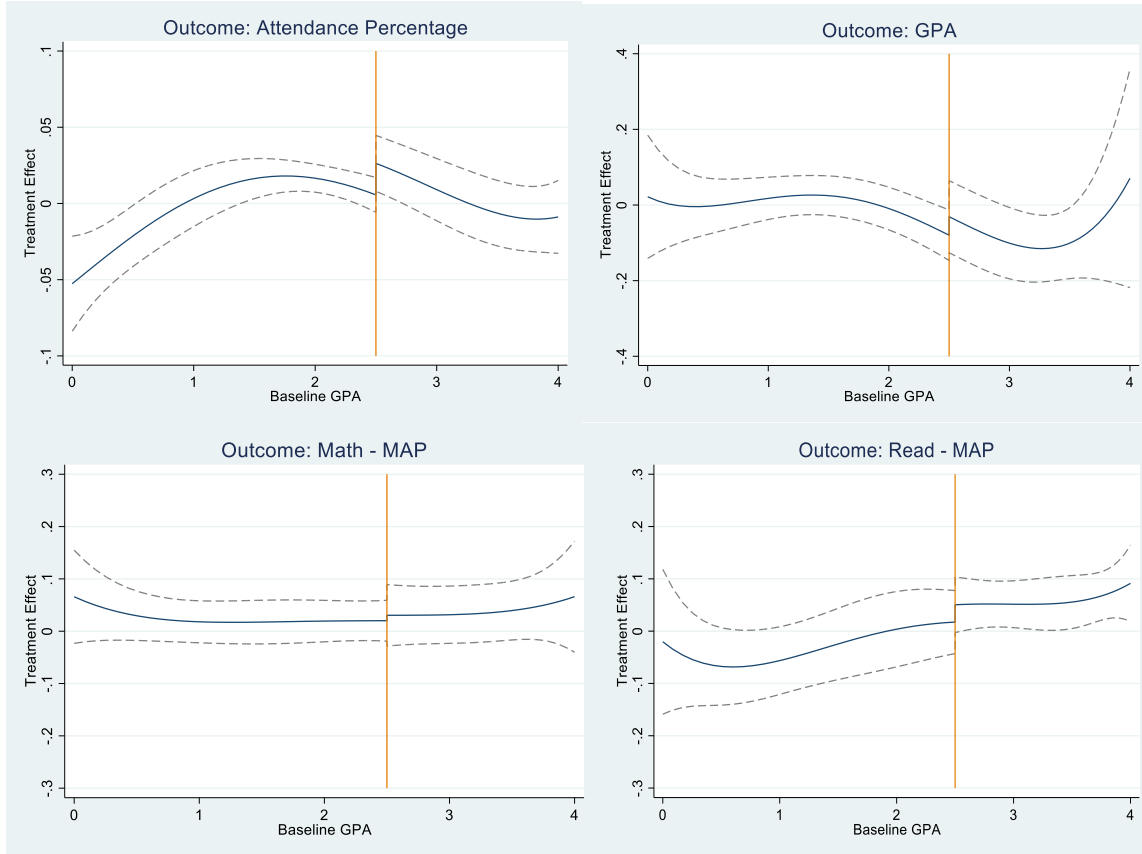
Notes: The solid black lines represent the estimated effect; gray dashed lines are 95% confidence intervals. Control group averages: Fall 2015 = 0.222; Spring 2016 = 0.211; Fall 2016 = 0.209; Spring 2017 = 0.181; Fall 2017 = 0.188; Spring 2018 = 0.177; Fall 2018 = 0.156; Spring 2019 = 0.142.

Figure H2: Treatment Effects by Grade



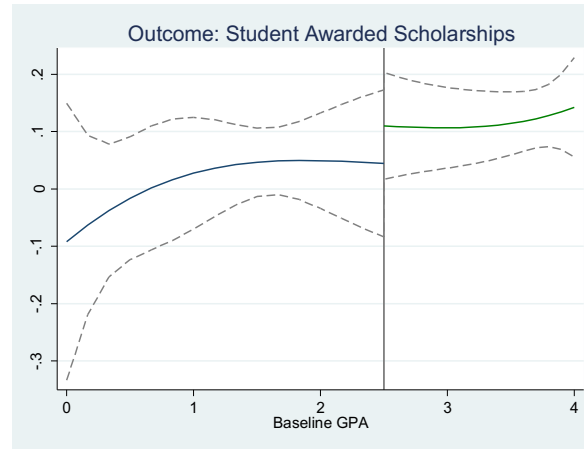
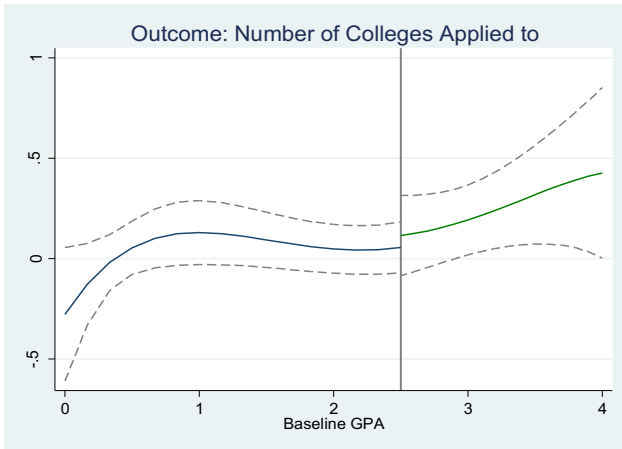
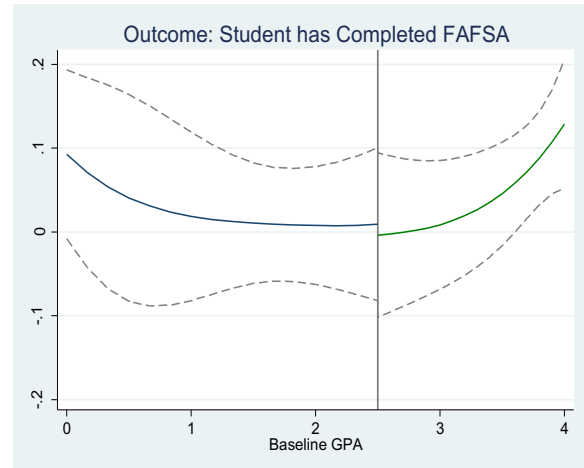
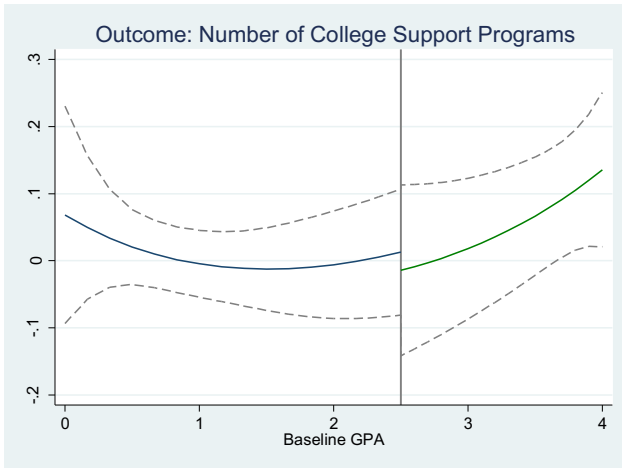
Notes: Figure 2 provides estimates on high school effort by grade as a test of the standard theory that discounting will lead students to delay their effort (see Figure 1B). These are based on the model 4 regression with a full set of covariates. 95% confidence intervals are shown.

*Figures H3: Test of Merit-Based Incentive Effects:
Local Linear Regression Effects by Baseline GPA on High School Academic Outcomes*

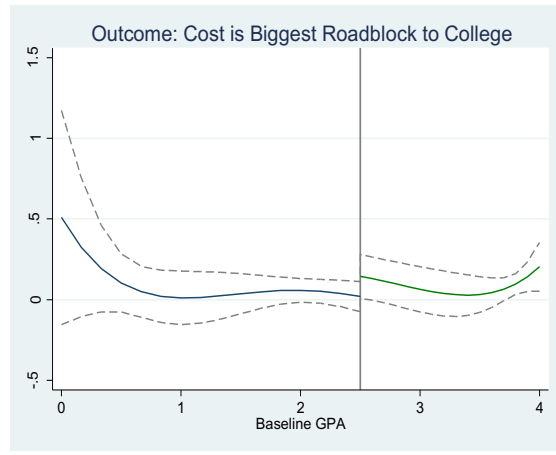
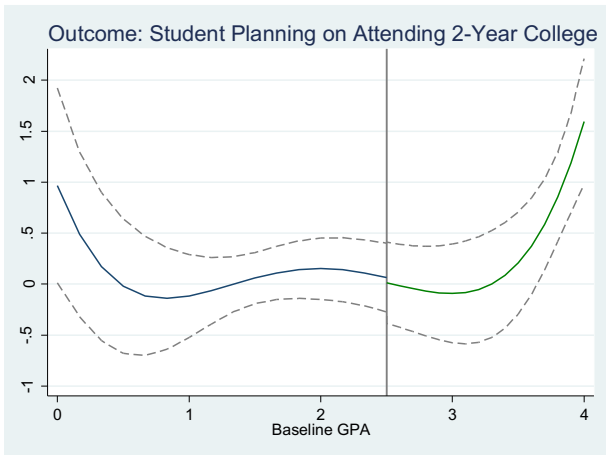
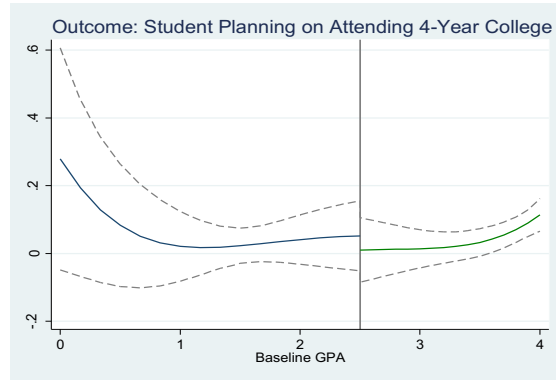
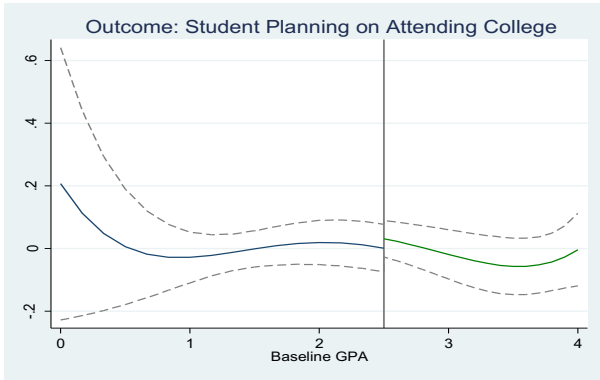


Notes: The panels of Figure 1A test the standard theory that the effort response function peaks near the threshold and is null in the tails (the inverse-U shape show in Figure 1A and *Hypothesis 4*). The academic endowment in this case is baseline GPA. (While TDP entailed other merit requirements, GPA was most likely to be binding on students.) As in all the effect heterogeneity analyses, these are based on the Model 4 regression with the full set of covariates, while also allowing the regression line to differ on either side of the merit threshold. 95% confidence intervals are shown.

*Figures H4: Effect Heterogeneity by Baseline GPA
Other Steps to College Taken During High School*



*Figures H5: Effect Heterogeneity, by Baseline GPA
College Expectations during High School*



Appendix I: Cost-Benefit Analysis Details

This section provides additional detail on the cost-benefit analysis in section V.C. of the main text. The benefits and costs of financial aid are characterized by:

$$\text{Benefits: } \beta_{2y} \sum_{t=1}^{45} (1 - \delta)^t w_{2y} + \beta_{4y} \sum_{t=1}^{45} (1 - \delta)^t w_{4y}$$

$$\text{Costs: } \beta_{2y} p_{2y} + \beta_{4y} p_{4y} + 100\eta g$$

where η is still the marginal cost of funds; g is the grant amount; β_{2y} and β_{4y} are the percentage point changes in degrees created by aid program, for two- and four-year degrees, respectively; and p_{2y} and p_{4y} are the resource costs for each type of degree. (We multiply the right-hand cost term by 100 because the grant effects are scaled such that a 10 percentage point effect means $\beta_{y} = 10$. In that case, 10 students benefit from the grant, but the marginal cost of funds is incurred for all 100 students who received the grant.)

The resource cost estimates come from the USDOE-NCES college spending data. We take the total relevant expenditures in each sector and dividing by the number of full-time equivalent (FTE) students. This includes the following cost categories: instructional, academic support, student services, and institutional support. Room and board are omitted because, for most students, similar costs are incurred regardless of whether students attend college.¹

For the most year available (the 2017-18 academic year), these costs figures were: \$10,370 (two-year) and \$22,885 (four-year) per FTE per year. Assuming 2.5/5 years of full-time enrollment for a 2/4-year degree, this yields total degree costs of: \$25,925/\$114,425. (Discounting would have a trivial effect on these cost estimates and is ignored.) Some students who receive degrees from the grant would have started college and never finished in the absence of the grant, in which cases the above cost measures would over-state the true costs. However, other students are likely induced to start college and do not finish and the costs of college attendance, short of a degree, are omitted. The above calculations assume that these effects countervailing cost effects cancel out.

¹ Even when students do live on campus (in four-year colleges), this is generally only for a one or two years, before students move to off-campus housing. While on- and off-campus room and board are usually more expensive financially than living at home with family, the opportunity costs are similar. Also, many students would move away from home even if they did not attend college. This differs from Angrist et al. (forthcoming) who use the entire cost of attendance, which includes room and board. In addition to room and board, we exclude the following college spending categories: auxiliary enterprises, hospitals, and research. The data can be found here: https://nces.ed.gov/programs/digest/d19/tables/dt19_334.10.asp.

Prior estimates for the present discounted value of earnings are surprisingly wide-ranging. For a bachelor's degree, these range from \$243,700 to \$629,000 (Belfield & Bailey, 2017). We use the mid-range value of \$436,350, which is somewhat lower than Avery and Turner (2012) who combine two- and four-year degrees; and is somewhat higher than the median major for four-year degrees reported by Webber (2016). While not shown in the equations, we also add to the cost equation the opportunity cost of attending college full-time, which we assume to be \$10,000 annually.

There are fewer estimates in the literature for the PDV of two-year degrees. Belfield and Bailey (2017) report \$94,030. However, this is lower than what is implied by Kane and Rouse (1995) who find that the return to a four-year degree is about twice as large as a two-year degree. We therefore use a PDV of \$150,000, which is also close to the median in Belfield and Bailey (2017).²

This yields the following benefit-cost ratio (BCR) formula:

$$\text{BCR: } (\beta_{2y}(\$150,000) + \beta_{4y}(\$436,350)) / (\beta_{2y}(\$25,925 + (2.5 \cdot \$10,000)) + \beta_{4y}(\$114,425 + (5 \cdot \$10,000)) + 100 \cdot 0.25 \cdot g).$$

With this, we can insert β_{2y} , β_{4y} , and g for each program and calculate separate BCRs.

We only report effects from programs when the estimates are statistically significant for either β_{2y} or β_{4y} , or both (i.e., the Wisconsin Scholars program is omitted because it had no effect on either outcome). In the base estimate, where only one of the two is significant, we count the effect of the imprecisely estimated parameter. However, we carry out robustness checks that use only the point estimates that are precisely estimated and set others to zero. Additional robustness checks include raising the MCF to 1.5 ($\eta = 0.5$) as in Heckman et al. (2010) and using the lower economic returns to college.

For two of the programs (Pittsburgh and Knox), the prior evidence does not directly measure effects on college graduation, but only initial enrollment or early persistence. However, given the evidence that the effects on graduation seem to operate through initial enrollment, this does not seem like a serious concern. In the cost-benefit analysis, we assume the enrollment/persistence effects are the same as the (unobserved) graduation effects.

² The variation in estimates, in both the two- and four-year sectors, does not appear to be due to differences in the discount rates.

The rationale for the above approach, which standardizes the costs and benefit calculations across programs, is to ensure comparability and to provide insight into the social welfare implications of taking all the programs to the scale. The analysis would be different if we were interested in the costs and benefits of each program as it is currently operated. For example, analysis of the Buffet Scholars, TDP, Kalamazoo, and Pittsburgh programs, individually, could justifiably exclude the MCF because these programs are funded philanthropically and do not require taxation. For purposes here, we are interested in the social welfare effects of scaling the programs up through government policies.

Table 4 in the main text summarizes the results. The second column summarizes the key properties of the study and program designs (see the theory in section II). The next three columns summarize the key parameters that are the basis for the BCR calculations. This is followed by our base/preferred estimate and the robustness checks. See the main text for interpretation.

These calculations rest on at least three additional assumptions: (a) the program is large enough that the average cost is a reasonable approximation of the marginal cost; (b) the return to the *average* student (as reflected in prior studies) is the same as the return for the *marginal* student induced to obtain a degree as a result of financial aid; and (c) the costs incurred by students who enter but drop out of college are exactly offset by the increased returns to education that these students receive.³ However, we note that these assumptions are applied equally to all of the programs, so that violations would affect the absolute size of all the BCRs, but likely have small effects on their relative sizes across programs. We also note that the degrees of freedom are low (i.e., too few studies for each combination of program design elements) and the standard errors of these estimates are no doubt very large (and difficult to estimate). The analysis is also partial equilibrium and neglects general equilibrium effects, e.g., political forces acting on the total resources provided, college responses regarding prices and quality, and labor market effects affecting the return to education.

³ This last assumption is due to the focus on degree attainment. The costs and benefits of aid for eventual college dropouts are omitted because they are so rarely measured.

Appendix J: Pre-Specification of the Analysis Plan

This experiment began before pre-registration of data collection and research methods was a common practice. Instead, we provide details from a grant proposal we submitted prior to the start of the experiment. This proposal was submitted to the U.S. Department of Education, Institute of Education Sciences (IES) in June of 2011. Given the funding source, the proposal is written more for an education, as opposed to economics, audience, which is reflected in some of the formatting and language (e.g., references to Hierarchical Linear Modelling). Since the proposal is quite lengthy, we provide the most pertinent text from the proposal, verbatim (reflected in quotation marks). We note that this initial proposal was rejected. The revised proposal, which was accepted for funding by IES, is similar, and also available upon request. However, we focus on the first proposal because this was submitted prior to the experiment and there are no major differences relative to the accepted proposal.

As noted in the main text, the proposal did not specify the method of estimation for the standard errors. After the proposal was submitted, we developed the paired cluster bootstrap method described in the main text to complement the Generalized Estimating Equation (GEE) standard errors, which we report in the main tables and figures.

Below is the complete text of the quantitative portion of the “Data Collection” section of the proposal. This shows that we did not change the data sources or key variables. One partial exception is that we proposed collecting FAFSA and financial aid package data, but we were only able to obtain some of these data from one college and in aggregated form. Also, the proposal suggests that the Wisconsin Department of Public Instruction would provide data on students who left Milwaukee, but we could not obtain these data.

We note that the program administrators originally planned to use the acronym “GLIMPS” to refer to the program, but later changed to “The Degree Project.” We have left in the references to GLIMPS to emphasize that the text is verbatim. Also, the references to appendices are not to the appendices of the current study, but, rather, appendices of the proposal, which are available upon request.

Data Collection Section of the Proposal

“We will collect data each year to answer our research questions. The majority of the data will come from existing administrative sources and will be de-identified (i.e., student names and other identifying information will be omitted). This approach allows for extensive analysis at a very low cost. The administrative data will be obtained from MPS (high school transcripts, surveys, test scores, and attendance and disciplinary records), the Wisconsin Department of Public Instruction (DPI) (high school transcripts for students who leave MPS), and the National Student Clearinghouse (college enrollment information). We will also seek parental consent and student assent for student interviews and collection of FAFSA and financial aid package data.

Administrative Data

MPS and DPI student records (high school academic preparation). The primary threat to validity in randomized trials is differential nonresponse/missing data. One way to avoid this problem is to obtain nearly complete administrative data. MPS has agreed to provide deidentified student records for all MPS ninth graders who remain in the MPS system. DPI has agreed to provide data on students who depart MPS but remain in public schools in Wisconsin. (See MPS and DPI letters of support in Appendix C.) The MPS and DPI data allow longitudinal linkages of individual student data using unique identifying numbers. MPS has used student identifiers for more than a decade, yielding accurate matches.

The MPS student record data include course names, grades, test scores on state standardized tests, ACT and SAT scores, attendance records, and disciplinary records. When students take the ACT, they also fill out surveys about their career interests and college plans; MPS collects these additional data from ACT and has agreed to provide them to us. The MPS course names are standardized across the district and include National Center for Education Statistics codes permitting comparisons with nationally representative data sets. MPS will add an indicator to its data system for the GLIMPS treatment status of each student, and we will cross-check this indicator against attendance data.

MPS student surveys (high school financial perceptions and social capital). MPS administers climate surveys each year to all students in grades 4-12. In 2011, the district also added a high school exit survey for students that includes many questions relevant to college-going (see latest versions in Appendix A). Surveys are administered via the Internet. The survey responses include the same student identifiers as the other MPS records. The MPS climate

surveys are based on the well-respected and research-based surveys developed and administered in the Chicago Public Schools in conjunction with the Consortium on Chicago School Research (the CCSR senior director is a consultant on this project; see Personnel section). One key construct of interest is students' peers and their interactions, as a form of social capital. The student climate survey includes items for "Students at my school focus on learning" and "Most of the students in my school are planning to go to college."

Regarding academic preparation, we are also interested in how student interactions with teachers; sample items for this include "At school I am expected to do my best all the time" and "Teachers at my school expect most students to go to college." Student responses to these questions have been shown to be strong predictors of college-going (e.g., Nagaoka et al., 2008). The high school exit survey includes questions about (a) perceptions of the expectations of, and support received from, parents, teachers, and counselors; (b) college plans, including planned course of study; (c) participation in other college access programs; and (d) parents' educational background. We are particularly interested in the items about another source of social capital: parents. Sample items include "my parents encouraged me to continue my education after high school" and "my parents talked to me about colleges/schools suited to my interests and abilities." MPS has agreed to allow us to add some questions to their surveys for this project; among these will be questions about parental expectations and our third mediating construct: perceptions about financial aid and college affordability.

MPS teacher surveys (high school social capital). One way in which GLIMPS might influence students, as noted above, is by changing interactions among students and teachers. The teacher climate survey and instructional practice survey will complement the student surveys in gauging these changes. MPS links teacher survey responses to teachers' unique identifiers. The pertinent constructs of the instructional practice surveys are (a) teacher demographics, (b) expectations for students, (c) orientation toward higher order thinking skills, and (d) student engagement. Key items include "Required students to interpret, analyze & evaluate information in their work" and "How much emphasis do you give [to] . . . tracking student progress toward expected outcomes?" The teacher climate survey includes measures of academic rigor and expectations for students' college education. Sample items include "Students at my school focus on learning" and "Teachers in this school expect most students to go to college." MPS specifies its own survey constructs, each of which is measured through multiple items.

We will conduct factor analysis to test the validity of the conceptual map and to construct new variables for the analysis. We will further develop measures through Rasch analysis so that constructs are comparable over time, have been checked for item fit and reliability, and produce standard errors of measurement for each student based on response pattern consistency.

With minimal district encouragement and no formal incentives, survey response rates in recent years have been 40% among students and 51% among teachers. MPS has agreed to take additional steps to increase these rates. In another research project conducted in MPS, researchers obtained at least 70% response rates from staff in every school (student surveys were not relevant to this other project) with incentives of just \$200 per school. Because student surveys are administered during the school day, student response rates are, according to MPS staff, heavily driven by whether teachers obtain computer lab time for students to fill out the survey. Therefore, to encourage high response rates, we will offer school staff incentives, tying them both to student and staff survey response rates. Schools in which student and teacher response rates exceed 70% will receive \$500 to be used for any purpose. Also, MPS has agreed to send school principals regular updates of their respective survey response rates and to remind them in weekly communications about the \$500 incentive and the importance of the surveys.

National Student Clearinghouse (college academic outcomes). We will obtain nearly complete data on college attendance and completion from the National Student Clearinghouse (NSC), a nonprofit organization that serves as the nation's only source for college enrollment and degree verification. This centralized reporting system collects information from the colleges and universities attended by 92% of U.S. undergraduates (this number likely will be higher in 2015 when GLIMPS students reach college age). Only 6 of the 64 eligible Wisconsin colleges do not participate in the NSC—all very small colleges that are not among those frequently attended by MPS students. Because NSC is a national data system, we will observe enrollments even for students who attend a college that is not GLIMPS eligible. Also, the fact that NSC is a near-census of college enrollment, means that it is generally reasonable to assume that if a student is not shown as enrolled in the NSC then the student is not enrolled anywhere. (We are currently testing the accuracy of NSC in the WSLs study using transcript data and are finding that it has very high accuracy.)

The NSC includes reliable data on college enrollment (including 2- and 4-year colleges), persistence, and graduation. Data are available for each individual college and term/semester a

student attends college. Enrollment intensity (part-time, full-time, etc.) is also included. At present, the intensity variables are considered unreliable, though this is expected to improve by the time the GLIMPS cohort reaches college. Because it is directory information (i.e., does not provide Social Security Number or other sensitive information), student consent is not necessary to use the NSC data for research purposes.

Establishing college enrollment in order to calculate persistence rates will require submitting a file of students' names and dates of birth to NSC, which will then execute a fuzzy matching process to search for enrollment records. MPS regularly makes requests to the NSC and will provide these data in addition to the student record data mentioned above. The PI is experienced with the NSC and has written about its many strengths and some weaknesses (Goldrick-Rab & Harris, 2010).

Proposal Discussion of Empirical Methods

“We organize our discussion of data analysis according to our four research topics. All of the analyses will to some degree rely on baseline equivalence or “balance” between the control and treatment groups. Before proceeding with any analysis, we will report the response rates and baseline (pre-randomization) characteristics of the control and treatment groups and test for differences between them using individual t-tests and multivariate F-tests. We will focus primarily on intent-to-treat (ITT) analyses because anyone who becomes aware of their treatment status is in some sense treated with this type of intervention.

Topic A1: Impact on High School Outcomes

To address our first research question, we will analyze MPS/DPI student record and survey data. While differences in means between control and treatment provide unbiased estimates of impacts, we have the advantage of longitudinal data and can therefore include lagged values of the dependent variables as explanatory variables in most analyses. In general, including lagged values as covariates improves precision (Raudenbush, Martinez, & Spybrook, 2007). For this reason, the minimum detectable effects (MDEs) for these analyses range from 0.14 (standard deviations) for ninth-grade student academic outcomes to 0.17 for the high school exit surveys.

Power is naturally lessened in teacher surveys, with an estimated MDE of 0.28. For high school graduation and college entry, we can identify (dichotomous) effects of 5 and 7 percentage points, respectively. Overall, the study has strong statistical power (see Power Analysis section).

Student surveys. If simple randomization had been used, we would begin with simple differences in means and, to obtain greater statistical power, follow with the usual covariate-adjusted models. As noted in the Sample Selection section, however, we are using paired randomization and this requires adjusting the usual analysis. Hedges (2009) and Imai, King, & Nall (2009) describe different ways to estimate impacts with paired randomization using simple differences-in-means analyses, which can be easily extended to covariate-adjusted models. We will use a Hierarchical Linear Modeling (HLM) framework, estimating a student-level model with school effects as Level 1 and then, with school-level randomization, estimating the treatment effect in Level 2:

$$\text{Level 1: } Y_{ist} = \alpha_0 + \alpha_1 Y_{is,8th} + \alpha_3 X_{it} + \varphi_s + \varepsilon_{ist} \quad (1)$$

where high school outcome measure Y_{ist} of student i in school s at time t is a function of the same lagged dependent variable (“pretest”) from eighth grade, X_{it} is a vector of other covariates (e.g., student demographics), and φ_s represents a vector of school effects. The school effects from Level 1 then enter as the dependent variable in Level 2:

$$\text{Level 2: } \varphi_s = \beta_0 + \beta_1 Y_{s,8th} + \beta_2 X_{st} + \beta_3 T_s + \kappa_p + \varepsilon_{st} \quad (2)$$

where κ_p is a vector of pair indicators with $p=1,2,\dots,18$ (i.e., indicating the pair used in the randomization). The variable T_s represents the treatment status of the schools so that the scalar β_3 is the average treatment effect. (As this is the intent-to-treat (ITT) analysis, students will be assigned to their original school in the coding of T_s .) This is similar to the approach used by Kane and Staiger (2008) who randomized pairs of teachers and then estimated a reduced form version of equations (1) and (2) with many covariates and indicators for each pair of teachers. The above approach will be applied to most of the data sources for high school outcomes discussed earlier, including student surveys, student academic records, and financial aid packages. With slight adjustments for the unit of analysis, this also extends to teacher surveys. Differential nonresponse is the primary threat to validity in the analysis of high school outcomes. In particular, if GLIMPS influences the high school dropout rate—which is one of the key objectives and a focus of our analysis—then this will create differential missing data between control and treatment groups, a potential serious threat to internal validity. We will therefore test whether the difference between exited and retained students (using eighth grade and other prior data) is the same in the control and treatment groups. In the remaining discussion, we assume the worst-case scenario that such differentials will exist, requiring additional steps.

Unfortunately, there is no agreement in the literature on how to handle missing data. We propose to follow the guidelines of an IES technical paper on missing data in the context of cluster randomized trials (Puma, Olsen, Bell, & Price, 2009). Specifically, we propose to (a) clearly specify the rate of missing data on each relevant variable by control and treatment; (b) delete student observations with missing dependent variables (case deletion); (c) apply nonresponse weights if certain hypothesized subgroups have different rates of missing data; (d) add missing data indicators to Equation (1) for all observations with a missing independent variables; and (e) employ multiple imputation when pretests are missing (Little & Rubin, 2002).

We will also take two additional steps: First, as is common in economics, we will estimate lower and upper bounds for impact estimates using the process outlined by Manski (1990). Second, we will seek additional funds from other sources to obtain complete data on a random subsample of students with missing survey data (stratified in this case by treatment condition), following a procedure outlined by DiNardo, McCrary, and Sanbonmatsu (2006). (This method has also been implemented by one of our key consultants, Dr. Larry Orr; see Orr et al., 2003; see also the Personnel section.) These data will provide a valid test of whether the differential response rates introduced bias into the impact estimates, and the direction and size of any bias.

Topic A2: Impact on College Outcomes

The period covered by this research proposal will extend from students' ninth-grade year through their first year in college (see Table 2 in Timeline section), allowing us to add to our high school-focused study with some analysis of initial college entry. The National Student Clearinghouse (NSC) data will include nearly all students in the original sample no matter where they go to college—public or private, in state or out of state. Moreover, any errors in the NSC data are likely to be unrelated to treatment status (see Goldrick-Rab & Harris, 2010) and therefore do not pose a significant threat to internal validity. With randomization and almost no missing data, these estimates will therefore have extremely high internal validity.

Most of the variables in the NSC are dichotomous, and for those variables we will estimate linear probability models, as well as the standard logit and probit. Although we do not have a lagged dependent variable in any of these NSC analyses, paired randomization and covariate-adjustments for other eighth grade measures improve power dramatically.”

Discussion of Mechanisms and Multiple Comparisons

What are the mechanisms and mediators of TDP impacts on high school graduation and college-going? One straightforward aspect of mediator analysis is to estimate equation (1) using the measures in Table 3 pertaining to the mediating variables: affordability, academic preparation, and social capital. These estimates are unbiased (Raudenbush, 2011), but the effects of the mediators on outcomes like college entry are generally biased (e.g., students are not randomly assigned to their academic trajectories). Some have proposed using instrumental variables to solve this problem, but this approach also requires restrictive assumptions (Gennetian, Morris, Bos, & Bloom, 2005).

We believe the best way to go about mediator analysis is to triangulate across different types of data and analysis. First, the patterns of effects on different college outcomes tell us something about the mechanisms. For example, if the promise scholarship has no influence on grades and test scores, then this would tend to suggest that academic preparation is not a mediator—and this interpretation would be reinforced if TDP has no influence on whether students are accepted to four-year colleges. We will also examine the possibility that the mediators are highly correlated, which would make it more difficult to disentangle their roles. Interview data also lend themselves well to mediator analysis and therefore we briefly discuss our approach to these data. As interview transcripts become available, the coding team will develop schemes that reflect both core research questions and emergent constructs. They will first read through a sample of interviews to decide on coding categories and then jointly code a second sample to ensure that the categories are comprehensive and valid. Each interview will be coded by two coders. This cross-sectional coding will allow for preparation of the data set for analysis and ensure that team members have immersed themselves in the sample of interviews and identified sets of themes and questions that need to be explored more thoroughly (Flick, von Kardoff, & Steinke, 2004). Codes will be entered into the Dedoose mixed methods software along with the survey and administrative data. Integration of all data into a single interactive database will facilitate an ongoing and iterative mixed-methods analysis. Interview data will be analyzed by (a) organizing transcripts chronologically by individual students and across cohort years; (b) manually coding respondents' references within the overall topical areas covered in the interview questions; (c) examining trends to discover key categories that emerged from the transcripts; (d) recoding transcripts within these key categories, and (e) compiling data within these coding categories into a tabular format to formulate themes. This

process will be used within and across the student, counselor, and principal interviews. Diamond and Farmer-Hinton will lead the qualitative data analysis with their considerable experience in using qualitative data analysis to understand college access.

We can also use some data from the counselor and principal interviews to carry out quasiexperimental analyses of program implementation. It may turn out that some schools are more aggressive than others in communicating about TDP. In that case, we can use difference-in-difference analyses, creating groups based on the treatment group schools' implementation behaviors. This is of course non-experimental and school-level randomization limits power, but we can complement this with analysis of student interviews that focus on TDP awareness.

We emphasize again that all of these approaches are individually non-experimental and exploratory. But we believe that we can learn something important from the combination of these different analyses. Since we should be able to obtain valid estimates of average treatment effects on college outcomes, it is incumbent on us to try and learn how those effects arise (Harris & Goldrick-Rab, 2012b). Combined with statistical analysis of student surveys, we can triangulate the analysis and perhaps develop plausible theories about mediators.

References

(only references used in the appendix but not the main text are listed)

- Belfield, C. & Bailey, T. (2017). *The Labor Market Returns to Sub-Baccalaureate College: A Review*. New York: Community College Research Center, Teachers College, Columbia University.
- Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2): 263–291.
- Kahneman, D. & Tversky, A. (2000). *Choices, Values and Frames*. Cambridge, MA: Cambridge University Press.
- Kane, T.J. & Rouse, C.E. (1995). Labor-Market Returns to Two- and Four-Year College. *The American Economic Review* 85(3): 600-614.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- Webber, D. (2016). Are college costs worth it? How ability, major, and debt affect the returns to Schooling. *Economics of Education Review* 53: 296-310.