



Optimal College Financial Aid: Theory and Evidence on Free College, Early Commitment, and Merit Aid from an Eight-Year Randomized Trial

Douglas N. Harris
Tulane University

Jonathan Mills
University of Arkansas

We provide theory and evidence about how the design of college financial aid programs affects a variety of high school, college, and life outcomes. The evidence comes from an eight-year randomized trial where 2,587 high school ninth graders received a \$12,000 merit-based grant offer. During high school, the program increased their college expectations and non-merit effort but had no effect on merit-related effort (e.g., GPA). After high school, the program increased graduation from two-year colleges only, apparently because of the free college design/framing in only that sector. But we see no effects on incarceration or teen pregnancy. Overall, the results suggest that free college affects student outcomes in ways similar to what advocates of free college suggest and making aid commitments early, well before college starts, increases some forms of high school effort. But we see no evidence that merit requirements are effective. Both the standard human capital model and behavioral economics are required to explain these results.

VERSION: May 2021

Suggested citation: Harris, Douglas N., and Jonathan Mills. (2021). Optimal College Financial Aid: Theory and Evidence on Free College, Early Commitment, and Merit Aid from an Eight-Year Randomized Trial . (EdWorkingPaper: 21-393). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/wz1m-v526>

Optimal College Financial Aid: Theory and Evidence on Free College, Early Commitment, and Merit Aid from an Eight-Year Randomized Trial

Douglas N. Harris
Tulane University

Jonathan Mills
University of Arkansas

May 1, 2021

Abstract: We provide theory and evidence about how the design of college financial aid programs affects a variety of high school, college, and life outcomes. The evidence comes from an eight-year randomized trial where 2,587 high school ninth graders received a \$12,000 merit-based grant offer. During high school, the program increased their college expectations and non-merit effort but had no effect on merit-related effort (e.g., GPA). After high school, the program increased graduation from two-year colleges only, apparently because of the free college design/framing in only that sector. But we see no effects on incarceration or teen pregnancy. Overall, the results suggest that free college affects student outcomes in ways similar to what advocates of free college suggest and making aid commitments early, well before college starts, increases some forms of high school effort. But we see no evidence that merit requirements are effective. Both the standard human capital model and behavioral economics are required to explain these results.

Acknowledgements: *The Degree Project* (TDP) scholarship experiment was developed in a partnership between the author, Milwaukee Public Schools (MPS), and the Great Lakes Higher Education Corporation (now called Ascendium), which administered the program and committed up to \$31 million for the TDP scholarships. The U.S. Department of Education's Institute of Education Sciences (IES) supported this work through Grant R305A130044. The author wishes to thank Richard George and Amy Kerwin of Great Lakes and Dr. Gregory Thornton, Dr. Heidi Ramirez, Kristin Kappelman, Dr. Deborah Lindsey, Matt Michala, and Marc Sanders of MPS. The University of Wisconsin at Madison, Institute for Research on Poverty provided outstanding assistance with data collection of life outcomes. Andrew Anderson, Paula Arce-Trigatti, Lindsay Bell Weixler, Christa Gibbs, Ryne Marksteiner, Fred Sakon, and Jon Valant contributed to the analysis. Tangela Blakely-Reavis, Bradley Carl, John Diamond, Raquel Farmer-Hinton, Debbie Kim, Hilary Lustick, and Kelly Rifelj are part of the broader TDP evaluation and provided useful comments. For their valuable feedback, We thank Joshua Angrist, Stefano Barbieri, James Benson, Howard Bloom, Matthew Chingos, Emily Cook, Susan Dynarski, Mathew Larsen, Larry Orr, Judith Scott-Clayton, Steven Sheffrin, Douglas Webber and participants in presentations at the University of Michigan, Tulane University, University of Wisconsin at Madison, and the Upjohn Institute, as well as meetings of the Association for Public Policy and Management and the Association for Education Finance and Policy. This study was initiated before pre-registration was common; however, the authors can share the grant proposal and originally proposed analysis plan upon request. The opinions expressed are those of the authors and do not represent views of IES or any other partner. The authors take responsibility for all remaining errors. The study is in memory of Regina Figueirido-Brown, a good friend and collaborator who passed away after years working on this project.

I. Introduction

The rising price of college is increasingly seen as a cause of stagnant and unequal educational attainment levels (Carneiro & Heckman, 2003; Havemen & Wilson, 2007; Goldin & Katz, 2008; Bailey & Dynarski, 2011). College costs, sticker prices, and net prices have been rising 3-4 percent per year for decades.¹ In part to pay these rising prices, real student loan debt more than doubled from 1990 to 2004 (Avery & Turner, 2012) and has increased another 56 percent since then (Institute for College Access and Success, 2020). These trends have raised enough public concern that nearly all candidates for U.S. president in 2016 and 2020 raised concern about college affordability and many proposed policies to address it.

Financial aid is one way to make college more affordable as it reduces the price of human capital investment and increases the internal rate of return to education. The federal Pell grant, targeted to low-income students and available at the vast majority of higher education institutions nationally, is one of the nation's oldest and the largest program by far (e.g., Bettinger, 2004; Marx & Turner, 2018). In the 1990s, states began supplementing federal aid with their own, mostly merit-based, programs (e.g., Dynarski, 2000; Cornwell, Mustard, & Sridhar, 2006; Fitzpatrick & Jones, 2016). "Promise scholarships" represent a newer form of financial aid, in which funding commitments are provided to students during their early K-12 years in particular localities (e.g., Carruthers & Fox, 2016; Swanson et al., 2017; Page et al., 2019; Bartik, Hershbein & Lachowska, 2021). Research on all of these and other programs has

¹ These figures are inflation-adjusted and come from two main sources: College Board (2016) and Delta Cost Project (2010). In some respects, price changes for low-income students have been relatively small. Increases in government financial aid such as the Pell grant program, combined with state and institutional programs, have offset rising costs for low-income students so that the net price (direct costs minus grants and scholarships) has remained relatively steady in recent decades for this group (College Board, 2012). On the other hand, the incomes of less educated parents have been stagnant compared to large increases at the top of income and education distributions, so that relative affordability for low-income families is declining.

generally shown positive effects on college enrollment and attainment (Deming & Dynarski, 2009; Scott-Clayton, 2011; Angrist et al., 2016; Nguyen et al., 2019; Angrist et al., 2020).²

Financial aid effects differ across programs, however, and this may be due to at least five dimensions along which programs vary: aid levels, student eligibility requirements, college eligibility requirements, commitment timing, and design/framing as “free college.”³ In this study, we provide theory and evidence about all five using an unusual, long-term randomized control trial (RCT). *The Degree Project* (TDP) provided a total of \$12,000 in grant funding (a maximum of \$6,000 per year) to students in Milwaukee, Wisconsin, sufficient to make community college free of tuition and fees, and to substantially reduce the price of four-year colleges. In 2011, the program administrators announced the aid offer to each first-time ninth grader⁴ in a randomly selected set of 18 high schools, half of those in the city. To receive the funding, students had to graduate high school on time, reach a cumulative high school GPA of 2.5 (4.0 scale), attend class 90 percent of the time, fill out the Free Application for Federal Student Aid (FAFSA), attend one of 66 eligible colleges in the state, and meet certain additional requirements.⁵

TDP was designed to make an early commitment of financial aid to students in 9th grade. This is potentially important because the largest aid programs do not make concrete commitments until it is too late to affect K-12 schooling behavior and preparation for college

² Exceptions include Angrist et al. (2016), Scott-Clayton & Zafar (2006), which find no aid effects. Another exception is the Wisconsin Scholars program. Goldrick-Rab et al. (2016) found positive effects from one cohort of students who had initially enrolled in four-year colleges; however, there were no effects across subsequent cohorts of four-year college students (Carlson et al., 2020) or on two-year college students (Anderson et al., 2019). A key reason for the absence of effect appears to be that the commitments occurred after students started college and therefore could not influence initial enrollment (Angrist et al., 2020).

³ With U.S. college financial aid system, the distinction between first dollar and last dollar aid is also relevant. This is fundamentally about the aid level and complexity, which are addressed later.

⁴ As in most urban school districts, many students do not fulfill the requirements necessary to be promoted to tenth grade and these students remain as non-first-time ninth graders who were not eligible to participate.

⁵ The merit requirements were patterned after the Pittsburgh Promise program and align well with at least 50 programs nationally, especially large state merit aid programs. Section III discusses additional requirements.

(Dynarski & Scott-Clayton, 2006). Committing aid in 9th grade (or earlier) could increase students' college expectations and do more to prepare before and during high school. The late commitment of traditional aid is also interconnected with student eligibility requirements; since eligibility is generally established at the time students enter college, any commitment made when students are young will be conditional on meeting requirements later, introducing uncertainty and perhaps blunting effects on high school behavior among risk-averse students. We test this hypothesis using a combination of the TDP merit requirements and extensive data collection.

This program was also designed to provide free community college to students, an idea that is gaining in popularity (Goldrick-Rab & Kendall, 2014). Free college programs not only provide more generous aid levels, but also reduce the psychological cost of debt (loan aversion), reduce the complexity of college decisions, and reduce uncertainty about the price of college.⁶ These benefits are mostly rooted in the behavioral economics of financial aid policies (Dynarski & Scott-Clayton, 2006), e.g., people place particular value on things that are free, beyond what we would expect from the price reduction (Shampanier, Mazar, & Ariely, 2007) and people are hindered by complex processes and decisions (French & Oreopolous, 2017). At least 23 states and 21 localities have established free college programs and there is growing interest in making this federal policy.

After tracking students for eight years, we find that the TDP aid offer increased two-year degree completion by 0.5 percentage points (3 percentage points in the treatment-on-treated). While these effects seem small in magnitude, they imply increases of more than 25 percent above baseline outcome levels and are precisely estimated. This graduation effect appears

⁶ Even with free college, students may still need to borrow when they go to college to cover living expenses as students have limited time to work while in college to cover these expenses. However, framing college as free may lead to a perception that all costs are covered and students may respond accordingly, with increased college attendance.

partially driven by a similarly sized, but less precise, effect on two-year college enrollment. We find no effects (or possibly negative effects), however, on four-year college outcomes, indicating substitution from four- to two-year colleges. The overall effect on college attendance is null.

Our analysis of TDP adds to the financial aid literature in six important ways. First, this is one of two studies to estimate the effects of financial aid on college enrollment and attainment in the U.S. using RCTs (see also Angrist et al., 2020), whereas the vast majority of studies use quasi-experimental research designs.⁷ Others have commented on the lack of randomized trials and the need for cleaner identification of effects (e.g., Deming & Dynarski, 2009).

Second, this is the first study to examine financial aid effects on a wide range of high school outcomes.⁸ This is important given the potential of early commitment to leverage human capital investments in the pre-college years. Our results show positive impacts on college expectations and non-merit behaviors that might enhance college outcomes (e.g., participating in other college access programs). However, we find no effect on the merit-related academic outcomes related to the merit requirements (GPA, attendance, high school graduation) or on high school outcome measures that involve costly, sustained effort from students (e.g., test scores).

Third, the early commitment design, combined with our rich annual data, allow us to test for unusual and important forms of effect heterogeneity, especially the effect timing. We find, for example, that the effects during high school are largest for students at the time of the TDP announcement (9th grade) and in the senior year of high school. We see no evidence of effect heterogeneity with regard to race, gender, or family income, however.

⁷ One other RCT has studied the effects of aid given to students after they have attended college (Goldrick-Rab et al., 2016; Carlson et al., 2019; Anderson et al., 2020).

⁸ Carruthers and Fox (2016) is a partial exception and includes high school graduation as an outcome.

Fourth, this is one of only a handful of studies to examine the effects of aid on life outcomes beyond education (see also: Bettinger et al., 2019; Scott Clayton & Zafar, 2019; Carlson et al., 2019)). Using data from eight years after they first learned about TDP, when students were 20-22 years old, we see no evidence of effects on incarceration or teen pregnancy. (We also report imprecise null effects for employment and earnings but it is likely too early for such effects to emerge.)

Fifth, we designed the experiment and data collection to learn how financial aid design and implementation may affect student outcomes. We find that: (a) the positive effects on high school outcomes provides some support for making aid commitments early; (b) the substitution of four-year for two-year college appears to reflect a response to the free college design/framing in the two-year sector; (c) the merit requirements attenuated effects on college and life outcomes; (d) the college eligibility requirements likely reduced the quality of colleges attended; and (e) the communications received by students varied widely and may have influenced student responses to the program, indicating that program implementation matters as much as design.

To provide additional evidence on aid design, we also compare TDP results with studies of other programs in a cost-benefit framework (Harris, 2013; Bartik et al., 2016; and Angrist et al., 2020). Addressing two limitations of prior cost-benefit studies, this analysis suggests that programs with both early and free college commitments, and limited merit requirements, do more to increase social welfare than other forms of financial aid.

Sixth, and finally, we outline and test theories about student responses to financial aid. On the one hand, the results are consistent with the standard economic theory in yielding positive average treatment effects on some college outcomes and effects on low-cost high school effects in 12th grade. On the other hand, the shift from four- to two-year colleges cannot be explained by

the standard theory, but can be explained by behavioral economics theories about loan aversion and decision complexity. In short, these results reinforce that a combination of standard human capital theory and behavioral economics is necessary to understand student responses to financial aid (Dynarski et al., 2021) and the behavior of young people more generally (Harbaugh, Krause, & Vesterlund, 2002).

Section II provides the theories and hypotheses for the different types of financial aid effects that we can test in a study with many different outcome measures. Section III describes *The Degree Project*, including the randomization process. The data for our high school, college, and life outcome analyses are discussed in section IV. The results, including intent-to-treat, treatment-on-treated, and effect heterogeneity are provided in section V. Section VI provides additional evidence about the costs and benefits of TDP and a variety of other financial aid programs. Section VII concludes.

II. Theory and Hypotheses

This section outlines theories about how students' college decisions are affected by college financial aid, starting with a standard theory regarding rational actors making decisions about college-going and high school effort to maximize utility subject to imperfect information, uncertainty, and risk-aversion. After each portion of the discussion, we identify hypotheses that are suggested by the theory, which we later test. (These are not theorems, which would require imposing many more assumptions and providing proofs that we not see as useful in this context.) At the end of the section, we show whether and how the hypotheses require modification when we add in behavioral economic theories.

II.A. Standard Economic Theory

We assume that students (indexed by i) are rational and maximize utility. Specifically, they choose whether to obtain a 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.⁹ 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:

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

Given the above assumptions, when students choose not to attend college, 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 (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

⁹ In reality, financial aid g is a function of p in the U.S. We add this later in the section.

degree choices. For example, it may be that high school effort increases the chances of gaining admittance 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 $d_i(m_i, e_i)$.¹⁰ 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:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - p + g) - c(e_i), \quad m_i \in \{0,1\}. \quad (2)$$

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 (Hoxby & Turner, 2015; Dynarski et al., 2021). One implication is that, if some students underestimate g , then communicating 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, intuitively, it might seem that we still arrive at *Hypothesis 2c* even if students do not underestimate g . This is especially

¹⁰ 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.

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).¹¹

But whether risk-aversion increases the benefit of early aid commitments depends on the return to effort.¹² 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).¹³ For this reason, uncertainty and risk-aversion alone are insufficient to obtain *Hypothesis 2c*.¹⁴

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

¹¹ Risk-aversion also increases with age (e.g., Harbaugh, Krause, & Vesterlund, 2002), so teenagers may be less risk-averse than adults.

¹² Formally, we can introduce risk aversion by defining the right-hand side of (2) as D and placing this within the function $V(D)$, which is increasing in D and strictly concave.

¹³ 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 analysis and conclusions in a tournament setting.

¹⁴ 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.”

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:

$$V(m_i, e_i) = d_i(m_i, e_i) \cdot (w_i - p + (1[\alpha_i + h(e_i) \geq A] \cdot g) - c(e_i)) \quad (3)$$

Students again choose (m_i, e_i) and, if they meet the merit requirements, then they 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 target 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 whether students obtain degrees is increasing in their academic endowment even apart from the merit requirements.¹⁵ This discussion of the incentive and targeting 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

¹⁵ We fix effort at zero to separate this from the incentive effect above.

proportionally larger. The ITT effect is the effect on 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, which would have the opposite effect and increase the number of degrees.

Hypothesis 3b: Merit requirements increase the treatment-on-treated (TOT) effect on college outcomes if the incentive and/or targeting effects are 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, which, by definition, is conditioned on meeting the merit requirement. So, any incentive or targeting effect is sufficient to increase the TOT.

Merit Aid Effects on High School Effort. Next, we show that the optimal effort level e_i^* is a non-linear function of 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 $a_i = A$; or, if α_i is sufficiently low, the disutility of effort exceeds the utility from w_i , in which case $e_i^* = 0$.

Figure 1A illustrates this with the high school effort response function. 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.¹⁶ The solid black line is the effort response function when students are certain about all the parameters.¹⁷

Now, suppose that students are risk-averse and that α_i for each student comes from a random draw from a well-defined distribution (all other parameters remain known).¹⁸ (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 $E[\alpha_i] = \alpha_i$; 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.

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.¹⁹

A further implication of Figure 1A is that when certainty about α_i is high, we can expect

¹⁶ 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 .

¹⁷ 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.

¹⁸ 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.

¹⁹ 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.

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, perhaps surprisingly, there a sharp change in effort at the threshold.

In using the above analysis to develop a hypothesis for our evaluation of TDP, two factors seem important. On the one hand, TDP creates 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 1A.

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.

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 productivity), e_i^A increases indirect utility in two ways, while e_i^o contributes in only one way, so that $\partial V/\partial e_i^A > \partial V/\partial e_i^o$.²⁰

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

²⁰ Another assumption required for this hypothesis is that e_i^A and e_i^o create costs through the same cost function.

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 assume above, so that $e_i = \sum_r e_{ir}$ and $c(e_i) = \sum_r (1 - \delta)^t c(e_{ir})$ where r reflects each grade. 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 functions under these two scenarios are shown in Figure 1B.

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

Since Hypothesis 6 is driven by the discount rate, note that, like risk-aversion, the discount rate decline 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, of the sort that are the subject of this study. For this reason, Figure 1B displays two lines, one for low- and high-discount-rate students.

II.B. The Behavioral Economics of Financial Aid

The above model focuses on how rational actors make college decisions under situations of imperfect information, uncertainty, risk-aversion, discounting, and effort cost. However, some have argued that students' college-going behavior is more consistent with behavioral economics (Dynarski & Scott-Clayton, 2006; French & Oreopolous, 2017). Below, we explain how behavioral economics might either reinforce the above hypotheses, alter them, or lead to new hypotheses because of loan aversion, salience, status quo bias, anchoring, and decision complexity. We provide a brief introduction to each theory and then explain their implications

for the hypotheses.

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 on 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).²¹

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 first re-state these 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 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.” This default can be formalized as additional indirect utility that arises when students’ decisions coincide with their expectation.

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,

²¹ 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 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.

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 grade. Salience reinforces the idea that effort will increase in 12th grade because this is when the (initial) decision to attend college is generally made. In addition, however, salience means that students are also likely to be respond with additional effort when they first learn about the grant. In the case of TDP, this was in the fall of 9th 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 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 direct effect of increasing g . (We would also frame ζ as a cost of decision complexity, which is also suddenly eliminated when $p = g$.²²)

²² 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.

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 decision. 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 illustrates a key point: that free college does little to reduce decision complexity or loan aversion in the presence of 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.²³

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, our interpretation of the results, and the analysis of optimal financial aid policy that follow.

III. Background: *The Degree Project*

The lead author designed *The Degree Project* program in partnership with the program funder and operator, the Great Lakes Higher Education Corporation and Affiliates (now known as Ascendium), and the Milwaukee Public Schools (MPS). Great Lakes committed \$31 million to fund the grants, enough to provide the full grant to every one of the 2,587 TDP promise

²³ In addition, the change in the default from the early commitment noted earlier could shift effort upwards, which, in turn would affect the probability that $1[\alpha_i + h(e_i) \geq A]$, so that students receive the funds.

recipients. TDP is a demonstration program designed to identify impacts therefore only one cohort of students was directly involved.

All schools serving ninth graders in MPS were included in the pool so long as schools provided GPA and attendance data to MPS and served students from ninth through twelfth grade. Within the 36 TDP schools, all first-time ninth graders enrolled as of November 1, 2011 were identified as eligible using administrative data.²⁴

We used paired cluster randomization to select treatment schools (Imai, King, & Nall, 2009). Pairing on a lagged value of the dependent variable maximizes precision because, compared with other pairing methods, this minimizes the variance of the baseline differences between the control and treatment groups on expected future outcomes. Specifically, we carried out the following steps: (a) ranked schools by the college attendance rates from recent cohorts; (b) created pairs of schools based on this prior college attendance rate (i.e., the two schools with the highest prior college enrollment rates form the first pair and so on); and (c) randomized one school in each pair to the treatment and the other to control. The selection of 18 treatment schools was carried out solely by the author, using data from MPS. (See Appendix B for additional detail on the data used in the paired randomization and related steps.)

On November 17, 2011, all of the 2,587 students selected for treatment via the above process were sent letters announcing their selection, describing the program, and directing them to the program website for additional information. There were 2,464 eligible students in the non-selected schools, which serve as the control group.

²⁴ Only a handful of people in the district knew about the program prior to the November 1 enrollment eligibility cut-off date.

TDP treatment students received TDP funds so long as they graduated from any MPS high school²⁵ on time (within four years of starting ninth grade) with at least a 2.5 cumulative GPA (C+/B-) and attend school 90 percent of the time.²⁶ Similar to those of the Pittsburgh Promise (Page et al., 2018), the cumulative nature of the requirements was intended to allow initially-low-achieving students time to catch up. Students remained eligible for the grant regardless of whether they switched high schools therefore this approach does not create a direct incentive for any MPS student to switch high schools. However, they still had to graduate from an eligible MPS school to receive funds and this created some incentive to stay within MPS.

In the graduating class of 2008-09, who were first-time ninth graders in 2005-06, 16.3 percent met these requirements (in the absence of the treatment) and 65 percent of those students went directly on to college in fall 2009.²⁷ The average ninth grade GPA for the TDP cohort was 2.2 with 81 percent attendance, both well below the program's eligibility thresholds.

Several additional program rules became relevant toward the end of high school and into the college years. Students had to complete a Free Application for Federal Student Aid (FAFSA) senior year and each year of college, which could be an impediment to college entry (Bettinger et al., 2009). Students need not start college immediately after high school, but had to start within 15 months of on-time high school graduation (fall of 2016).

The funds could be used at any of 66 two- or four-year public college and many private colleges in Wisconsin. Students could spend up to half the total grant per year if they attend full-time (≥ 12 credits) and half this amount if they attended at least half-time. There were no GPA

²⁵ There were a few exceptions, as a few very small schools do not report sufficient data to MPS to check eligibility requirements. Students were regularly updated on which schools they can graduate from in order to remain eligible.

²⁶ In practice, these were operationalized by the program operator as 2.45 GPA and above and 89.5 percent attendance to address rounding. We follow this same rule in our analysis.

²⁷ Author's calculations with assistance from MPS.

requirements during college, but students had to use the grant funds within four years of expected high school graduation (i.e., by spring of 2019). If students obtained a two-year degree, they could still use the remaining funds for other degrees. Further, students could use the funds if they transferred institutions, so long as the receiving institution was eligible.

To receive the funds, students also had to be first-time college enrollees, degree-seeking and have at least \$1 of unmet need.²⁸ Therefore, while income did not directly affect initial eligibility, it did affect the level and form in which students received the funds. Financial aid offices disbursed the aid following the same process used to disburse state grant aid. TDP grant are “last dollar”²⁹ and covered up to the cost of attendance.³⁰

To place the grant amount in perspective, half of MPS high school graduates who went on to college typically attended either Milwaukee Area Technical College or University of Wisconsin–Milwaukee, where full-time annual tuition and fees in 2012 were \$3,184 and \$8,675, respectively. Given these figures, the communications to students indicated that this was enough to cover all tuition and fees for a two-year degree (i.e., free college), an especially positive framing for students whose families appear to be loan averse.³¹

Great Lakes, as the program administrator, sent two letters in the first year, and four letters per year in the other three years, not only to remind students about their eligibility but

²⁸ Unmet need is the cost of attendance minus the expected family contribution and existing grant and scholarship aid (excluding loans and work study).

²⁹ Last dollar means that students draw on all other forms of public funding first before using the TDP funds; that is, the TDP funds are the last dollars considered. A key implication is that some students will receive fewer dollars than others even if they attend the same college at the same time with the same credits. Also, last dollar funds do not have the potential to reduce living expenses and other costs related to college in the way that first dollar funds do.

³⁰ Because the scholarship is last dollar, funds sometimes substituted out loans and other forms of aid depending on income levels and cost of attendance, but essentially all students will see substantial price reductions. In another study of a Wisconsin financial program administered in similar ways and for a similar population of students, most substitution was in the form of loan reduction (Goldrick-Rab, Harris, Kelchen, & Benson, 2012).

³¹ In our baseline survey, half of the students’ parents reported that it was very or extremely “wrong . . . to owe money” and almost three-quarters reported that it was very or extremely “hard . . . to get our of debt.”

about whether they were meeting the requirements at each point in time (see a sample communication in the appendix). Great Lakes also encouraged parents, counselors, and teachers to talk to eligible students about the program. Because students are imperfectly informed about the cost of college (Ikenberry & Hartle 1998; Avery & Kane, 2004; Dynarski et al., 2021), some of the communications to students include information about full-time tuition costs at institutions typically attended by MPS graduates.

IV. Empirical Framework

IV.A. Sample Description and Data

The Milwaukee Public Schools (MPS) is a large urban school district that enrolls an average of roughly 80,000 students, making it the 36th largest district in the nation. The student body is majority African American and one-quarter Hispanic. Test scores are well below the national average and below the urban district average (U.S. Department of Education, 2011). Parent income is slightly below the urban district average (UW-Milwaukee, 2010) and, as a result, four out of five students are eligible for free or reduced-price lunches. (See Appendix A for additional background on Milwaukee.)

We study the effects of TDP in Milwaukee on a wide variety of outcomes. While it is common to study effects on college outcomes, our data allow us to study effects on high school outcomes and life outcomes as well. The variables, sources, years, and missingness are discussed below and summarized in Table 1.

IV.A.1. High School Academic Outcomes

The first three outcomes of interest are those directly tied to the program requirements: GPA, attendance, and high school graduation. We include both on-time graduation with a regular

diploma (as required for the scholarship funds) and any graduation within five years of starting high school (including late graduation and complete with a GED or other credential).

Students must pass the state standardized test (or another test) in order to graduate from high school, but exceptions are routine, and the tests are not part of the TDP requirements. Also, the state standardized test is administered only in 10th grade. So, we instead examine test scores on the low-stakes Measures of Academic Progress (MAP) test, which is aligned to the state standardized test, and administered by MPS three times per year in all high school grades.

IV.A.2. High School Outcomes from Student Surveys

The district administers an annual climate survey to all MPS students each spring and allowed the research team to add ten questions each year. Compared with the overall sample, the survey sample had higher pre-treatment levels of attendance, GPA, and test scores (Table 1).

MPS also administers a Senior Exit Survey MPS administers each year to measure students' college plans and steps they are taking to prepare, just before leaving high school. Like the climate survey, MPS also allowed us to add some questions for the evaluation of TDP. The survey for the TDP cohort was administered during January-June of 2015 (in the senior year). The four main categories of college-related measures are: general college plans, college applications, other steps taken to prepare for college, and measures related to college finances.³²

IV.A.3. College Enrollment Data

We measure college entry using the National Student Clearinghouse (NSC), a near census of students in college (Dynarski et al., 2013). The NSC data were provided through MPS for the entire TDP cohort (control and treatment, unconditional on high school graduation)

³² If students did not plan to attend college, they were not asked to fill out the remaining college-related questions. Therefore the analyses of these questions are censored. This is in some sense necessary because students not planning to attend college could not possibly answer questions about two- and four-year college or specific colleges (or selectivity), for example.

through four years of post-on-time high school college enrollment. We use these variables to identify any college enrollment and distinguish two- from four-year colleges, TDP-eligible from ineligible colleges, and in-state from out-state colleges. By combining the Senior Exit Survey with the NSC, we can also obtain direct measures of “summer melt,” i.e., the fact that students who say they plan to go to college in their senior year end up not attending perhaps because of something that occurs during the intervening summer (Castleman & Page, 2015).

Table 2 provides the control group baseline means for the TDP sample, which mirror the academic and income reported above for the district as a whole. In addition, note that, among the prior cohort of first-time MPS ninth graders in fall 2002, 67.8 percent completed high school on time in 2006, and 44.4 percent of those high school graduates directly transition to college, somewhat less than the national average.³³

Twenty of the 36 TDP schools are some form of traditional public schools and the others are charter schools, which operate semi-autonomously from MPS and have smaller enrollments (see Appendix Table A2). Four of the 36 TDP sample schools can be considered selective admissions in that students have to apply and meet academic requirements. Partly as a result, the college-going rates varied greatly across high schools prior to TDP, ranging from 10 to 88 percent for on-time college enrollment.³⁴

During the TDP cohort’s sophomore through senior years (post-treatment), MPS closed or re-organized eight of the 36 TDP schools (3 treatment and 5 control), which apparently reduced outcomes for these students (Larsen, 2016). Great Lakes sent letters to treatment

³³ 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). Subsequently, 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.

³⁴ The choice of variables in Table 2 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.

students and their parents to convey their continued eligibility for the program, as they did with students who transfer to another MPS school.³⁵ We account for school closure/re-organization as a covariate in some of the analyses that follow.

IV.A.4. Life Outcomes: Labor Market, Incarceration, and Pregnancy

We measure life outcomes using data from the University of Wisconsin at Madison, Institute for Research on Poverty's (UW-IRP) Multi-Sample Person File (MSPF) Data System (Brown et al., 2020). The MSPF integrates Wisconsin state agency data on public assistance, child support, child welfare, unemployment benefits, and incarceration administrative data systems to create a unique record for each participant across years. We code students as employed if they have any earnings information from summer of 2015 (just after on-time high school graduation) through December 2017 (the end of the panel) and code them as not employed otherwise. We also created a separate variable for whether they were employed *or* enrolled in college since there may be substitution between the two and both are positive outcomes. Earnings are calculated by summing quarterly data across the most two years available (January, 2015 to December, 2017).³⁶ We emphasize that effects on labor outcomes are unlikely to emerge at this point except perhaps for students who did not attend college.

Individuals are identified as “ever incarcerated” if they are registered as incarcerated at any point in the panel.³⁷ Finally, “ever had a child during high school” identifies students who reported having a child in Wisconsin's child-care services data during high school (between

³⁵ A ninth school ended its relationship with MPS and stopped sharing data with the district. As a result, the students who graduated from this school were no longer eligible to receive this scholarship (eligibility criteria could no longer be checked), though students could transfer to other MPS high schools and continue their eligibility.

³⁶ Individuals who do not appear in the MSPF data or for whom quarterly earnings are missing are given an earning value of 0. We account for the large number of observations with 0 earnings in our analysis using an inverse hyperbolic sine (IHS) transformation.

³⁷ Note that there is considerable evidence of bias against black people in the criminal justice system that may be reflected in these numbers (e.g., Rehavi & Starr, 2014).

January 2012 and July 2015). These life outcomes are more useful than the labor outcomes because we can capture pregnancy during the teen years and incarceration rates generally peak between the ages of 16-24, which are included in the analysis. With all of these life outcome variables, missing data are minimal.³⁸

IV.A.5. Missing Data and Attrition

In a randomized trial, the primary threats to identification is missing data/attrition. NSC data were requested for all the originally selected students so there is no missing data, except for the very small (and likely random) error rates in the NSC itself (Dynarski et al., 2013). For the high school administrative data (GPA, attendance, etc.), data are available for all but 11 of the 5,052 students.³⁹ However, attrition worsens over time in the administrative data as students drop out or transfer to schools outside the MPS system. We use cumulative GPA and attendance using the last observed data to minimize the effect of attrition. In contrast to the administrative data, which is nearly complete, the overall survey response rate was 53 percent averaged across the four years, with 3-5 percentage point higher rates for the treatment group.

IV.B. Baseline Equivalence

Table 2 reports estimates of equation (1) with baseline measures as dependent variables (and only pair effects as covariates) to test for baseline equivalence between the control and treatment. The baseline differences usually favor the treatment group, though only one of seven baseline equivalence tests for academic outcomes rejects the null (math scores). The same

³⁸ Like the NSC, it is not possible to establish a match rate in the MSPF because people only show up in the state data when the specified event (e.g., employment or incarceration occurs). However, both the state data and MPS data included extensive identifying information, including Social Security Numbers and home addresses, so we expect the match rate is very high and balanced by control and treatment groups.

³⁹ We also regressed a treatment status indicator on a vector of indicators for missing variables. The joint test rejects the null that missingness is the same in control and treatment.

general pattern emerges, albeit more strongly, in the survey sample. On the other hand, the treatment students are more likely to be eligible for free or reduced-price lunches.

Figure 2 provides a visual representation of baseline equivalence, comparing schools within each pair on the pairing variable itself (college entry rate of prior cohorts). The fact that prior college going rate is the pairing variable is visually evident from the close overlap within each pair on that variable. Since college outcomes are arguably the outcomes of greatest interest, the fact that we have baseline equivalence on this measure is important.

Figure 3 shows kernel density plots for attendance and GPA, also by control and treatment.⁴⁰ There are no baseline differences in the individual survey measures (available upon request).⁴¹ As shown later, and in equation (1), covariates are sometimes included in the impact estimation to address these relatively small imbalances.

IV.C. Estimation

TDP impacts are identified from a paired cluster random control trial. Our preferred estimator therefore is simply:

$$Y_{it} = \beta_0 + \beta_1 T_i + \kappa_{ip} + \varepsilon_{ist} \quad (6)$$

where Y_{it} is outcome measure for individual i at time t (up to eight years post-treatment) is a function of treatment offer status (T_i). Other than T_i , the only term included in all our specifications is κ_{ip} , a vector of pair indicators with $p=1, 2, \dots, 18$, which are necessary to account for the randomization design and obtain correct standard errors. We are primarily

⁴⁰ Since we are primarily interested in whether students are different overall, we also regressed the treatment indicator on the vector of demographic and/or lagged dependent variables (see bottom of Table 3). The joint test rejects the null and differences, again, tend to favor the treatment group. Pair indicators are included in the estimation, but are not part of the null hypothesis in the joint test.

⁴¹ For the baseline test with survey measures, we use survey data from the prior cohort of ninth graders in the same 36 schools (i.e., those students in ninth grade in fall, 2010) and only for the two composite measures that were created based on survey questions that were asked the prior year. We cannot distinguish first-time from non-first-time ninth graders, which somewhat limits the validity of this test.

interested in average treatment effect parameter β_I .

Given that there were some baseline differences between the control and treatment grounds, we include in some specifications non-linear functions of the baseline *individual*-level dependent variables $f(Y_{i0})$, non-linear functions of the baseline *school*-level dependent variables $f(Y_{s0})$, and other individual- and school-level variables (X_{ist}), including student demographics.⁴² The X_{ist} vector also includes an indicator for whether an announcement was made to close or re-organize the school after treatment began.⁴³

While covariate adjustment can introduce bias when there is effect heterogeneity (Freedman, 2008), we see little evidence of effect heterogeneity with respect to the covariates we include in our analysis, so bias is unlikely. Rather, if the coefficients change with the addition of covariates, it is likely because of the slight baseline imbalance. Additional covariates can also increase statistical power, as they do in this case. Most reported effects are from Ordinary Least Squares (OLS). We also estimate logits for binary dependent variables (some survey measures and thresholds effects). Major deviations between the OLS and other estimation techniques are noted in discussion of results, though these are rare.

IV.D. Standard Errors and Statistical Power

We estimate models at the student level and generally report cluster-robust standard errors based on the usual Generalized Estimating Equation (GEE) method (Liang and Zeger,

⁴² We do not have pre-treatment values of all the dependent variables at the individual level. This is especially true with the survey measures, in which case we used lagged values of individual-level attendance and test scores as covariates in each of the models. We include both eighth grade and ninth grade lags in Y_{it} because both are imperfect: the latter measures include only two months and are therefore noisy while the former measures are missing at higher rates and subject to other forms of measurement error. Nearly all the students in the sample attended different schools in eighth grade than they did in ninth grade. Many students attended non-MPS schools, so the missing data rates are higher.

⁴³ We recognize the problem of including post-treatment covariates, but, in this case, school closure is almost certainly exogenous and we use the covariate of school closure to test whether this influenced student outcomes in ways that happened to be correlated with treatment status.

1986) with clustering at the level of the original school students attended just prior to the program announcement. This requires asymptotic assumptions that do not hold in small samples, though the number of clusters in this case (36) is similar to the minimum number recommended by some researchers (Cameron, Gelbach, & Miller, 2008).⁴⁴ In any event, our results are generally robust to alternative inference (MacKinnon & White, 1985; Cameron, Gelbach, & Miller, 2008), adapted to this paired randomization design.⁴⁵ They are also robust to estimation at the school-level of aggregation.

Statistical power is driven more by the number of clusters in this type of RCT than the number of observations within each cluster. For this reason, the usual rule of thumb is that 40 clusters are required to achieve sufficient power with cluster RCTs. However, pairing on the lagged dependent variable largely offsets the loss of power from cluster randomization. Therefore, we are able to identify effects on overall college entry, for example, as small as 0.6 percentage points.

IV.E. Treatment Dosage

We consider two types of treatment effects: (a) the aid offer itself could influence high school effort, so the offer is the treatment (ITT); and (b) the receipt of funding may influence college and subsequent outcomes (TOT). We discuss below the dosage (or treatment contrast) associated with each.

⁴⁴ Cameron, Gelbach, and Miller (2008) show that the approach yields valid inference beyond 30 clusters, at least under somewhat simple forms of heteroskedasticity. Kezdi (2004) finds that the standard errors are accurate when there are more than 50 clusters, a standard espoused by Angrist and Pishke (2009).

⁴⁵ Bootstrapping is an alternative. In a simple cluster randomized trial, the block is the cluster, so the bootstrap re-sampling is at the cluster level even when microdata are available. In a paired cluster RCT such as TDP, the block is the pair and we refer to this as the “stratified sample bootstrap” (SSB). As a robustness check, we calculated two additional standard errors for some analyses: the wild bootstrap-se (Cameron, Gelbach, & Miller, 2008) and the bias-reduced linearization proposed by MacKinnon and White (1985) and Bell and McCaffrey (2002), and recently advanced by Pustejovsky and Tipton (2016). The standard errors are 25-50 percent larger in the case of the BRL method though this rarely changes the conclusions (exceptions are noted later). Adjusting for multiple comparisons increases the standard errors further.

With regard to the ITT, the dosage is arguably the level of awareness of the offer. We measure this dosage using student-level spring survey data. Two-thirds of the treatment student respondents correctly reported their treatment status and, of those, students correctly identified the eligibility criteria 80 percent of the time (compared with a 50 percent expected by chance). Only about five percent of the control group responded that they were offered the scholarship. We carried out the same test each year and obtained similar results, suggesting that students did not forget about the scholarship, perhaps because of the regular communications.

The treatment contrast in aid receipt, for the TOT, might be limited because the merit requirements restricted the share of treatment students who received funds (see theory in section II). The aid formula also affected how much funding students actually received. We estimate that the average TDP grant at four-year colleges was \$4,262 and average reduction in loans was \$1,407 annually.⁴⁶

The percentage of students self-reporting that they planned to use “scholarship funds” for college was only eight percentage points higher for the treatment group.⁴⁷ While the expected level of scholarship funds was no doubt higher for the TDP treatment students, compared with the control group, this small contrast likely dampened the treatment effects. Formally, we consider students treated if they met the merit requirements by the summer after on-time high

⁴⁶ These numbers are based on anonymized group-level data from the University of Wisconsin at Milwaukee only. The loan reduction is expected given that the grants reduce the need for loans (Angrist et al., 2020). A similar ratio of grant-to-loan reduction was found in another financial aid experiment (Goldrick-Rab, Harris, Kelchen, & Benson, 2016). Estimating the TDP effects on aid packages of students attending two-year colleges is more difficult. Based on financial aid rules, it is almost certain that most students receiving TDP funds and attending two-year colleges paid no tuition and fees over the first several years. Some likely also received checks covering part of their living expenses (up to the cost of attendance).

⁴⁷ Student perceptions also appeared to be in line with what eventually happened as only 21 percent eventually met the requirements. The fact that this is somewhat higher than the survey is probably explained by the fact that the survey was carried out in the spring, before students received congratulatory letters. We cannot be sure of the size of the scholarships the control group was eligible for.

school graduation and were sent a congratulatory letter by the program operator, so that they had the opportunity to use the TDP funding.

V. Results

Below, we present results for the various high school, college, and life outcomes using four econometric models. We begin with just the treatment and vector of pair indicators (equation (5); Model 1), then add various combinations of individual- and/or school-level pre-treatment dependent variables (Models 2 and 3). Finally, we add a vector of student demographics and a school closure indicator (Model 4). Given both the school closures that occurred in the middle of the experiment and the baseline nonequivalence (Table 2), we give the greatest weight to estimates that are robust across the four columns.

V.A. Effects on High School Performance Metrics

Table 3A reports average effects for the first-year academic measures that are included in the merit requirements. Only five of the 44 coefficients are significant in Table 3A. With regard to *Hypotheses 2a-2c* and 5, we see little evidence that students responded to the performance incentives by increasing merit-related effort. First, we see no effects on any of the three academic merit requirements: GPA, attendance, or high school graduation. For GPA and attendance, Table 3A also tests for threshold effects, where we are most likely to see incentive effects, but there is no evidence of effort effects there either. While this might seem to contradict Scott-Clayton (2011) who did find incentive effects in a study of the West Virginia (WV) Promise, but that study focused on performance requirements for *maintaining the aid* during college. Performance requirements provide more powerful incentives because they leverage loss

aversion in ways that merit requirements do not.⁴⁸ Also, Scott-Clayton (2011) found that the largest effects were on credit hours, which requires effort but almost no knowledge of the production function (Fryer, 2010).⁴⁹

We initially see some evidence that the scholarship increased math scores; however, this is likely a result of baseline non-equivalence noted earlier. The effects on math shown here are similar in magnitude to the baseline differences; and they become small and significant when we control for these differences in columns (2)-(4). There are some signs of increases in high school degrees, but these are not robust to specification.⁵⁰

V.B. Effects on Non-Merit High School Effort and College Expectations

More consistent with *Hypothesis 2a*, the effects of TDP are more positive when we turn to the survey measures of non-merit metrics. Table 3B shows that treatment assignment increased the number of college access programs students reported participating in and the number of colleges students applied to. TDP also increased college expectations, overall and especially for attending college full-time. This is noteworthy given the behavioral economic theory of reference points and defaults. Our evidence of increased college expectations seemed

⁴⁸ The WV Promise was also more generous than TDP, covering all tuition and fees at both two- and four-year colleges. The four-year college students therefore had much more to lose than those in TDP.

⁴⁹ To be initially eligible for the WV Promise, students had to score at least a 21 overall on the ACT or 1000 on the SAT. To maintain eligibility, students had to complete at least 30 credits per year and maintain a 3.0 cumulative GPA. Students meeting these merit requirements likely had more knowledge of the production function, making it even easier to meet the merit requirements necessary to maintain eligibility. Scott-Clayton (2011) focused on the college performance requirements not the high school merit requirements.

⁵⁰ The largest and most robust effect on high school academic outcomes in Table 3A is an increase in transfer to other MPS high schools. Students may have tried to move to high schools where receiving a 2.5 GPA was easier to achieve or moved to more challenging schools that might better prepare them for college. Additional analysis, however, shows no difference in the (pre-treatment) school-level college entry rates between the sending and receiving schools of mobile students. The null effects on academic outcomes remain even after controlling for student transfer. Since transfer is endogenous, the models that control for this analysis are not our preferred estimates and are not shown.

to change students' default options, which may partially explain why we also see the above increases in students' college-related effort.

Mixed effects emerge on students' finance-related steps and perceptions. Students were more likely to fill out the FAFSA (a requirement to receive scholarship funds), but the program had no effect on their concerns about the cost of college. This may be because, while the program reduced costs for many students, the frequent discussion of college cost in the communication from program administrators may have also raised more attention to costs and created concerns that students did not originally have.

The above, generally positive, effects on student behaviors and perceptions are largely confirmed by additional descriptive analysis; 56 percent of treatment students reported that TDP led them to work harder (39 percent reported no change and five percent reported a drop in effort). It therefore appears, taking Tables 3A and 3B, together with the descriptive analysis, that students made some response to the merit requirements, but only low-cost, non-sustained effort, i.e., FAFSA completion (consistent with *Hypothesis 2a*). We also seem some effort response with non-merit metrics, but all of these are low-cost as well (e.g., participating in other college access programs). This suggests that the cost of effort is a significant factor makes students reluctant to put forth effort in the more sustained ways required to increase GPA and attendance and/or that students are so uninformed about the education production function that their efforts are just not reflected in the data.

V.C. Effects on College Entry and College Type

Table 3C shows no effect on overall college entry. The point estimates for two-year college attendance are only as high as one percentage point, but these are not statistically significant in any of the models and any positive effect that may have occurred was apparently

offset by a similar-sized reduction in four-year college attendance, yielding no effect on overall attendance. This substitution across sectors may seem surprising given that TDP provided identical nominal funds for either two- or four-year college. In fact, students attending two-year colleges likely could not use all the funds, due to the other aid they were eligible for and the capping of aid at the cost of attendance. In this sense, TDP reduced the net price more for four-year colleges and yet the impact of the free two-year college framing was still strong enough to get students to shift to two-year colleges.

The above findings support the relevant hypotheses. While the lack of an effect on entry/matriculation seems inconsistent with *Hypothesis 1*, recall that *Hypothesis 3a* suggested the effect should be attenuated because of the merit requirements, as only 21 percent of those offered TDP ultimately met the requirements. Moreover, the shift from four- to two-year colleges is broadly consistent with our hypothesis regarding free college (*Hypothesis 7*).

Prior research suggests that students are quite willing to substitute the specific colleges they attend in order to reduce the price and that this can also entail reductions in college quality (Cohodes & Goodman, 2014). While the estimates are imprecise, TDP seems to have induced students to shift from out-state/ineligible to in-state/eligible colleges and to colleges with lower graduation rates.⁵¹ We also find evidence of increased undermatching, as in Hoxby and Avery (2012).⁵²

Finally, we see no effect on “summer melt.” In fact, the estimates consistently point toward a drop in college enrollment conditional on expectations. This may reflect that summer

⁵¹ We used the graduation rate for the two- and four-year colleges that students attend and the competitiveness of four-year colleges only. Whether students attend college, and in which sector, is endogenous in this analysis of college type, but these results are still suggestive of effects on the types of colleges attended.

⁵² One reason that states might nevertheless include in-state requirements, as well as merit requirements, is to keep their best students in state, i.e., to prevent “brain drain” (Cornwell, Lee, & Mustard, 2005; Fitzpatrick & Jones, 2013; Zhang & Ness, 2010).

melt is based on a comparison of expectations and behavior and TDP increased expectations (see above Table 2B).

V.D. Effects on College Completion and Life Outcomes

Table 3D focuses on college completion and life outcomes, measured up to four years after on-time high school graduation. The top rows show uniformly positive point estimates for students graduating from college within this time period and these are significant in the two-year college sector. The program apparently increased college completion by a half percentage point, compared with a baseline mean of two percentage points, a 25 percent increase.⁵³ The estimates are concentrated in (and precisely estimated for) for two-year colleges. Note, too, that the point estimates for college graduation are similar to those for on-time college attendance (Table 3C) and the main difference here is that the estimates are now more precise.

Combined with Table 3C, this evidence reinforces growing evidence that the effects of college financial aid arise primarily through initial enrollment rather than persistence (conditional on enrollment). Angrist et al. (2020) draw the same conclusion from the Buffet Scholars program; and recall that the Wisconsin Scholars program, which estimates effects on persistence conditional on enrollment, is one of the rare studies to find no effects on completion. This reinforces the idea that decisions during high school, a key contribution of the present study, are critical to college outcomes.

The bottom of Table 3D also reports effects on employment, earnings, incarceration, and teen pregnancy observed up to four years after on-time high school graduation (roughly age 22).⁵⁴ The point estimates generally suggest positive effects on employment/enrollment, though

⁵³ The Appendix also reports effects on college enrollment by semester, which are fairly flat and insignificantly different from zero in each year.

⁵⁴ With pregnancy, we limit to the high school years to align with the idea of teen pregnancy; however, we see no effects on pregnancy in later years either.

these are imprecise. The results across life outcomes are also not robust across specifications, which may reflect baseline nonequivalence that is not well accounted for with these covariates.⁵⁵ Given the positive effects on college completion, and the short window of post-high school time we can observe, it may simply be too soon for effects on labor market outcomes, in particular, to emerge.

V.E. Treatment on Treated (TOT)

As is now standard in analyses of RCTs, our TOT estimates use assignment to treatment as an instrument for being treated, in the sense that students met the requirements by the end of high school and were sent a congratulatory letter indicating eligibility to receive the funds. In an RCT, this instrument generally satisfies the exclusion restriction and easily passes the first stage, so that we obtain a LATE for compliers. Since only 21 percent of the assigned students met the requirements, the treatment effect is, predictably, about 4-5 times larger for treated students in Table 4 relative to the ITT estimates in Table 3C.⁵⁶ The effects on two-year college graduation, in particular, average roughly three percentage points and remain precisely estimated.

In this study, however, even the offer of funds might have had some effect on who attended college (see the hypotheses in section II pertaining to high school effort), which would violate the exclusion restriction for the TOT. This likely introduces little bias given the small high school effects. As an alternative, we also attempted to estimate the effects using a regression discontinuity (RD) design, as in Scott-Clayton (2011), using only the treatment group. However, this method suffers from the same problem. The same high school effects from the offer that call into question the exclusion restriction in the TOT also raise the possibility of

⁵⁵ Recall that the covariates in Models 1-3 are academic and demographic measures. We cannot account for baseline differences in neighborhood crime rates, for example, which could alter the results.

⁵⁶ Quan (2020) also estimates effects using a regression discontinuity (RD) design in a within-study comparison with the IV-TOT estimation. Her results are similar.

endogenous forcing variables in the RD. In addition, the RD analysis is underpowered because of the small number of observations near all the various merit thresholds.⁵⁷ Moreover, the RD estimate is just a LATE and the local linear regressions (see Figure 4B) therefore show all the LATEs across the GPA distribution.

In any event, the RD results are qualitatively similar to the ITT and TOT (see Appendix C), suggesting null effects on college attendance and the same substitution pattern from four- to-year colleges shown earlier.

V.F. Effect Heterogeneity

There are good reasons to expect effect heterogeneity in the above effects; however, this type of analysis poses distinctive problems in paired cluster RCTs. Some clusters lack student observations because: (a) some clusters are small; and (b) students sort themselves along a variety of dimensions so that some subgroups simply do not exist in some schools.⁵⁸ When a cluster is dropped for any combination of these reasons, the whole pair has to be dropped in this design. In addition to the reduced statistical power, this creates a further complication that each subgroup effect is estimated from a different sample of schools. To ensure that the results are most comparable across subgroups, we identify a subset of schools where all of the estimates are feasible, re-estimate the average treatment effects to ensure comparability, and estimate the effect heterogeneity for every subgroup using the same subset. We also conduct baseline tests for each subgroup analogous to those summarized in Table 2 (available upon request).⁵⁹ In the

⁵⁷ This discussion focuses on college enrollment, which is the best-case scenario. The estimates are even more imprecise when we focus on college completion, given how rare completion is in this case.

⁵⁸ For these reasons, when estimating models (3) and (4), I re-estimate the school averages for the covariates using only the values in the subgroup. Also, note that missing data in the survey measures compounds the above problems.

⁵⁹ These baseline equivalence tests are not possible with the survey measures because: (a) only a few of the relevant survey measures were available pre-treatment; and (b) students were not linked to their responses at the individual level, so there is no way to place students into sub-groups in a way that allows estimates for the survey measures.

discussion of each type of effect heterogeneity, all instances of baseline nonequivalence are noted in the text. In all of the subgroup analyses, we focus on regression model 4 with the full set of covariate adjustments.

V.F.1. Effects by Academic Endowment: GPA and Test Scores

In section II, our theory suggested that one potential reason for using merit requirements is the incentive for effort during high school (*Hypotheses 3a* and *3b*; and Figure 1A). We see only weak evidence of this. Figure 4A reports piecewise local linear regressions.⁶⁰ When the baseline GPA was near or below the threshold, the regression line shows a small upward shift in (post-treatment) GPA and attendance, but these are very imprecise. We also see no evidence of larger near-threshold effects on test scores.

A second reason for using merit incentives is targeting aid to those who respond most to it, which would suggest a general increase in treatment effects with baseline GPA.⁶¹ We see no evidence of this in Figure 4B. These patterns may be an artifact of imprecision, but this still provides suggestive evidence that contradict the arguments for performance incentives. We also test for targeting effects more directly in Appendix D, estimating all four regression models with interactions of treatment and baseline academic measures that might serve as eligibility requirements (GPA and test scores), but again we find no evidence of targeting and some evidence of “reverse-targeting,” i.e., giving the funds to students who are least responsive to it.

⁶⁰ The estimates are piecewise in the sense that we allowed the estimates to be separate on either side of the academic threshold. The local linear regression interacts the treatment effects with a quartic polynomial of the academic endowment variables.

⁶¹ Attendance was also a requirement, but far more students were below the GPA bar than the attendance bar, and very few students were above the GPA bar and not the attendance bar. This makes GPA the more relevant merit requirement.

V.F.2. Effects by Age/Grade

In section II, standard economic theory suggested that high school effort would be largest for 12th graders and salience suggested that we might also see effects in 9th grade, near the initial program announcement (modified *Hypothesis 6*). To test this, Figure 5 shows results by grade. Since the composition of the sample also changes across grades due to drop out and transfer out of the district, this is restricted to the sample of students whose data are available throughout the panel. These results are consistent with Figure 1B.⁶²

V.F.3. Effects by Program Implementation (Aid Communication)

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.⁶³ 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 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 person, such as Great Lakes, 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 was designed to provide both types of communications.

⁶² We also note that most of the low-cost efforts that increased as a result of TDP (FAFSA and participating in other college access programs) are also behaviors that are only feasible at the end of high school.

⁶³ Others have also emphasized the specific messages participants receive (Bloom, Hill, & Riccio, 2003).

To test for complementarities between treatment assignment and communications, we estimated a version of equation (5), adding an interaction term between treatment and the number of communications students reported receiving about TDP.⁶⁴ The mean number of communications is three (range: 0-19). This variation in communications is also associated with the TDP effects. Consistent with some prior research, the effects of TDP are more positive when accompanied by more communications (Table 4).⁶⁵ The results imply that if students had experienced 10 communications on average (a number that is above the actual mean but below the number of letters sent), then the predicted effect of TDP on initial college enrollment would have been 13 percentage points. 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. All students were sent the same letters. The program administrator also made the same communications with all high school counselors.

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 students did not receive those communications due to attrition (dropping out or leaving MPS schools). Second, interviews with school counselors suggest that some consciously avoided communicating with specific students who were below the thresholds; in fact, they even came to

⁶⁴ 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 (e.g., 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).

⁶⁵ Research by Oreopoulos and Dunn (2013) reinforces the potential importance of information in taking the steps toward college entry. Program implementation seems to have influences treatment effects in other areas of public policy (Bloom et al., 2003).

view the students themselves in a more negative light because of their failure to respond academically to meet the requirements (Rifelj & Kuttner, 2020). The communications from counselors also tended to lack substance of concrete guidance or advice (Kim & Rifelj, in press). 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.

V.F.4. Effects by Demographics: Race, Gender, and Income

Prior research suggests the possibility of variation in effects by gender, race, and family income (Angrist, Lang, & Oreopoulos, 2009; Scott-Clayton, 2011; Harris & Goldrick-Rab, 2012; Dynarski, 2013; Bartick et al., 2019), although few consistent patterns have emerged in the literature. Family income is the factor we would most expect to influence student responses, but we see no effect heterogeneity by income, race, or gender (results available upon request).

VI. Analysis of Optimal Financial Aid Policy

A key purpose of this study is to provide theory and evidence about optimal financial aid design. In this section, we combined evidence from above, and from other studies, regarding average treatment effects, with evidence on costs to understand the roles of the various policy design features.

The prior literature has generally expressed the effects of financial aid as degrees per \$1,000 in aid or as aid per degree (Deming & Dynarski, 2009). Harris (2013) and Bartik et al. (2016) carry out more formal cost-benefit analyses, but also focus on fiscal costs. As Angrist et al. (2020) point out, however, these approaches do not yield much information about social

welfare because financial aid is mainly a transfer payment. To address this, Angrist et al. (2020) focused instead on what colleges report as the official cost of attendance (COA) for the additional students who attended college as a result of the Buffet Scholars program. This approach, too, has limitations: First, prior studies do not include the marginal cost of funds (MCF), which arise because of the distortionary effects of taxation required to cover fiscal costs. Also, their use of COA over-states costs by including room and board, even though these costs are required even when students do not attend college.

We address all of these concerns and compare the social welfare gains of different program designs. For a government-funded program, the total cost of degrees is $(N \cdot p \cdot \varphi_d^*(g, A)) + (N \cdot \eta \cdot g)$ where N is the number of students receiving grant aid, $\varphi_d^*(g, A)$ is the share of students choosing college because of g , and η is the marginal cost of funds (MCF).⁶⁶

The core economic assumptions are as follows: $p = \$25,925 / \$114,425$ and the (PDV) of $w_i = \$150,000 / \$436,350$, for two- and four-year colleges, respectively.⁶⁷ The institutional costs exclude room and board and certain other costs unrelated to instruction. We take the MCF to be 1.25 (Ballard, Shoven, and Whalley, 1985). Appendix F provides additional explanation.

We apply this framework to TDP as well as other city/county-based programs in Kalamazoo, MI (Bartik et al., 2021), Knox County, TN (Carruthers & Fox, 2016), and Pittsburgh, PA (Page et al., 2018); the other U.S., broad-based RCT to find an effect of grant aid,

⁶⁶ Also, in section II, we defined p as the price that students faced. However, with government subsidies, this differs from the (shadow) price faced by the social planner. We therefore redefine p as the college resource costs incurred for those students induced by aid to attend college.

⁶⁷ Bartik et al. (2016) find that the equivalent figures are similar so that $w_i = \$133,800 / \$314,800$, but there estimates for the average student are much higher (and much higher than any other estimate we found). As discussed in Appendix F, the figures for p focus on instructional costs (excluding room and board) and come from NCES and the figures for w_i come from various other studies.

Buffett Scholars (Angrist et al., 2020); and the average aid effects from two literature reviews (Deming and Dynarski, 2009; Nguyen et al., 2019). In some of these studies, the effect on either two- or four-year sector graduation is precisely estimated but the effect on the other sector is imprecise. In the base case, we take the point estimate as the expected effect. However, we carry out robustness checks that treat all insignificant results as zero. Additional robustness checks assume lower returns to college degrees or a higher MCF of 1.50, as in Heckman et al. (2010).

The results are summarized in Table 5. As in Harris (2013), all of the programs pass a benefit-cost test and, with the base assumptions, the BCRs are in the range of 1.502-2.581, which suggests that financial aid consistently increases social welfare.⁶⁸ In the robustness checks, only two of the 21 BCRs are less than unity (TDP with low returns or a higher MCF).

We are mostly interested in *which* programs improve social welfare most. It is not obvious a priori. Free college reduces loan aversion and decision complexity, but free college requires larger grants on average than other kinds of programs and there might be diminishing returns to aid for most students. Early commitment might be helpful, but, given teenagers' apparently high discount rates and limited information about the production, these benefits may be negligible. Finally, while merit requirements seemed counter-productive in TDP, maybe this is not the case in other programs. By comparing BCRs across studies, we can get additional traction on the issue of optimal policy design.

The pattern of results across programs is generally supportive of the idea that financial aid programs improve social welfare more when designed with early commitment, free college, and limited merit requirements. The largest BCR comes from the Knox program (2.581), which is has all three open access elements. At the other end of the spectrum, TDP and Buffet, which

⁶⁸ Bartik et al., (2016) also find that the Kalamazoo program passed a cost-benefit test.

are merit-based and, consistently have the lowest BCRs. The Pittsburgh program also has merit requirements but a much higher BCR, which may be because the merit thresholds are much lower in Pittsburgh than in either TDP or Buffet and are therefore less binding for Pittsburgh students versus those in TDP.

It is also important to emphasize that merit requirements limit the number of students who benefit from a program (*Hypothesis 3a*). The BCRs do not account for this and really reflect the increase in social welfare per student. This means that, to increase social welfare, the BCRs for more restricted aid have to be higher to make up for the smaller share of students receiving the aid. The evidence is to the contrary. The BCRs tend to be smaller when aid is more restricted. Also, in the Nguyen et al. (2019) review, the effects per \$1,000 for merit-based programs are at or below the average program across all studies.

This welfare analysis entails several caveats. The degrees of freedom are low (i.e., too few studies for each combination of design elements). The standard errors of these estimates are no doubt very large (but difficult to estimate). The analysis is also partial equilibrium and neglects general equilibrium effects, e.g., political forces on the total resources provided, college responses with regard to quality, and labor market effects affecting w_i and therefore d^* .

The partial versus general equilibrium differences are also exacerbated when relying on evidence from RCTs of temporary, small-scale programs. Communication is more limited, narrow, and challenging in small-scale programs studied with RCTs, while scaled-up programs are promoted on billboards and on TV news. Also, the small-scale programs with RCTs limit the potential of spillover effects across students, e.g., the cost of each student's effort may be declining in the effort level of other students. Another potential spillover involves the behavior by K-12 educators. If generous financial aid substantially increases students' expectations, then

we might expect educators to change what they do to help students prepare. Some of these efforts, such as revamping curricula and lesson plans, and offering new college preparation courses, require upfront effort investments by teachers and counselors. Qualitative evidence suggests that, with the scaled-up Kalamazoo program, high schools changed how they served students to meet these new and higher expectations from the city's promise program (Miron et al., 2008, 2009). In contrast, we see little evidence of such effects with TDP (Rifelj & Kuttner, 2020; Kim & Rifelj, in press).⁶⁹ More generally, the effects of any program that would produce positive (negative) externalities in scale-up will be under-estimated (over-estimated) in RCTs of small, temporary programs. The single-cohort, RCT design of TDP may be why the effects and BCRs are smaller than the others, even if still above unity.

VII. Conclusion

With consistently rising costs and debt per student, and large gaps in college access by income, college affordability is a significant issue of modern economic policy. In this study, we have made six main contributions to understanding the policy issues involved, providing both theory and evidence from a rare RCT that captures effects of college financial aid from high school through college and beyond. We posed nine main hypotheses (14 in total counting the sub-theories and behavioral economic additions and amendments) and the results are consistent with all the hypotheses except two (*Hypothesis 4* regarding the inverse-U effort response and *Hypothesis 5* regarding incentive effects on merit requirement metrics). Overall, we find that college financial aid increases low-cost forms of effort at the end of high school and increases

⁶⁹ Using student surveys, we also estimated effects on school climate and educator expectations for students to attend college. We see positive point estimates, but these are imprecisely estimated (available upon request).

college graduation from two-year colleges, but has yet to show discernable effects on life outcomes.

The theory and evidence inform all five key dimensions of financial aid policy design. In addition to the aid level, merit requirements played a key role. We find almost no evidence of incentive or targeting effects, so that the ITT effects are almost certainly smaller than they would have been without the merit requirements. Also, consistent with prior research, restrictions on college eligibility may shift students to lower quality colleges. This has important implications state merit aid programs and for federal proposals, such as that being put forth by President Biden, that involve federal-state partnerships.

Free college is another key dimension of the current debate and our results are generally consistent with the arguments made by its advocates. The substitution of four-year for two-year college appears to reflect a response to the free college design/framing in the two-year sector. In shifting to two-year colleges, most students actually sacrificed funding they would have received at four-year colleges, but they switched anyway, apparently because of loan aversion.

Early commitment is another feature of financial aid that could affect program efficiency through students' high school effort, which, in turn, could influence college and life outcomes. We find that TDP succeeded in increasing low-cost student efforts, especially at the end of high school, but had no influence on the academic measures that were part of the merit requirements and that demanded more sustained effort and changes in student habits in earlier grades. These high school effort effects were likely attenuated because of the merit requirements and the small-scale RCT design, but the high discount rates of teenagers also suggest some limits on the potential of college financial aid to improve high school outcomes by changing college policies.

As *Hypothesis 8* showed, these three main policy decisions are intertwined. “Early commitment” is a misnomer if it comes with eligibility requirements that can only be verified at the end of high school. Eligibility requirements also blunt the effect of free college and other forms of aid by reducing the number of students who receive aid, without any accompanying incentive and targeting effects that could improve program efficiency.

In addition to policy design, financial aid implementation appears important. Aid requires communication, which varies in terms of frequency, framing, and sources. If the letters from program administrators had not mentioned that TDP would cover all tuition and fees at the local technical college, and instead focused on how students going to two-year colleges would generally get to use fewer TDP funds, then the results of this experiment likely would have been different. The communications students received also varied considerably across students, partly because of how counselors made sense of the policies and reacted to them.

It does seem clear that financial aid policies of almost all sorts generally improve social welfare. The average effects are positive (Deming & Dynarski, 2009; Nguyen et al., 2019) and those averages estimates pass our cost-benefit test. The distribution of those benefits is an additional consideration, especially given the wide disparities in college outcomes across income groups, even when controlling for academic background (Bailey & Dynarski, 2011). If we define equity as the share of funds going to low-income families, then need-based aid will likely be most equitable, almost by definition. Merit requirements, in contrast, reduce this form of equity because of the positive correlation between income, academic endowments, and college attendance probabilities. However, if we define equity in terms of what aid programs likely do the most to improve the welfare of low-income families, regardless of what they do for high-income families, the calculus is different. In that case, open-access aid, without need or other

requirements, will apparently do the most to help low-income families. This is because, for example, low-income families face the greatest hurdles in navigating complex processes, have higher discount rates, and are most averse to risk and loans, so that the effect of targeting aid by income, however well-intentioned, may reduce program effectiveness for those the most disadvantaged students.

The above results for TDP cannot be explained by either the standard economic theory or behavioral economics alone. In section II, we showed how the two frameworks can be used to create and test alternative hypotheses. Standard economic theory can explain how aid increases college attendance and why the effects on high school effort are delayed until 12th grade. But behavioral economics is necessary to explain the response to the free college design/framing and the pattern of responses across baseline GPA levels. Understanding responses to student aid programs, and likely other youth interventions, appears to require both theoretical frameworks.

None of this proves that one policy is better than all others. Some other programs have larger social benefits than financial aid (Harris, 2013). Some have made the same argument about income-contingent loans represent an alternative that may have benefits similar to free college, at lower cost. Also, over the next decade, additional studies will emerge regarding the plethora of a free college programs being adopted in states and cities throughout the country. This study provides a framework within which to design and interpret these studies and to understand their implications. In the meantime, as policymakers debate what may be the most important change in federal higher education policy since the Pell grant or GI Bill, the lessons of this unusual experiment may help draw attention to the key design issues and provide guidance about designing financial aid to best serve the needs of students and society.

References

- Anderson, D.M., Broton, K.M., Goldrick-Rab, S. & Kelchen, R. (2020) Experimental Evidence on the Impacts of Need-Based Financial Aid: Longitudinal Assessment of the Wisconsin Scholars Grant. *Journal of Policy Analysis and Management* 39(3): 720-739.
- Angrist, J., Autor, D., & Pallais, A. (2020) Marginal Effects of Merit Aid for Low-Income Students. *Working Paper 27834*. Cambridge, MA: National Bureau of Economic Research.
- Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1), 1–28.
- Angrist, J., Oreopoulos, P., & Williams, T. (2010). *When opportunity knocks, who answers? New evidence on college achievement awards* (NBER Working Paper No. 16643). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w16643>.
- Angrist, J. & Pischke, J-S. (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.
- Avery, C. & Kane, T.J. (2004). Student Perceptions of College Opportunities: The Boston COACH Program.” In *College Choices: The Economics of Where to Go, When to Go, and How To Pay for It*, edited by Caroline Hoxby, 355-394. Chicago: University of Chicago Press.
- Avery, C. & Turner, S. (2012). Student Loans: Do College Students Borrow Too Much—Or Not Enough? *Journal of Economic Perspectives* 26(1): 165–192.
- Bailey, M.J. & Dynarski, S.M. (2011). Gains and gaps: Changing inequality in U.S. college entry and completion. NBER Working Paper 17633. Cambridge, MA: National Bureau of Economic Research.
- Ballard, C.L., Shoven, J.B., and Whalley, J. (1985). General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States. *The American Economic Review* 75(1): 128-138.
- Barrow, L., & Malamud, O. (2013). Is College a Worthwhile Investment? *Annual Review of Economics* 7:519–55.
- Barrow, L., Richburg-Hayes, L., Rouse, C. A., & Brock, T. (2014). Paying for Performance: The Education Impacts of a Community College Scholarship Program for Low-Income Adults. *Journal of Labor Economics* 32(3): 563-599.
- Bartik, T.J., Hershbein, B., & Lachowska, M. (2016). The Merits of Universal Scholarships: Benefit-Cost Evidence from the Kalamazoo Promise. *Journal of Benefit Cost Analysis* 7(3):400–433.
- Bartik, T.J., Hershbein, B., & Lachowska, M. (2021). The Effects of the Kalamazoo Promise Scholarship on College Enrollment and Completion. *Journal of Human Resources* 56(1): 269-310.
- Bell, R.M., & McCaffrey, D.F. (2002). Bias Reduction in Standard Errors for Linear regression with Multi-stage Samples. *Survey Methodology* 28.
- Benhassine, Nagy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Poulouen (2013) Turning a Shove into a Nudge? A “Labeled Cash Transfer” for Education. NBER Working Paper No. 19227. Cambridge, MA: National Bureau of Economic Research.
- Bettinger, E. (2004). How financial aid affects persistence. In C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 207–238). Chicago, IL: University of Chicago Press.

- Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2009). *The role of simplification and information in college decisions: Results from the H&R Block FAFSA experiment* (NBER Working Paper No. 15361). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w15361>
- Bettinger, E., Gurantz, O., Kawano, L., Sacerdote, B., & Stevens M. (2019). The Long-Run Impacts of Financial Aid: Evidence from California's Cal Grant. *American Economic Journal: Economic Policy* 11(1): 64-94.
- Bloom, H.S., Hill, C.J., & Riccio, J.A. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*. 22(4): 551-575.
- Boatman, A., Evans, B.J., & Soliz, A. (2017). Understanding Loan Aversion in Education: Evidence from High School Seniors, Community College Students, and Adults. *AERA Open* 3(1): 1-16.
- Brown, P. R., Thornton, K., Ross, D., Smith, J. A., & Wimer, L. (2020). *Technical report on lessons learned in the development of the Institute for Research on Poverty's Wisconsin Administrative Data Core*. Institute for Research on Poverty, University of Wisconsin.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). Bootstrap-based improvement for inference with clustered errors. *Review of Economics and Statistics*. 90(3): 414-427.
- Carlson, D.E., Elwert, F., Hillman, N., Schmidt, A., & Wolfe, B.L. (2019). The Effects of Financial Aid Grant Offers on Postsecondary Educational Outcomes: New Experimental Evidence from the Fund for Wisconsin Scholars. *NBER Working Paper 26419*. Cambridge, MA: National Bureau of Economic Research.
- Carneiro, P.M., Heckman, J.J., 2003. Human Capital Policy. In: Heckman, J.J., Krueger, A.B. (Eds.), *Inequality in America: What Role for Human Capital Policies?* MIT Press, Cambridge, MA.
- Carruthers, C.K. & W.F. Fox (2016). Aid for all: College coaching, financial aid, and post-secondary persistence in Tennessee. *Economics of Education Review* 51: 97-112.
- Castleman, B. & Page, L. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization* 115, 144-160.
- Chetty, R., Looney, A., & Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review* 99(4): 1145-77.
- Cohodes, S.R. & J.S. Goodman (2014). Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics* 6(4): 251-85.
- College Board (2016). *Trends in College Pricing 2016*. New York City.
- Cornwell, C., Mustard, D., & Sridhar, D. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE Scholarship. *Journal of Labor Economics*, 24, 761-786.
- Delta Cost Project (2010). *Trends in College Spending 1998-2008*. Lumina Foundation.
- Deming, D., & Dynarski, S. (2009). *Into college, out of poverty? Policies to increase the postsecondary attainment of the poor* (NBER Working Paper No. 15387). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w15387>
- Dynarski, S. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3), 629-661.

- Dynarski, S. (2002). The behavioral and distributional implications of subsidies for college. *American Economic Review*, 92(2), 279–285.
- Dynarski, S. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1), 279–288.
- Dynarski, S. (2005). *Building the stock of college-educated labor* (NBER Working Paper No. 11604). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w11604>
- Dynarski, S., & Scott-Clayton, J. (2006). The cost of complexity in federal student aid: Lessons from optimal tax theory and behavioral economics. *National Tax Journal*, 59(2), 319–356.
- Dynarski, S., & Scott-Clayton, J. (2008). Complexity and Targeting in Federal Student Aid: A Quantitative Analysis. *NBER Working Paper No. 13801*. Cambridge, MA; National Bureau of Economic Research.
- Dynarski, S. Hemelt, S.W., & Hyman, J.M. (2013). The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes. NBER Working Paper 19552. Cambridge, MA: National Bureau of Economic Research.
- Dynarski, S., & Scott-Clayton, J. (2013). Financial Aid Policy: Lessons from Research. NBER Working Paper 18710. Cambridge, MA: National Bureau of Economic Research.
- Fitzpatrick, M.D. & D. Jones (2016) Post-baccalaureate migration and merit-based scholarships. *Economics of Education Review* 54: 155-172.
- French, R. & Oreopoulos, P. (2017). Behavioral barriers transitioning to college. *Labour Economics* 47: 48-63.
- Freedman, D.A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics* 40, 180–93.
- Fryer, R. G., Jr. (2010). *Financial incentives and student achievement: Evidence from randomized trials* (NBER Working Paper No. 15898). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w15898>
- Goldin, C., & Katz, L. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.
- Goldrick-Rab, S., Harris, D., & Trostel, P. (2009). Why financial aid matters (or does not) for college success: Toward a new interdisciplinary approach. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 24, pp. 1–45). New York, NY: Springer.
- Goldrick-Rab, S., Kelchen, R., Harris, D.N., & Benson, J. (2016). Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology* 121(6), 1762-1817.
- Goldrick-Rab & Kendall (2014), *Redefining College Affordability: Securing America's Future with a Free Two Year College Option*. Lumina Foundation. https://www.luminafoundation.org/files/publications/ideas_summit/Redefining_College_Affordability.pdf
- Goodman, J. (2008). Who merits financial aid?: Massachusetts' Adams Scholarship. *Journal of Public Economics* 92, 2121–31.
- Harbaugh, W., Krause, K., and Vesterlund, L. (2002). Risk attitudes of children and adults: choices over small and large probability gains and losses. *Experimental Economics* 5: 53–84.
- Harris, D. (2013). Applying cost-effectiveness analysis in higher education. In A. Kelly and K. Carey (eds.). *Stretching the Higher Education Dollar*. (pp. 45-66). Washington, DC: American Enterprise Institute.

- Harris, D. & Goldrick-Rab, S. (2012). Improving the productivity of educational experiments: Lessons from a randomized trial of need-based financial aid. *Education Finance and Policy*, 7(2): 143-169.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94 (1–2), 114–128.
- Hoxby, C. M. & Avery, C. (2012). The missing “one-offs”: The hidden supply of high-achieving, low-income students (NBER Working Paper No. 18586). Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C.M. & Turner, S. (2015). What High-Achieving Low-Income Students Know about College. *American Economic Review* 105(5): 514-17.
- Hoxby, C.M. & S. Turner (2013). Expanding College Opportunities. *Education Next* 13(4).
- Hryshko, D., Luengo-Prado, M., & Sorensen, B. (2011). Childhood determinants of risk aversion: The long shadow of compulsory education. *Quantitative Economics* 2: 37-72
- Ikenberry, S. O., & Hartle, T. W. (1998). *Too Little knowledge is a dangerous thing: What the public thinks and knows about paying for college*. Washington, DC: American Council on Education.
- Imai, K., King, G., & Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*, 24(1), 29–53.
- 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.
- Kim, D., & Rifelj, K.K. (in press). Packaging the promise: Money, messaging, and misalignment. *Teachers College Record*.
- Lawrence, E. C. (1991). Poverty and the rate of time preference: Evidence from panel data. *Journal of Political Economy*, 99, 54–77.
- Liang, Kung-ye, and Scott L. Zeger, (1986). Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* 73, 13-22.
- MacKinnon, J.G., and H. White, (1985) Some Heteroscedasticity-Consistent Covariance Matrix Estimators with Improved Finite-Sample Properties, *Journal of Econometrics* 29, 305-325.
- Marx, B. M., & Turner, L. J. (2018). Borrowing trouble? Human capital investment with opt-in costs and implications for the effectiveness of grant aid. *American Economic Journal: Applied Economics*, 10(2), 163–201.
- McGuire, M. Pratt, J., & Zeckhauser (1991). Paying to Improve Your Chances: Gambling or Insurance? *Journal of Risk and Uncertainty* 4: 329-338.
- Miron, G., Jones, J. N., & Kelaher Young, A. J. (2009). *The impact of the Kalamazoo Promise on student attitudes, goals, and aspirations* (Evaluation of the Kalamazoo Promise: Working Paper #6).
- Miron, G., Spybrook, J., & Evergreen, S. (2008). *Key findings from the 2007 survey of high school students* (Evaluation of the Kalamazoo Promise: Working Paper #3).
- Mulhern, C. (2019). Beyond teachers: Estimating individual guidance counselors’ effects on educational attainment. *Unpublished manuscript*.
- Nguyen, T.D., Kramer, J.W., & B.J. Evans (2019). The Effects of Grant Aid on Student Persistence and Degree Attainment: A Systematic Review and Meta-Analysis of the Causal Evidence. *Review of Educational Research* 89(6): 831-874.

- Oreopoulos, P. & Dunn, R. (2013). Information and College Access: Evidence from a Randomized Field Experiment. *The Scandinavian Journal of Economics* 115(1): pp. 3-26.
- Page, L. C., Iriti, J. E., Lowry, D. J., & Anthony, A. M. (2019). The promise of place-based investment in postsecondary access and success: Investigating the impact of the Pittsburgh Promise. *Education Finance and Policy* 14(4): 572-600.
- Pustejovsky, J.E. & Tipton, E. (2016). Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. Unpublished manuscript.
- Quan, S. & Harris, D.N. (2020). Do Regression Discontinuity Methods Yield The Same Result As Randomized Control Trials? Evidence from an Early Aid College Scholarship Program. *Unpublished working paper*.
- Rehavi, M.M. & Starr, S.B. (2014). Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy* 122(6): 1320-1354.
- Rifelj, K.K. & Kuttner, P. (2020). Evidence of Failure: How high school counselors and administrators make sense of promise scholarship merit requirements. *Teachers College Record* 122(7).
- Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty* 1(1): 7-59.
- Samwick, A.A. (1998). Discount rate heterogeneity and social security reform. *Journal of Development Economics*: 117–146.
- Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*.46(3), 614-646.
- Scott-Clayton, J. & Zafar, B. (2019). Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid. *Journal of Public Economics* 170: 68-82.
- Seftor, N., & Turner, S. (2002). Back to School: Federal Student Aid Policy and Adult College Enrollment. *Journal of Human Resources*. Vol. 37, No. 2, pp. 336-352.
- Shampanier, K., Mazar, N., Ariely, D. (2007). Zero as a Special Price: The True Value of Free Products. *Marketing Science* 26(6).
- Sjoquist, D.L. & Winters, J.V. (2012). Building the stock of college-educated labor revisited. *The Journal of Human Resources*, 47(1): 270-285.
- Skaperdas, S. & Gan, L. (1995). Risk Aversion in Contests. *The Economic Journal*, 105 (July), 951-962.
- Swanson, E., Watson, A., Ritter, G.W., & Nichols, M. (2017). *Promises Fulfilled? A Systematic Review of the Impacts of Promise Programs (December 18, 2016)*. EDRE Working Paper No. 2016-16. <http://dx.doi.org/10.2139/ssrn.2849194>.
- The Institute for College Access and Success (2020). *Student Debt and the Class of 2019*. Retrieved from: <https://ticas.org/wp-content/uploads/2020/10/classof2019.pdf>.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- U.S. Department of Education, National Center on Education Statistics (2011). Nation’s Report Card Trial Urban District Assessment: Mathematics (NCES 2012-452). Retrieved from: <http://nces.ed.gov/nationsreportcard/pdf/dst2011/2012452.pdf>.
- Valant, J. & Newark, D. (2020). The Word on the Street or the Number from the State? Government-Provided Information and Americans’ Opinions of Schools. *Journal of Public Administration Research and Theory* 30(4): 674–692.

Figure 1A: High School Effort Response by Academic Endowment in in Response to Merit Requirements (Standard Theory)

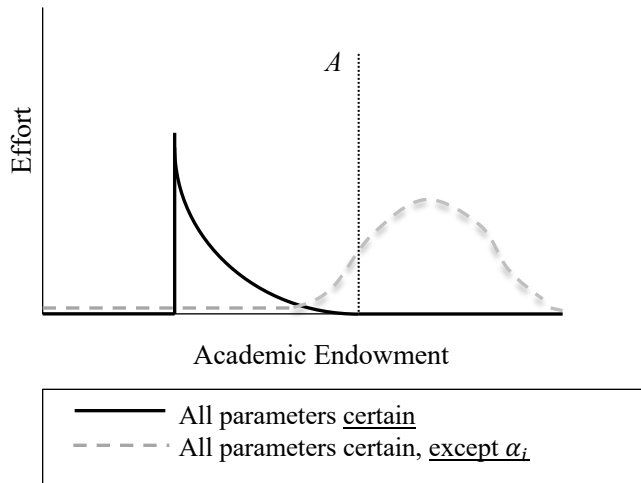


Figure 1B: High School Effort by Age/Grade (Standard Theory)

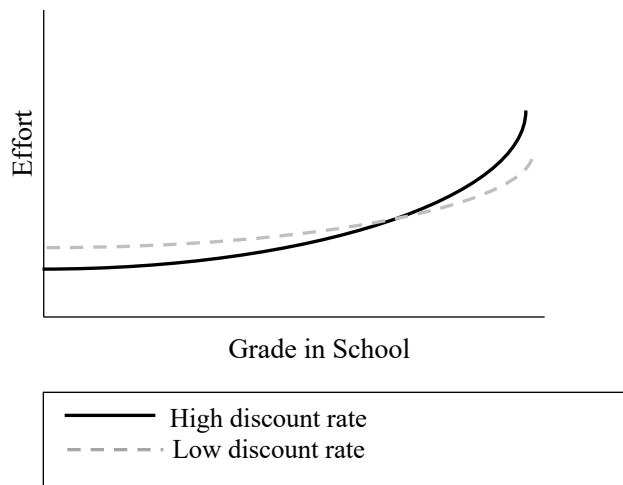
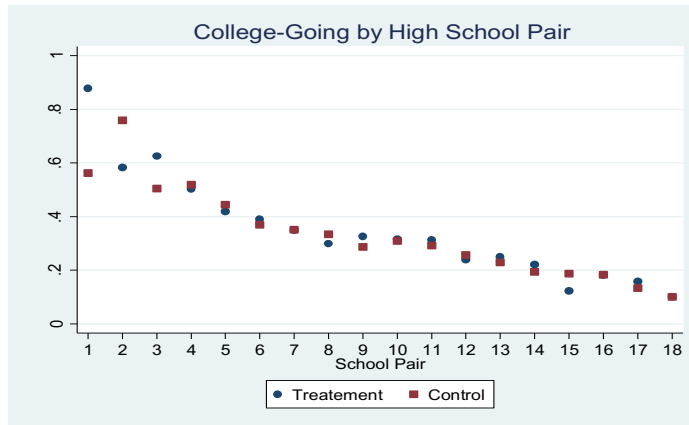
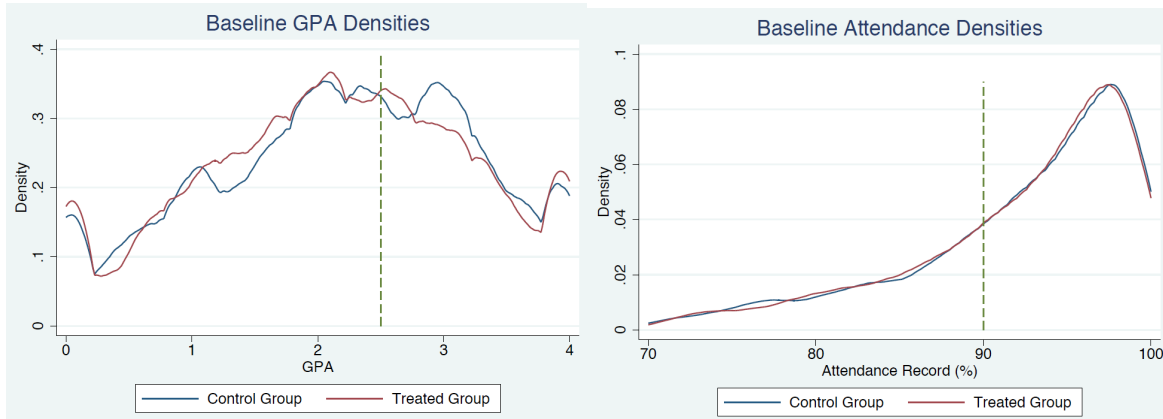


Figure 2: Baseline Equivalence



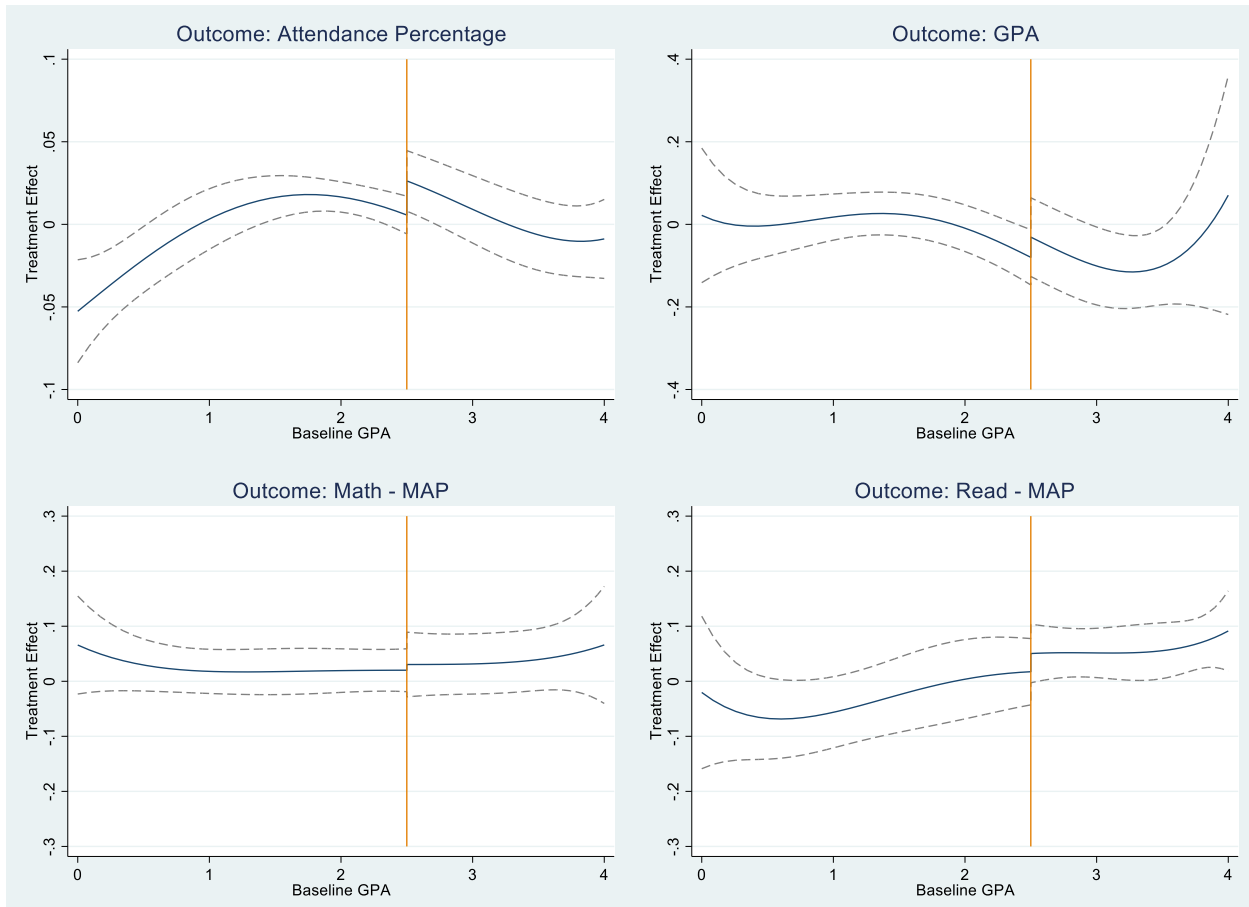
Notes: In Figure 2, each circle (square) is the pre-treatment on-time college entry rate for each treatment (control) school. These are shown for each pair of schools where randomization occurred within each pair. In the pairing process, we stratified according to whether the school had actual versus predicted attendance rate, which explains why the control-treatment differential is large in the first three pairs.

Figure 3: Baseline Equivalence Distributions for GPA and Attendance



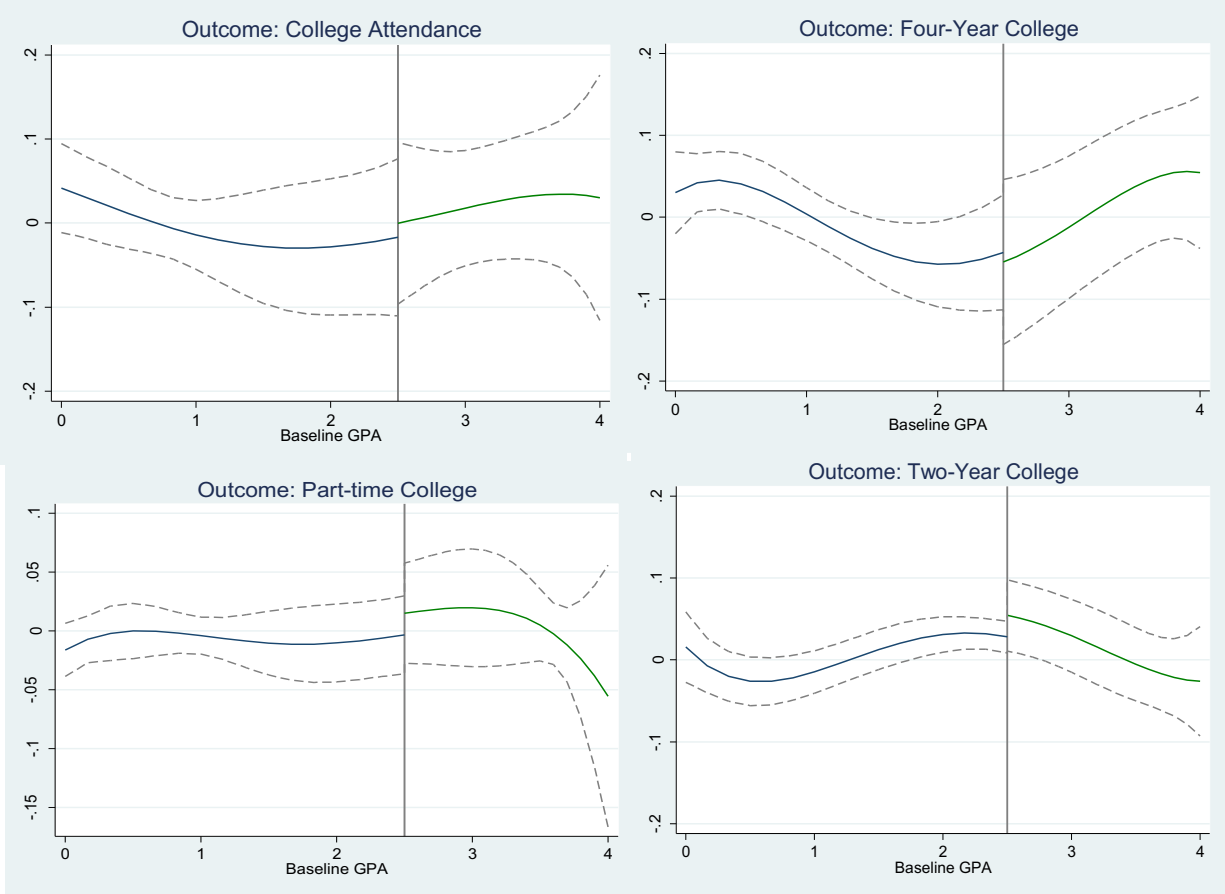
Notes: Figure 3 shows kernel density plots for GPA and attendance for the control and treatment groups. These are the main two academic requirements that can be tested using baseline data (high school graduation is the main additional requirement).

*Figures 4A: Test of Incentive Effects:
Local Linear Regression Effects by Baseline GPA on High School Academic Outcomes*



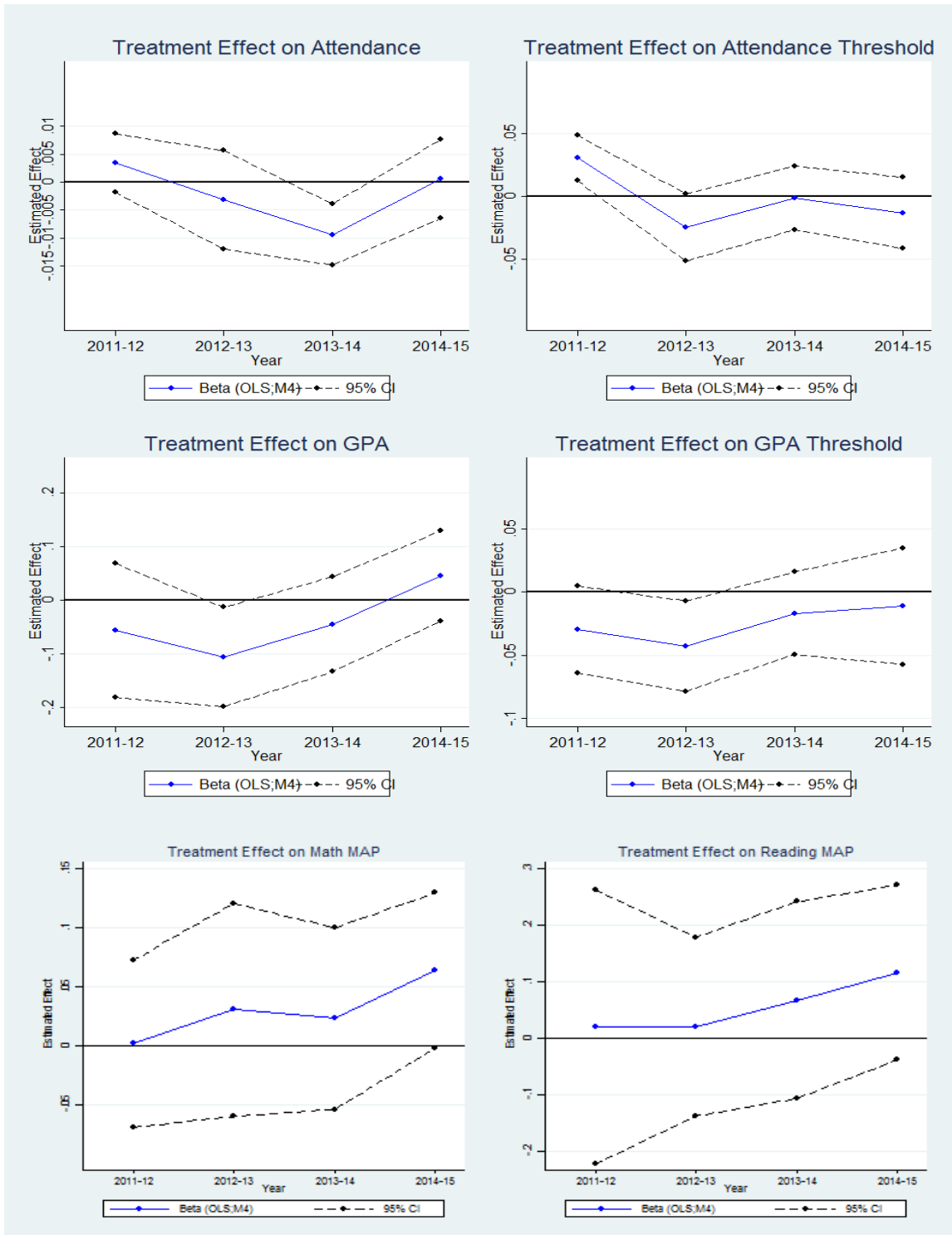
Notes: The panels of Figure 4A test the standard theory that the effort response function peaks near the threshold and is null in the tails (the inverse-U shape show in Figure 1A). The academic endowment in this case is baseline GPA. (While TDP entailed other merit requirements, GPA was most likely, by far, 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 4B: Test of Aid Targeting:
Local Linear Regression Effects on College Outcomes by Baseline GPA*



Notes: The panels of Figure 4B 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 GPA to the right of the threshold. See additional notes in Figure 4A. 95% confidence intervals are shown.

Figure 5: Treatment Effects by Grade



Notes: Figure 5 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.

Table 1: Data and Variables

<i>Outcomes</i>	<i>Sources</i>	<i>Years</i>	<i>Attrition/ Missing Data</i>
<i>High School (Academic)</i> <i>Variables tied to TDP Req:</i> GPA, attendance, grad <i>Other variables:</i> 10 th grade test scores, transfer	MPS (admin)	2012-2016 (freshmen to one year post-on-time- HS-grad)	Minimal (see notes)
<i>High School (Other)</i> FAFSA completion, college access programs, college expectations	MPS (climate survey, senior exit survey)	2012-2015 (freshmen thru senior year)	53% response rate (control-treatment differential: 3-5 pp)
<i>College</i> College entry, persistence, graduation, institution/sector	NSC (through MPS)	2009, 2015-2019 (four years post- on-time-HS-grad)	Almost none
<i>Life</i> Employment, earnings, crime, pregnancy	UW-IRP-MSPF	2015-2017 (two years post- on-time-HS-grad)	Almost none

Notes: The high school academic outcomes come from school transcripts that are provided to MPS. For this reason, we lose students when they switch school districts, but this is not likely to lead to differential attrition unless TDP induced some students not to leave the city. The Senior Exit Survey data on college-going plans is only available for students who said they intended to attend college. The NSC data were collected for all MPS ninth graders (not only high school graduates). The years in Table 1 pertain to the spring year (e.g., 2011 means 2011-2012). The 2009 college data were used to calculate the baseline college-going rates for purposes of paired randomization (see details of that process in section III).

Table 2: Baseline Equivalence

	Full Sample			Senior Exit Survey Sample			
	N	Control Mean	Treatment Diff	Diff w/ Pairs	Control Mean	Treatment Diff	Diff w/ Pairs
Panel A: Dependent Variables (at Baseline)							
Attendance	4989	0.886	0.003 (0.027)	0.010 (0.008)	0.951	0.001 (0.010)	0.006 (0.003)
90% Threshold	4989	0.704	-0.001 (0.078)	0.019 (0.023)	0.886	-0.010 (0.041)	0.001 (0.015)
GPA	3199	2.164	-0.034 (0.242)	0.060 (0.041)	2.573	-0.019 (0.227)	0.120** (0.036)
2.5 Threshold	3199	0.421	-0.017 (0.106)	0.027 (0.021)	0.587	-0.022 (0.107)	0.042* (0.018)
Math-MAP	4460	0.009	-0.016 (0.233)	0.093*** (0.021)	0.382	-0.100 (0.218)	0.057** (0.020)
Read-MAP	4401	0.029	-0.056 (0.219)	0.050 (0.048)	0.391	-0.160 (0.186)	-0.035 (0.044)
Panel B: Independent Variables (at Baseline)							
Female	4995	0.501	-0.019 (0.037)	-0.024 (0.021)	0.577	-0.036 (0.037)	-0.044 (0.028)
Age	4995	15.147	0.005 (0.092)	0.001 (0.027)	14.929	0.034 (0.050)	0.014 (0.019)
Free/Red. Price Lunch	4995	0.803	0.055 (0.053)	0.036* (0.013)	0.740	0.083 (0.066)	0.039* (0.018)
Special Education	4995	0.220	-0.018 (0.033)	-0.023 (0.013)	0.122	-0.010 (0.023)	-0.017 (0.012)
ELL	4995	0.083	0.052 (0.051)	0.035 (0.035)	0.072	0.073 (0.042)	0.056* (0.025)
School Closed	5038	0.049	-0.013 (0.047)	-0.011 (0.038)	0.047	-0.027 (0.043)	-0.023 (0.036)
English at Home	4995	0.897	-0.078 (0.053)	-0.065 (0.039)	0.892	-0.104* (0.050)	-0.097** (0.032)
Spanish at Home	4995	0.068	0.0563 (0.042)	0.0444 (0.034)	0.06	0.0617 (0.039)	0.0519 (0.035)
Black	4995	0.669	-0.087 (0.105)	-0.088 (0.073)	0.626	-0.075 (0.115)	-0.099 (0.083)
Asian	4995	0.048	0.022 (0.028)	0.025 (0.022)	0.078	0.040 (0.051)	0.051 (0.045)
Hispanic	4995	0.153	0.089 (0.077)	0.076 (0.062)	0.135	0.090 (0.080)	0.082 (0.077)
White	4995	0.121	-0.022 (0.035)	-0.013 (0.019)	0.154	-0.053 (0.042)	-0.032 (0.032)
College-Going Probability	4818	0.297	-0.034 (0.079)	-0.022 (0.013)	0.423	-0.039 (0.087)	-0.006 (0.017)
Joint F-statistic	2763			11.54***			30.13***

Notes: Treatment differences are from OLS regression (with GEE standard errors) clustered at the school level. The last row provides the joint significance test from a regression of treatment status on the full vector of lagged dependent variables and student demographics using logistic regression. The results are qualitatively similar with logit. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3A: Average Treatment Effects on High School Academic Outcomes (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Attendance pct	4075 - 5033	0.813	0.004 (0.008)	-0.003 (0.004)	-0.004 (0.003)	-0.002 (0.004)
90% or above	4075 - 5033	0.469	0.002 (0.017)	-0.012 (0.010)	-0.010 (0.009)	-0.009 (0.009)
GPA	3158 - 4948	1.801	-0.006 (0.032)	-0.039 (0.030)	-0.009 (0.036)	-0.028 (0.027)
2.5 or above	3158 - 4948	0.263	-0.009 (0.011)	-0.017* (0.008)	-0.008 (0.009)	-0.015* (0.006)
Meets Both DP reqs.	3090 - 4948	0.228	-0.008 (0.012)	-0.017 (0.010)	-0.006 (0.010)	-0.015 (0.009)
Math MAP	3587 - 4761	-0.057	0.085** (0.024)	0.023 (0.021)	0.002 (0.015)	0.022 (0.022)
Read MAP	3534 - 4753	-0.021	0.026 (0.037)	0.002 (0.028)	-0.014 (0.033)	-0.001 (0.028)
Transferred schools	3944 - 5038	0.407	0.057 (0.043)	0.052 (0.044)	0.108*** (0.023)	0.052 (0.043)
Missing in 2014-15 data	3944 - 5038	0.261	0.001 (0.020)	-0.016 (0.019)	-0.004 (0.017)	-0.019 (0.017)
Grad On Time w/ Reg Diploma	3944 - 5038	0.505	0.006 (0.018)	0.006 (0.015)	-0.018* (0.007)	0.007 (0.015)
Grad w/ Any Credential, Anytime	3944 - 5038	0.550	0.019 (0.016)	0.021 (0.012)	-0.001 (0.010)	0.022 (0.012)
Baseline Performance (i)				X		X
Baseline Performance (j)					X	
Student Covariates (i)						X

Notes: The table reports effects based on treatment assignment from Ordinary Least Squares (OLS) estimation. All models include randomization pair indicators. Each coefficient from a separate regression. Standard errors are clustered (GEE) by original high school attended. “Baseline Performance (i)” include cubic models of student-level math performance and attendance percentage calculated prior to random assignment. “Baseline Performance (j)” is the same but at the school-level. “Student Covariates” include the student-level controls listed in Table 2 (Panel B) and an indicator for whether the school students were originally assigned closed or re-organized after the 2011-12 school year, possibly forcing students to switch schools. Control means are from the full TDP sample, but vary only slightly in the sub-samples used when covariates are added. There are 5,038 students in the TDP sample, but the low end of this range is often smaller due to missing data on the covariates. Results are robust with logit for dichotomous dependent variables.

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

Table 3B: Average Treatment Effects on College Expectations and Non-Merit Effort (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
College expectations						
Planning to attend college	1839 - 2146	0.680	0.041* (0.015)	0.014 (0.015)	0.023 (0.013)	0.014 (0.014)
Planning on 4-year college	1839 - 2146	0.465	0.025 (0.024)	0.022 (0.028)	0.008 (0.018)	0.024 (0.027)
Planning on 2-year college	1839 - 2146	0.216	0.016 (0.020)	-0.008 (0.021)	0.015 (0.017)	-0.010 (0.022)
Planning full-time college	1837 - 2144	0.565	0.055* (0.020)	0.038 (0.023)	0.037* (0.014)	0.040 (0.021)
Steps to College						
# college support prog participated in	1839 - 2146	0.463	0.149** (0.054)	0.175** (0.062)	0.056 (0.043)	0.152* (0.058)
# colleges applied to	1839 - 2146	2.354	0.309* (0.096)	0.201 (0.101)	0.262* (0.064)	0.176 (0.97)
First choice college is highest prestige	1645 - 1920	0.140	-0.029 (0.018)	-0.026 (0.018)	-0.027* (0.012)	-0.022 (0.018)
Financial aid and college cost (senior year)						
Completed FAFSA	1826 - 2129	0.519	0.044** 0.015	0.023 0.015	0.029 0.015	0.025 0.015
Applied for scholarships	1250 - 1461	0.596	0.079** 0.026	0.075* 0.030	0.046 0.025	0.0718* 0.031
Awarded scholarships	694 - 808	0.393	0.066*** 0.017	0.078*** 0.018	0.077*** 0.019	0.075*** 0.016
Intends to use scholarships	1250 - 1461	0.589	0.095*** 0.013	0.085*** 0.010	0.081*** 0.008	0.079*** 0.011
Biggest roadblock to college is cost	1829 - 2135	0.557	-0.001 0.015	-0.015 0.013	-0.006 0.018	-0.019 0.012
Summer Melt						
Planned on college, but did not enroll	1839 - 2146	0.419	0.042 0.023	0.019 0.026	0.022 0.023	0.017 0.025
Planned on 4-year college, but enrolled in 2-year	1839 - 2146	0.027	0.003 0.006	0.002 0.007	0.011 0.006	0.001 0.007
Lagged Dependent Variable (i)				X		
Lagged Dependent Variable (j)					X	X
School Covariates (i)						X

Notes: See notes to Table 3A. The number of observations drops considerably here because of the survey response rate (combined with covariate adjustment). Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 3C: Average Treatment Effects on Initial College Outcomes (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Any College Enrollment	3624 - 5038	0.334	0.005 (0.016)	-0.006 (0.017)	-0.016 (0.013)	0.001 (0.016)
2-year college	3624 - 5038	0.198	0.012 (0.009)	0.006 (0.009)	0.001 (0.010)	0.011 (0.009)
4-year college	3624 - 5038	0.188	-0.004 (0.017)	-0.008 (0.022)	-0.017 (0.014)	-0.004 (0.021)
Part-time status	3624 - 5038	0.214	-0.002 (0.016)	-0.017 (0.017)	-0.020* (0.008)	-0.013 (0.017)
Full-time status	3624 - 5038	0.224	0.009 (0.016)	0.002 (0.019)	-0.011 (0.013)	0.009 (0.018)
In-state college	3624 - 5038	0.290	0.008 (0.012)	-0.001 (0.011)	-0.014 (0.008)	0.005 (0.009)
Out-of-state college	3624 - 5038	0.074	-0.012 (0.014)	-0.014 (0.016)	-0.013 (0.012)	-0.012 (0.016)
TDP eligible college	1294 - 1594	0.828	0.018 (0.028)	0.014 (0.029)	-0.032 (0.020)	0.014 (0.031)
Coll. competitiveness	769 - 935	0.929	0.015 (0.048)	-0.002 (0.042)	-0.025 (0.058)	-0.013 (0.042)
Coll. grad. rate	1273-1566	0.440	-0.011 (0.011)	-0.013 (0.013)	-0.010 (0.006)	-0.006 (0.013)
Undermatched college entry	666 - 803	0.079	0.017 (0.021)	0.003 (0.015)	-0.001 (0.013)	0.006 (0.016)
Lagged Dependent Variable (i)				X		
Lagged Dependent Variable (j)					X	X
School Covariates (i)						X

Notes: See notes to Table 3A. The college competitiveness and college graduation rate measures are institution-level and only available for students who attended a four-year college, according to the NSC data. The number of observations for summer melt are small because this requires the college expectations data. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 3D: Average Treatment Effects on College Graduation and Life Outcomes (ITT)

	N	Control mean	Model 1	Model 2	Model 3	Model 4
College graduation						
Ever graduated from college	3624 - 5037	0.020	0.005 (0.003)	0.006 (0.004)	0.003 (0.003)	0.004 (0.004)
Ever graduated from 2-year college	3624 - 5038	0.012	0.005* (0.002)	0.007** (0.002)	0.003* (0.001)	0.006* (0.002)
Ever graduated from 4-year college	3624 - 5038	0.007	0.000 (0.003)	-0.001 (0.003)	0.000 (0.002)	-0.002 (0.003)
Employment outcomes						
Ever employed	3624 - 5037	0.602	0.011 (0.014)	0.006 (0.017)	0.016 (0.011)	0.006 (0.016)
Ever employed or enrolled in college	3624 - 5037	0.719	0.008 (0.009)	0.000 (0.012)	0.004 (0.010)	0.005 (0.011)
Employed or enrolled in college in 2017	3624 - 5037	0.639	0.005 (0.009)	-0.004 (0.011)	0.003 (0.009)	-0.002 (0.009)
Earnings						
Full sample	3624 - 5037	5.061	0.141 (0.144)	0.004 (0.156)	0.167 (0.098)	-0.024 (0.157)
Excluding college enrollees in 2017	3036 - 4342	5.122	0.112 (0.106)	-0.026 (0.129)	0.164 (0.102)	-0.028 (0.115)
Ever incarcerated	3624 - 5037	0.011	-0.001 (0.002)	0.001 (0.002)	-0.004** (0.001)	0.000 (0.002)
Ever had child during High School	3624 - 5037	0.018	-0.002 (0.004)	-0.002 (0.003)	0.003 (0.004)	-0.001 (0.004)
Lagged Dependent Variable (i)				X		
Lagged Dependent Variable (j)					X	X
School Covariates (i)						X

Notes: See notes to Table 3A. College graduation outcomes last observed as of May, 2019. Employment, earnings, incarceration, and pregnancy data last observed in December, 2017. Earnings models estimated using an inverse hyperbolic sign transformation. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 4: 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	0.334	0.029 (0.090)	-0.031 (0.081)	-0.083 (0.068)	0.003 (0.075)
Ever enrolled in 2-year college	3601 - 4948	0.198	0.071 (0.054)	0.029 (0.047)	0.012 (0.059)	0.043 (0.048)
Ever enrolled in 4-year college	3601 - 4948	0.188	-0.022 (0.093)	-0.042 (0.103)	-0.096 (0.076)	0.078 (0.096)
College graduation						
Ever graduated from college	3601 - 4948	0.02	0.027 (0.015)	0.028 (0.019)	0.019 (0.017)	0.021 (0.019)
Ever graduated from 2-year college	3601 - 4948	0.012	0.030*** (0.008)	0.035*** (0.010)	0.020** (0.007)	0.030** (0.011)
Ever graduated from 4-year college	3601 - 4948	0.007	-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	0.602	0.071 (0.074)	0.043 (0.082)	0.092 (0.065)	0.043 (0.061)
Ever employed or enrolled in college	3601 - 4948	0.719	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	0.639	0.038 (0.053)	-0.009 (0.053)	0.019 (0.074)	0.034 (0.038)
Earnings (full sample; IHS)	3601 - 4948	5.061	0.877 (0.705)	0.122 (0.757)	0.919 (0.564)	0.596 (0.720)
Ever incarcerated	3601 - 4948	0.011	-0.008 (0.012)	0.005 (0.012)	-0.017** (0.006)	0.016 (0.013)
Ever had child during High School	3601 - 4948	0.018	-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.

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

Table 5: Benefit-Cost Analysis Results

Program	Study/Program Design	Fiscal Cost/ Student	Effects on 2y degree	Effects on 4y degree	Benefit-Cost Ratios			
					Base	Low Return to Ed	High MCF	Insignif. effects = 0
The Degree Project	RCT; Merit Req; Free 2y, Covers 4y; Last \$	\$3,357	3.00*	-0.50	1.502	0.984	0.973	1.901
Kalamazoo (Bartik et al., 2021)	DD; No Merit; Free 2y/4y; First \$	\$6,800	2.60	7.40*	2.381	1.427	2.141	2.381
TN-Knox (Carruthers & Fox, 2016)	DD; No Merit; Free 2y; Last \$	\$971	4.00*	-1.00	2.581	1.825	1.867	2.720
Pittsburgh (Page et al., 2019)	RD/DD; Merit Req; Covers 2y/4y; Not Free; Last \$	\$3,934	-3.00	7.70*	2.399	1.420	2.219	2.700
Buffet Scholars (Angrist et al., 2020)	RCT; Merit Req; Covers 2y/4y; Nearly Free; Last \$	\$8,200	-3.00	8.40*	2.241	1.327	1.961	2.702
Average 1 (Nguyen et al., 2019)	Mostly QED & No Merit; Last \$	\$1,000		2.00*	2.464	1.470	2.302	2.464
Average 2 (Deming & Dynarski, 2009)	Mostly QED & No Merit; Last \$	\$1,000		4.00*	2.555	1.523	2.464	2.555

Notes: Fiscal costs are the cost per grant recipient per year. Effects are in percentage points. Significant effect estimate are indicated with *. The bottom two rows are from summaries or “averages” from the literature and these do not distinguish two- from four-year degrees. The base BCRs are followed by several robustness checks. See above discussion and main text for details.

Appendices for “Optimal Financial Aid Policy”

Appendix A. Additional Background on Milwaukee

Table A1 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 A2, 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 Four of the 36 TDP sample schools can be considered selective admissions in that students have to apply and meet academic requirements. Partly as a result, the college-going rates varied greatly across high schools prior to TDP, ranging from 10 to 88 percent for on-time college enrollment.³

In addition to common programs such as the federal TRIO college access programs and other community-based efforts, MPS has attempted to emulate Chicago 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 (Farmer-Hinton, 2016).

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 While the sample includes almost all publicly funded schools, there were a few exceptions, as a few very small schools do not report sufficient data to MPS to check eligibility requirements.

³ 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.

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.

The MPS district leadership was also been engaged, around the time of TDP, in many other efforts to increase college-going during the experiment. These efforts were apparently successful as the percentage of students who met the TDP requirements (control and treatment) who went on to college increased by roughly 10 percentage points over the course of the project.⁴ This is informative regarding the counterfactual.

⁴ 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) met the requirements. These translate into 74.6 and 76.7 percent, respectively.

Table A1: Descriptive statistics for dependent variables

Variable	Full Sample					Senior Exit Survey Sample	
	N	Mean	s.d.	Min	Max	Mean	p-value diff w/ full sample
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		
<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

Variable	Full Sample					Senior Exit Survey Sample	
	N	Mean	s.d.	Min	Max	Mean	p-value diff w/ full sample
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
Native American	4995	0.009	0.092	0.000	1.000	0.006	0.043
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.

Table A2: 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 serving students with special needs); and (e) Partnership (same as Alternative except operated by a private provider under MPS contract). “Attendance” and “GPA” refer to the eighth grade information. “Prior College Attendance” is the college-going rate from a prior cohort (see text). “# TDP Comm.” refers to the number of communications about TDP students reported receiving (in treatment schools) in the first year of the program. Other information, such as demographics, are omitted to avoid identifying the specific schools.

Appendix B: Additional Information about Experiment/Randomization

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. 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 large in a few cases in Figure 1.

While the college attendance rate was the main criterion for pairing, we also considered school size and school test scores. Having roughly equal school sizes was desirable for Great Lakes to limit their risk exposure; if we happened to select larger schools, then this would increase the cost of the program. Accounting for school size was also beneficial for the analysis of impacts because balanced designs are somewhat more powerful.

Next, we provide 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 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

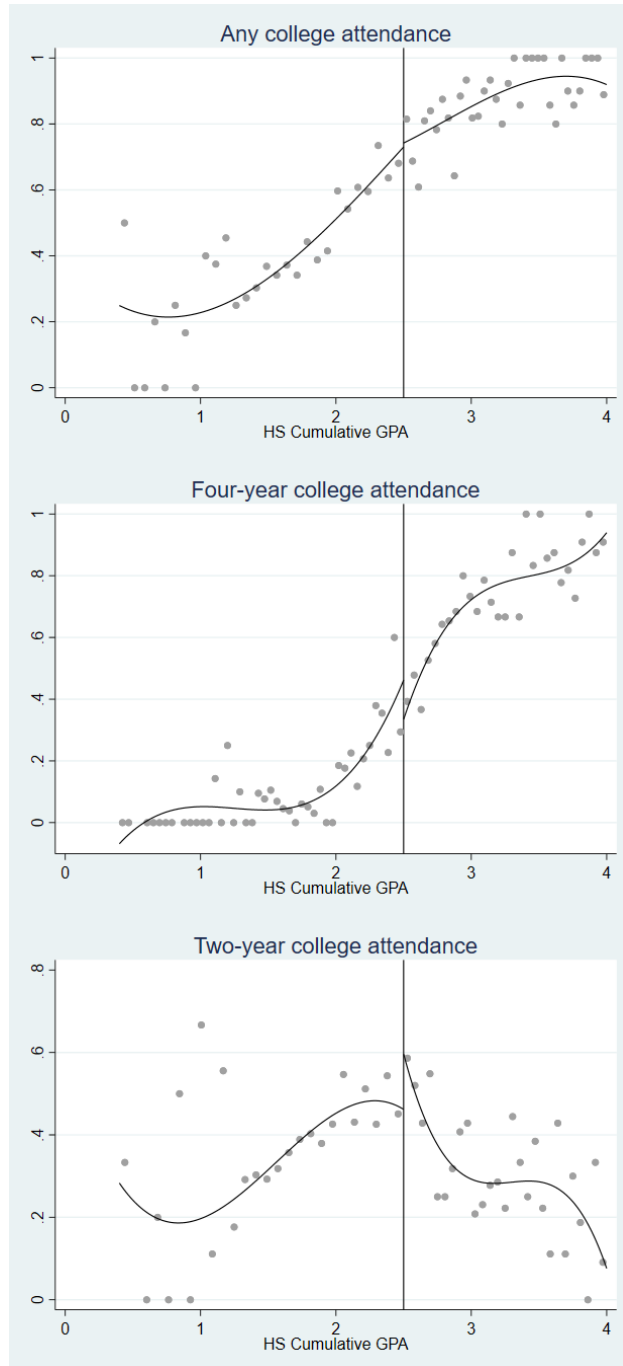
one of the letters. Students were also regularly updated on which schools they can graduate from in order to remain eligible.

Appendix C. Regression Discontinuity Analysis

Randomization provides evidence of average marginal effects. Also, Figure 2 provides evidence of the distribution of effects by GPA. Another way to estimate the effects of TDP on college outcomes is using regression discontinuity (RD), which yields the local average treatment effect (LATE) around the threshold. We do not view this as especially informative given that the local linear regression regressions already provide the distribution of LATEs, including that which arises near the threshold, but some readers may find the RD of interest and we report this below. (These are based on Model 4.)

Since we are examining college entry and almost no students, go to college who do not graduate. 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.

Figures C1: Test of Aid Targeting: Effect Heterogeneity by Baseline GPA on College Outcomes



Notes: See Quan and Harris (2020) for McCrary tests and optimal bandwidth estimates.

Appendix D. Testing for Targeting Effects of Merit Requirements

The tables below provide evidence regarding targeting effects of the merit requirements. Since we are interested in 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. We focus on 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.

The tables provide no evidence that such effects. All three show a mix of positive and negative effects. These are also imprecisely estimated.

Table D1: 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*
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 2A. 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 D2: 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 2A. 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 D3: Effect heterogeneity by baseline performance for initially qualified, long run outcomes

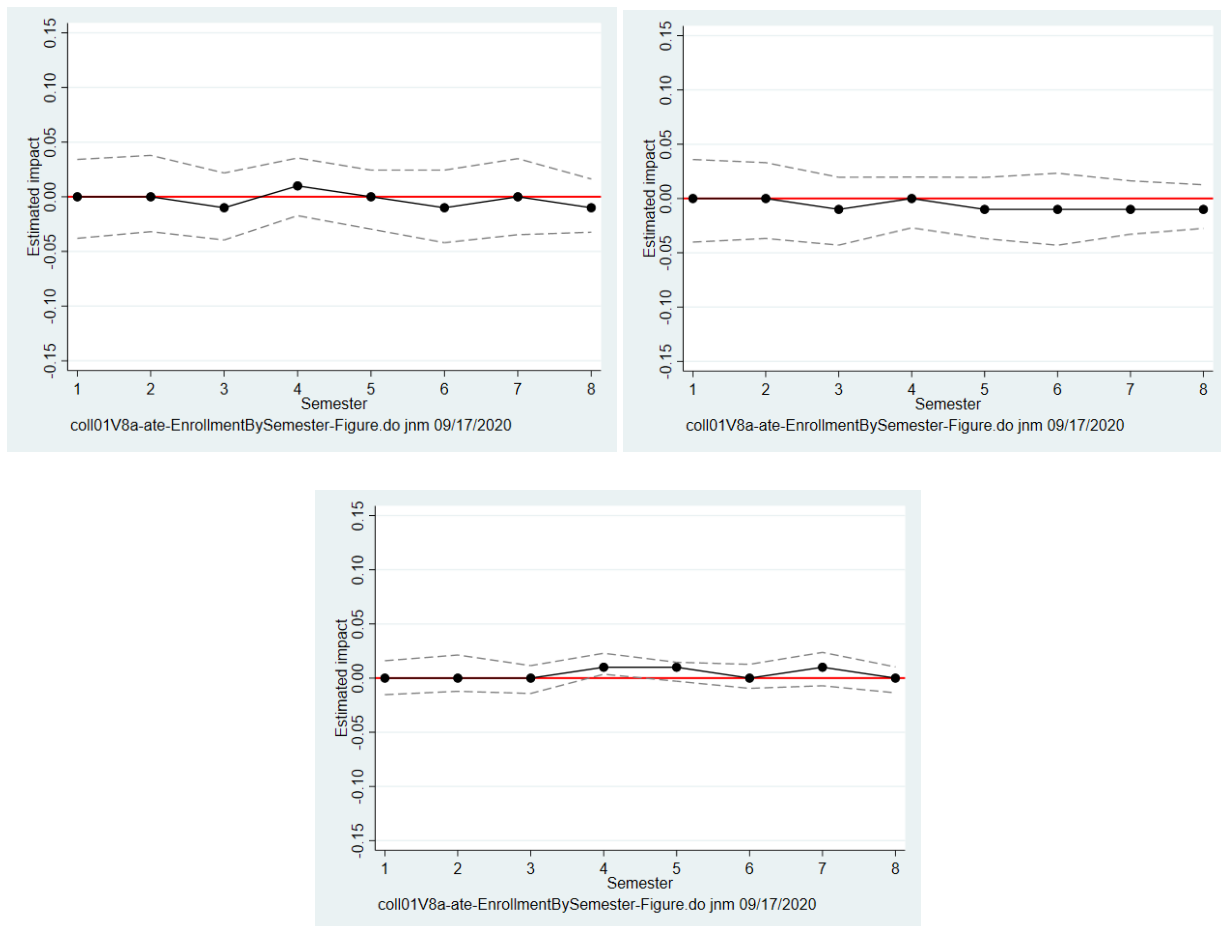
	TreatedXGPA	TreatedXMath	TreatedXReading
College graduation			
Ever graduated from college	0.010 (0.021)	0.018 (0.010)	0.007 (0.009)
Ever graduated from 2-year college	0.007 (0.017)	0.021 (0.009)	0.012 (0.007)
Ever graduated from 4-year college	0.003 (0.013)	-0.003 (0.006)	-0.005 (0.007)
Employment outcomes			
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 2A and Table 2D. 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

Appendix E. Additional Effect Heterogeneity

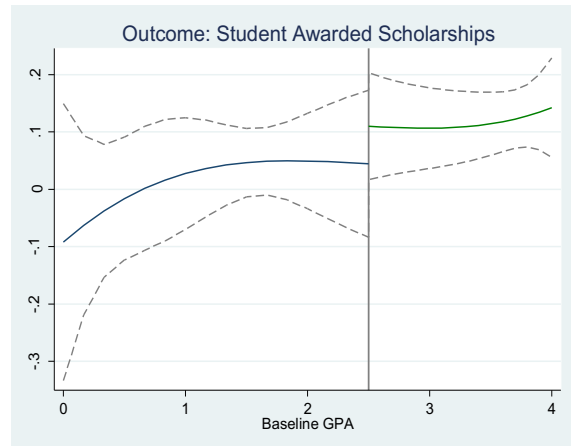
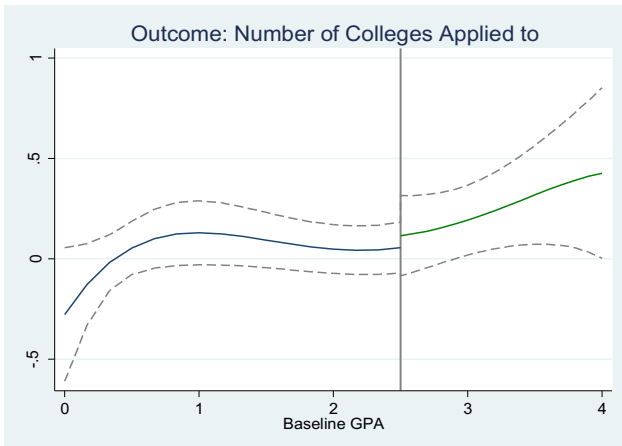
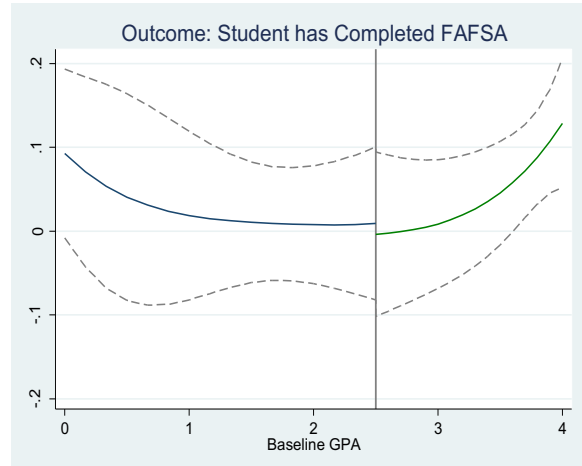
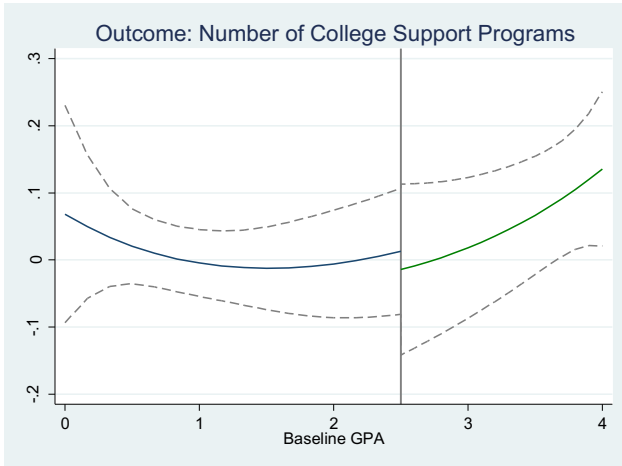
In this section, we provide evidence regarding effect heterogeneity. Figure E1 provides evidence about effects on college enrollment by semester. Figure E2 and E3 provide additional evidence regarding high school effort by academic endowment, extending Figures 4A and 4B in the main text to additional outcomes. Table E1 provides evidence regarding effects by level of communication. All results in this section are based on regression model 4 from the main text.

Figure E1. Enrollment Effects by Year, Post-College



Notes: The figure reports effects based on treatment assignment from Ordinary Least Squares (OLS) estimation of fully specified models with cubic pre-treatment controls. Solid black line represents 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.

*Figures E2: Effect Heterogeneity by Baseline GPA
Other Steps to College Taken During High School*



*Figures E3: Effect Heterogeneity, by Baseline GPA
College Expectations during High School*

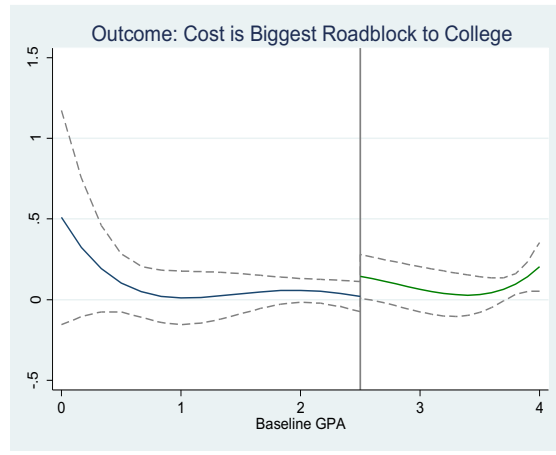
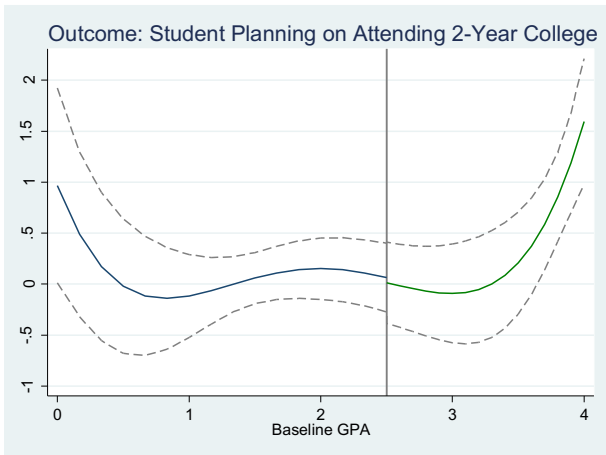
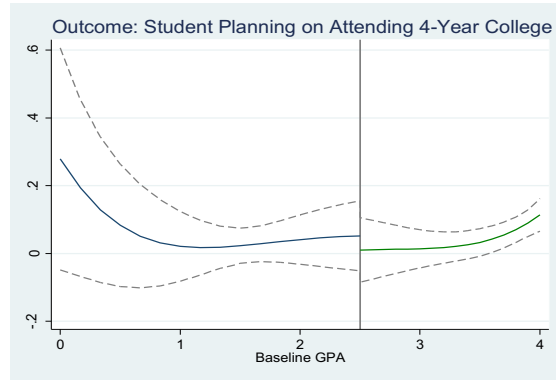
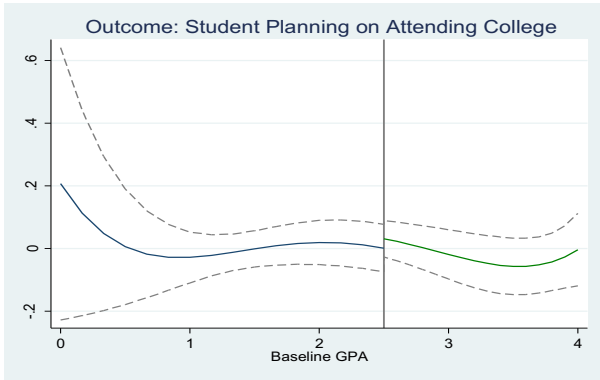


Table E1: Average treatment effect by number of communications received

	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 The Degree Project, and the number of times an adult in the school spoke with them about the project, which I summed across the two measures and four years (e.g., for student who stayed in school for four years and reporting receipt a letter and two mentions by a school counselor each year would yield $3 \times 4 = 12$ communications) in surveys conducted each year by MPS. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix F: Additional Details on Cost-Benefit Analysis

This section provides additional detail on the cost-benefit analysis. The main text provided a general formula. Here we elaborate:

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

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

where η is still the marginal cost of funds; g is still 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.

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, including 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.⁵

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 costs would over-state the true costs. However, other

⁵ 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 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. (2020) 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.

students are likely induced to start and do not finish, creating additional costs. The above calculations assume that these effects countervailing cost effects cancel out.

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), which combines 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 (across sectors).

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).⁶

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). In the base estimate, where only one of the two is significant, we count the effect of the imprecisely estimated

⁶ The variation in estimates, in both the two- and four-year sectors, does not appear to be due to differences in the discount rates.

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 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.

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 5 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.

In addition to the above assumptions, these calculations rest on two additional ones: (a) the program is large enough that the average cost is a reasonable approximation of the marginal cost; and (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. We will be able to test (b) in the future once we are able to observe longer-term effects on employment and earnings within TDP.

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.
- Kane, T.J. & Rouse, C.E. (1995). Labor-Market Returns to Two- and Four-Year College. *The American Economic Review* 85(3): 600-614
- 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.