ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-952

Should College Be "Free"? Evidence On Free College, Early Commitment, and Merit Aid From An Eight-Year Randomized Trial

Douglas N. Harris
Tulane University

Jonathan Mills
The Coleridge Initiative

We provide evidence about college financial aid from an eight-year randomized trial where high school ninth graders received a $12,000 merit-based grant offer. The program was designed to be free of tuition/fees at community colleges and substantially lower the cost of four-year colleges. During high school, it increased students' college expectations and low-cost effort, but not higher-cost effort, such as class attendance. The program likely increased two-year college graduation, perhaps because of the free college framing, but did not affect overall college entry, graduation, employment, incarceration, or teen pregnancy. Additional analysis helps explain these modest effects and variation in results across prior studies.

VERSION: April 2024

Suggested citation: Harris, Douglas N., and Jonathan Mills. (2024). Should College Be "Free"? Evidence On Free College, Early Commitment, and Merit Aid From An Eight-Year Randomized Trial. (EdWorkingPaper: 24-952). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/r4nr-hr77

SHOULD COLLEGE BE "FREE"? EVIDENCE ON FREE COLLEGE, EARLY COMMITMENT, AND MERIT AID FROM AN EIGHT-YEAR RANDOMIZED TRIAL

Douglas N. Harris
Tulane University

Jonathan Mills
The Coleridge Initiative

January 30, 2024

Abstract: We provide evidence about college financial aid from an eight-year randomized trial where high school ninth graders received a $12,000 merit-based grant offer. The program was designed to be free of tuition/fees at community colleges and substantially lower the cost of four-year colleges. During high school, it increased students' college expectations and low-cost effort, but not higher-cost effort, such as class attendance. The program likely increased two-year college graduation, perhaps because of the free college framing, but did not affect overall college entry, graduation, employment, incarceration, or teen pregnancy. Additional analysis helps explain these modest effects and variation in results across prior studies.

JEL Codes: I2, J24, D6

Corresponding Author: Douglas N. Harris, 6823 St. Charles Ave., 208 Tilton Hall, New Orleans, LA 70118. 504-862-8352, dharri5@tulane.edu.

Acknowledgements: *The Degree Project* (TDP) scholarship experiment was developed in a partnership between the lead 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. 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 many 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). Sticker prices have been rising for decades. Real student loan debt more than doubled from 1990 to 2004 (Avery & Turner, 2012) and has increased another 56 percent in real terms since then (Institute for College Access and Success, 2020). While the trends in *net* prices—the prices students actually pay—have been relatively flat over the past 15 years (College Board, 2021),[1] there is 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, is one of the nation's oldest programs and is, by far, the largest (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 students in particular localities are told early in their K-12 years that college will be free of tuition and fees (e.g., Carruthers & Fox, 2016; Swanson et al., 2017; Page et al., 2019; Bartik, Hershbein & Lachowska, 2021; Long, Goldhaber, & Gratz, 2021). Research on all of these and other programs has 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., forthcoming).[2]

Financial aid effects vary across programs, however, and this may be due to at least six dimensions along which programs vary: aid levels, administrative barriers, student eligibility

[1] The net price is lower than the sticker price because of government and institutional aid that reduces what students pay. These forms of aid have been growing particularly for low-income students (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 college affordability for low-income families is more complex.
[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 two-year college students (Anderson et al., 2019). A key reason for the absence of effects appears to be that the commitments occurred after students started college and therefore could not influence initial enrollment (Angrist et al., forthcoming).

requirements, college eligibility requirements, commitment timing, and design/framing as "free college." In this study, we provide theory and evidence about all six dimensions 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 randomly selected 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 every first-time ninth grader[3] in a randomly selected set of 18 high schools, half of those in the city. To receive the funds, 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 discussed later. These requirements are common in financial aid, especially state merit aid and promise scholarship programs (Swanson et al., 2017; Nguyen et al., 2019).

TDP was designed to make an early commitment of financial aid to students in ninth grade. This is potentially important because the largest aid programs, with their complex aid formulas, do not make concrete commitments until it may be too late to affect K-12 schooling behavior and preparation for college (Dynarski & Scott-Clayton, 2006). Committing aid in ninth grade (or earlier) could increase students' college expectations and improve college preparation before and during high school. The late commitment of traditional aid is also interconnected with student eligibility requirements. Eligibility is generally established near the time students enter college, so 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 during and after high school.

The program was also designed to minimize administrative barriers. The importance of these barriers is increasingly recognized with respect to complex FAFSA forms in higher education (e.g., Bettinger et al., 2009; Dynarski & Scott-Clayton, 2008) and other areas of public policy (Herd and Moynihan, 2018). Unlike some other financial aid programs, such as Buffet

[3] As in most urban school districts, many ninth graders 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.

Scholars (Angrist et al., 2021), TDP did not require that students sign up for the program, which broadened access of the program to more students.

The program was also designed to provide free community college to students, an idea that has recently become popular (Goldrick-Rab & Kendall, 2014). Partially rooted in the behavioral economics of financial aid policies (Dynarski & Scott-Clayton, 2006; Goldrick-Rab, Harris & Trostel, 2009), free college programs not only provide more generous aid levels, but also reduce the psychological burden of debt (Boatman, Evans, and Soliz, 2017), the complexity of college decisions (French & Oreopolous, 2017), and uncertainty about the price of college. At least 348 college promise and free college programs now exist across 47 states[4] and, reflecting this growing national interest, President Biden included free community college as part of his initial Build Back Better proposal.

After tracking students for eight years, we estimate that the TDP aid offer increased two-year degree completion by 0.5 percentage points (intent-to-treat). These estimates only reach conventional thresholds of statistical significance with the bootstrap standard errors and fall a bit short of those thresholds with randomization inference. We also find negative point effects, however, on four-year college outcomes, providing some suggestive evidence of substitution from four- to two-year colleges (the difference between the two- and four-year coefficients is very imprecise). The net effect on overall college attendance is therefore null. This is not due to inadequate statistical power as the standard errors are small enough to rule out the typical effects found in the prior literature.[5]

Our analysis of TDP adds to the financial aid literature in six 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, the other being the evaluation of the Buffet Scholars program (Angrist et al., 2021). Two other sets of RCTs have also examined the effects of financial aid on persistence, conditional on college entry, i.e., the evaluation of Opening Doors program (Barrow et al., 2014) and the Wisconsin Scholars program (Goldrick-Rab et al., 2016; Carlson et al., 2019; Anderson et al., 2020). The RCT evaluation of the Michigan HAIL program could also be included in this

[4] https://www.collegepromise.org/.

[5] This statement is based on the bootstrap standard errors. Prior research reviews have found average effects of 2-4 percentage points on college enrollment per $1,000 in grant aid. TDP aid amounted to $3,357 per recipient and the standard errors allow us to rule out enrollment effects as small as of 3.7 percentage points, or just one percentage point per $1,000.

list, though the intervention in that case provided information about existing aid, but not new aid (Dynarski et al., 2021). Others have commented on the prevalence of quasi-experimental research designs in the financial aid literature and the need for randomized designs and cleaner identification (e.g., Deming & Dynarski, 2009).

Our second contribution is that we have unusually rich data to explain why the effects were so small. We identify five main likely explanations: (1) the experimental design, which, while reducing bias, decreased positive spillover effects on other students and educators that seem to arise in scaled-up programs (Miron et al., 2008, 2009) due to its temporary and small-scale nature; (2) the scholarship amount was more modest than many other more effective aid programs; (3) the TDP merit requirements severely restricted the share of the treatment group that ultimately received any money; (4) the merit requirements pertained to high school behavior that established eligibility to receive aid in the future, in contrast to prior research suggesting that merit requirements are effective when they target college effort that prevents the loss of aid already granted (Scott-Clayton, 2011; Barrow et al., 2014); and (5) the program targeted a population that may face too many other, non-financial barriers to college access (Dynarski, et al., 2022).

Third, this is the first study to examine financial aid effects on a wide range of *high school* outcomes.[6] Exploratory analysis of these outcomes is important given the potential of early aid commitments to leverage human capital investments in the pre-college years. Our results show positive impacts on college expectations and low-cost, non-merit behaviors that might enhance college outcomes (e.g., participating in other college access programs). However, we find no effect on the high-cost, academic outcomes related to the merit requirements (GPA, attendance, high school graduation). We also find that the effects during high school are largest for students at the time of the TDP announcement (ninth grade) and in the senior year of high school, when aid is most salient. This apparent delay in student responses suggests that there are some limits on the effectiveness of early commitments of aid to increase academic and other forms of preparation during high school.

Fourth, this is one of only a handful of studies (Bettinger et al., 2019; Scott Clayton & Zafar, 2019; Carlson et al., 2019) to examine the effects of aid on life outcomes beyond education. Using data from six years after they first learned about TDP, when students were 20-

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

22 years old, we see no effects on incarceration or teen pregnancy. We also report imprecise null effects for employment.

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 possible substitution of four-year for two-year college may reflect a response to the free college design/framing in the two-year sector; (b) the positive effects on high school outcomes provide some, albeit limited, support for making aid commitments early; (c) the merit requirements did not have their intended targeting and incentive effects and therefore attenuated the average treatment effects; (d) the college eligibility requirements may have reduced the quality of colleges attended; and (e) implementation of the program varied widely, even within schools, indicating that implementation may matter as much as program design.

Sixth, we compare our results with other studies using comparative cost-benefit analysis to understand why the TDP effects were so small, whether TDP and other programs nevertheless improve social welfare, and how those welfare gains vary by program design. Prior studies have examined the social welfare implications of individual aid programs (Bartik et al., 2016; Angrist et al., 2021), but we are able to show that financial aid generally passes a cost-benefit test by a wide margin—TDP is the exception where the benefits and costs seem to largely cancel out. This analysis also suggests that open-access aid programs (i.e., those with few eligibility criteria) do more to increase social welfare. The net benefits per student are at least as large in open-access programs and these net benefits accrue to a larger number of students.

Section II describes *The Degree Project*, including the randomization process. The data for our high school, college, and life outcome analyses are discussed in section III. The results, including intent-to-treat, treatment-on-treated, and effect heterogeneity are in section IV. Section V uses theory and cross-study analysis to explain both the small size of the TDP effects and the variation in effects generally across programs. Section VI concludes.

II. 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 2,586 TDP promise 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 (i.e., there was no sign-up process).[7]

We used paired cluster randomization to select treatment schools (Imai, King, & Nall, 2009). Pairing on a lagged value of the dependent variable greatly increases 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. While this approach complicates statistical inference (Athey & Imbens, 2017), this problem is solvable with randomization inference (Bai et al., 2022) and this method provided statistical power well above average relative to the prior financial literature.

Specifically, we carried out the following steps to randomize treatment: (a) ranked schools by the college attendance rates from recent cohorts;[8] (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 lead author, using data from MPS.[9] (See Appendix C for additional detail on the data used in the paired randomization and related steps.)

On November 17, 2011, all of the 2,586 ninth graders 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 ninth graders in the non-selected schools who serve as the control group.

TDP treatment students received TDP funds so long as they graduated from any MPS high school[10] on time (within four years of starting ninth grade) with at least a 2.5 cumulative

---

[7] Only a handful of district-level employees knew about the program prior to the November 1 cut-off date.
[8] Six schools were new and did not have a prior cohort of students with which to create the matched pairs. In those cases, we imputed college-going rates using detailed demographic and academic outcome data and then matched within the set of schools that had missing data.
[9] Randomization was carried out using the random number generator in Stata (with set seed). Within each pair, the school with the larger number was assigned to the treatment condition.
[10] There were a few small MPS schools that did not report sufficient data to MPS to check eligibility requirements. Great Lakes regularly updated students regarding which schools they could attend to remain eligible.

GPA (C+/B-) and attended school 90 percent of the time.[11] 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 for students to stay within MPS.

In the (pre-treatment) graduating class of 2009 who were first-time ninth graders in 2005-06, 16.3 percent met these requirements and 65 percent of those students went directly on to college in fall 2009.[12] The average ninth grader in the TDP cohort had a 2.2 GPA and 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 did not have to 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 of the $12,000 each year if they attended full-time (≥ 12 credits) and one-quarter of the funds each year if they attended at least half-time. There were no GPA 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 to other eligible institutions.

To receive the funds, students also had to be first-time college enrollees, degree-seeking and have at least $1 of unmet need.[13] 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 grants

[11] 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.
[12] Author's calculations with assistance from MPS.
[13] Unmet need is the cost of attendance minus the expected family contribution and existing grant and scholarship aid (excluding loans and work study).

were "last dollar"[14] and covered up to the full cost of attendance.

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 explicitly that the $12,000 was enough to cover all tuition and fees for a two-year degree, i.e., free college. This is an especially positive framing for students whose families appear to be loan averse. 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 out of debt." This reinforces the potential value of the free college design/framing for these students.

Great Lakes, as the program administrator, sent two letters in the first year, and four letters per year in the other three years, to remind students about their eligibility and about whether they were meeting the requirements at each point in time (see a sample communication in the Appendix D). 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.

III. Empirical Framework

This section provides detail on data and research methods starting with a sample description and followed by discussion of baseline equivalence, estimation, inference, and treatment dosage. Most of our decisions on the empirical analysis were pre-specified in advance

[14] "Last dollar" means that unmet need is filled after other forms of eligible aid have already been applied to reduce that gap. If there is still a gap after that, called unmet need, then last dollar aid can be used, up to the point that the gap is eliminated, i.e., this aid is applied as a "last" step (https://www.acct.org/page/first-dollar-vs-last-dollar-promise-models). For example, for a student with unmet need of $2,000, then the student could use up to $2,000 in last dollar aid to cover that gap. One key implication of last dollar grants is that students, with their varying incomes, EFCs, and COAs, receive differing aid levels even when the face value of the grant (in this case, $12,000) is the same for everyone. Last dollar programs are also more complicated to administer because they require completion of the FAFSA to determine the EFC and other aid levels. (In some cases, though not with TDP, last dollar also means that the additional aid is sufficient to cover *all* direct costs remaining after other forms of aid have been accounted for, up to the COA, but that definition does not apply to TDP.) Last dollar aid contrasts with "first dollar" aid, which does not depend on other aid and therefore does not necessarily require the FAFSA. With first dollar aid, it is also sometimes possible to receive aid to cover living expenses that are not included in the COA, which is not possible with last dollar aid such as TDP.

of data collection (see Appendix J).[15] Some additional elements of the analysis were not pre-specified and we note these in the discussion that follows.

III.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 national 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 B for additional background on Milwaukee and its schools.)

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 briefly in Table 1. Appendix Table B1 provides full descriptive statistics.

III.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 completion with a GED or other credential).

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

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

The district administered an annual climate survey to all MPS students each spring and allowed us 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 2).

[15] The experiment began before pre-registration was a common practice. However, the analysis is laid out in a 25-page grant proposal to the USDOE Institute of Education Sciences prior to the experiment's launch. Appendix J provides a verbatim summary of the key methodological plans, including the choice of dependent variables and estimation methods.

MPS also administered a Senior Exit Survey MPS administers each year to measure students' college plans and steps they were taking to prepare, just before leaving high school. Like the climate survey, MPS also allowed us to add some questions to the Senior Exit Survey 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.[16]

III.A.3. College Enrollment Data

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

III.A.4. Life Outcomes: Employment, Incarceration, and Pregnancy

We measure life outcomes using near-census 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 UW-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. MPS sent student names, Social Security Numbers, and addresses to UW-IRP whose staff then manually searched for matches in these comprehensive state data files.

We code students as employed if they have any UW-MSPF 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 indicator for whether they were employed *or* enrolled in college since there may be substitution between the two and both are positive outcomes. While it might seem premature to measure employment,

[16] 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 necessary because students not planning to attend college could not answer questions about two- and four-year college or specific colleges, for example.

note that students had six potential years of potential employment during this panel (from ages 16-22) and that employment is the main alternative college for those who did not attend.

Individuals are identified as "ever incarcerated" if they are registered as incarcerated at any point in the panel. 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 January 2012 and July 2015). Note that incarceration rates generally peak between the ages of 16-24 and our analysis includes most of this age range.

III.A.5. Descriptive Statistics and Other Information

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. For comparison, 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.

Twenty of the 36 TDP schools are traditional public schools and the others are charter schools, which operate semi-autonomously from MPS and have smaller enrollments (see Appendix Table B2 for information about each school). Four of the 36 TDP sample schools are selective admissions as students have to apply and meet academic requirements (i.e., magnet schools). Partly because of these selective schools, the college-going rates varied greatly across high schools prior to TDP, ranging from 10 to 88 percent for on-time college enrollment.

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, 2020). However, these closed schools were quite small compared with most other schools, so this affected only 7 percent of the total sample and this number is almost exactly evenly split between control and treatment. Great Lakes also sent letters to treatment students and their parents to convey their continued eligibility for the program, as they did with students who transferred to another MPS school.[17] The students in these closed schools are still included in all of the analyses and we account for possible effects of

[17] 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.

this exogenous shock by adding a school closure/re-organization indicator as a covariate in some of the analyses that follow.

III.A.6. Missing Data and Attrition

In a randomized trial, the primary threat to identification is missing data/attrition. The outcomes we pre-specified as being of primary interest have minimal attrition. We discuss attrition with the high school, college, and life outcomes separately below. At the end of the section, we also discuss missing data on covariates, though this is less important since we can estimate unbiased effects without them.

The control and treatment groups include 5,050 students in total. For high school academic outcomes, we are missing data for 0-6 percent of the sample. Attrition in the high school data worsens somewhat 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 (e.g., we use 10^{th} grade GPA when students are missing in 11^{th} grade).

The largest attrition issue by far in the study is with the high school survey measures; the overall survey response rate was 53 percent, and we are missing data for 58-84 percent of specific items that are included in the analysis. The response rate was 3-5 percentage points higher for the treatment group. This missingness is due to a combination of: (a) high school dropouts; (b) transfers out of MPS; and (c) many of the high school exit survey questions are only asked of students who initially indicate they plan to attend college. Given these high missing data rates, and our pre-specification plan, we view the survey data is exploratory.

The college and life outcome data collection, as described above, is detached from TDP program administration and includes rich information to find individuals in the data, which keeps missing data to a minimum and ensures that the limited remaining missingness is balanced between control and treatment. MPS requested NSC and UW-MSPF data for all of the 5,050 originally selected students.

Prior research suggests that the NSC covers 96.5 percent of all college-going students in the state of Wisconsin (Dynarski et al., 2013). The implied missing data rate of 3.5 percent may be an understatement of the actual missing data for purposes, but NSC missing data is generally thought to be very low.[18] It is not possible, however, to identify specific missing observations or to calculate a specific missing data rate in this sample, given the method of constructing the data. Students who do not show up in the NSC as having attended college are simply coded as zero for college outcomes. Given the very low missing data rate overall (Dynarski et al., 2013), and the fact that there is no reason to expect differential attrition by treatment status, it seems unlikely that attrition biases the effect estimates for college outcomes.

The UW-MSPF data for the life outcomes data have properties similar to the NSC. For example, people only show up as employed in the state data when an employer reports them as employed in the unemployment insurance system; and those who do not show up as employed in the data are assumed to be not employed.[19] Given the extensive identifying information used in the data-matching process (see above), the missing data rate is thought to be very low. However, unlike the NSC (Dynarski et al., 2013), there is no way to estimate a missing data rate.

We considered calculating bounds of the effect estimates, at least for the survey results, but standard methods are infeasible or uninformative in this case. Lee bounds involve trimming the dependent variable to obtain similar distributions in the control and treatment groups, but this trimming is infeasible with dichotomous dependent variables. Manski bounds, in contrast, are feasible but would be so wide as to be uninformative. Instead, we rely on the limited differential in missingness between control and treatment groups and reiterate that the analysis of the survey-based outcomes is exploratory.

III.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) and the treatment students are actually more likely to be eligible for free or reduced-price lunches. The same

[18] The "coverage rate" reported by Dynarski et al. (2013) is the total enrollment of colleges included in the NSC divided by the enrollments of all colleges reporting to the U.S. Department of Education as of 2011. This is likely a lower bound on the missing data rate in our analysis because, while coverage has likely improved in recent years that are the subject of this analysis (through 2019), individual students within colleges covered in the NSC, and matching, is imperfect.

[19] A small share of employers, as well as self-employed workers, are not required to report to the unemployment insurance data system.

general pattern emerges, albeit more strongly, in the survey sample.[20] As shown later, covariates are sometimes included in the impact estimation to address these relatively small imbalances.

Appendix Figure C1 provides a visual representation of baseline equivalence, comparing schools within each pair on the pairing variable itself (college entry rate of prior cohorts). Since college outcomes are arguably the dependent variables of greatest interest, the fact that we have baseline equivalence on this measure is important. Appendix Figure C2 shows kernel density plots for attendance and GPA, as well as by control and treatment, showing that the distributions are also very similar between the two groups.[21]

III.C. Estimation

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

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 X_{ist} + \kappa_{ip} + \varepsilon_{ist} \tag{1}$$

where Y_{it} is outcome measure for student *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} where *p*=1, 2,…, 17 indexes pairs (recall that one pair had to be dropped). These pair indicators are necessary to account for the randomization design and obtain correct standard errors. We are primarily interested in average treatment effect parameter β_1. We estimate all the models using Ordinary Least Squares (OLS).

Given that there were some baseline differences between the control and treatment grounds, we include in some specifications a vector of student covariates X_{ist}. In some cases, these are non-linear functions of the baseline *individual*-level dependent variables, non-linear functions of the baseline *school*-level dependent variables, and other time-varying, individual- and school-level variables, such as student demographics.[22] The pre-specification in Appendix J

[20] We tested for baseline differences in the survey measures using survey data from the prior cohort of ninth graders in the same 36 schools (i.e., those students in ninth grade in fall, 2010). We found no baseline differences.
[21] 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 2). The joint test rejects the null. Pair indicators are included in the estimation but are not part of the null hypothesis in the joint test.
[22] 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 therefore include both eighth grade and ninth grade lags in Y_{it}; 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.

is nearly identical to equation (1). The X_{ist} vector also includes an indicator for whether an announcement was made to close or re-organize the school after treatment began.[23] As noted earlier, this affected 7 percent of the student sample, although the numbers were almost exactly evenly balanced between control and treatment.

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 imbalances noted earlier. Additional covariates can also increase statistical power.

III.D. Inference and Statistical Power

With estimation at the student level and randomization at the school level, one option for calculating the standard errors is the usual Generalized Estimating Equation (GEE) method with school-level clustering (Liang and Zeger, 1986). However, we avoid this because it requires asymptotic assumptions that do not hold in small samples, and the number of clusters in this case (36) is on the borderline of what is recommended by some researchers (Kezdi, 2004; Cameron, Gelbach, & Miller, 2008; Angrist & Pishke, 2009).

An alternative, increasingly preferred, option is randomization inference (RI), which involves re-randomizing the treatment condition and re-estimating the treatment effect in each iteration to obtain a distribution of potential effects. This yields a *p*-value that is the share of estimates that are at least as large as that from the actual randomization (e.g., Athey & Imbens, 2017). This is an exact test and is specifically recommended in the case of matched-pair randomization designs such as this (Bai et al. 2022).[24] We report the RI *p*-values in the main tables and the bootstrap standard errors in Appendix E ((MacKinnon & White, 1985; Cameron, Gelbach, & Miller, 2008). As we will show, these two sets of results are very similar.

Statistical power is very similar to some recent studies such as Buffet Scholars. While using a clustered design reduces power, this is almost made up by increased power from the pairing process. We can identify effects on two-year college entry, for example, as small as three

[23] We recognize that school closure occurs post-treatment and controlling for post-treatment factors can introduce bias, but this is a simple test for whether school closures affected the estimates. Our evidence suggests that school closure does not influence the findings, likely because the closures affected the control and treatment groups in nearly identical percentages.

[24] This is implemented with Stata's *ritest* command with the *strata(pair)* and *cluster(school)* using 1,000 replications.

percentage points, or less than one percentage point per $1,000 in aid, which is smaller than what is typically found in the literature (Deming & Dynarski, 2009; Nguyen et al., 2019).

The calculation of the standard errors is also potentially affected by the large number of dependent variables in the study, which increases the probability of individual Type I errors (i.e., of finding precisely estimated effects for some outcomes by chance alone). In the case of the usual confirmatory analyses, this situation normally calls for adjustments to standard errors to address multiple comparisons. In this case, however, we have relatively few outcomes of primary interest: GPA, attendance, high school graduation, and college-going. Almost all of the remaining analyses focus on the mechanisms behind these effects, which we pre-specified as exploratory and where multiple comparison adjustments are not typically applied.

III.E. Treatment Dosage

The size of the program effect depends on the contrast in treatment between the control and treatment groups with respect to dosage. The first aspect of dosage in this case is the level of awareness of the offer, which we analyze using student-level spring survey data. Two-thirds of the treatment student respondents correctly reported their treatment status and, of those, 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 from the program administrators. We also surveyed all students about whether planned to use some form of "scholarship funds" for college. As expected, the treatment group reported in the affirmative at higher rates, though the differential was only eight percentage points. 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.

To understand why the survey treatment contrast was so small, we next turn to the data on actual aid receipt and some general facts about the cost of attendance and grant aid that students typically received without TDP. In addition to posted tuition rates ($8,675), living expenses, fees and other costs raised the UW-Milwaukee cost of attendance to approximately $22,000 in the early 2010s when the TDP students were in high school. The average TDP control group student at UW-Milwaukee received $4,541 in total grant aid (mostly from Pell for those eligible for it), leaving roughly $17,500 in direct costs.

For two-year colleges, in addition to the $3,184 in tuition, living expenses for the typical two-year college student were $12,280 for those living on their own and $5,420 for the roughly one-third living at home (Anderson, 2020). Also, the average two-year college student, regardless of income, received grants of $6,700 per student (Anderson & Goldrick-Rab, 2018), so that the remaining costs were $15,464-$6,700=$8,764 (living on their own) or $8,604-$6,700=$1,904 (living at home). While these remaining costs are lower than for four-year colleges, they are still high enough that most Milwaukee students could use all the TDP funds.[25]

The above discussion focuses on the baseline aid of students in the TDP sample (and others like them). Next, we consider the TDP effects on college costs, i.e. dosage. Great Lakes reported that 312 students (12 percent of the treatment group), or about 57 percent of those who met the requirements, received any TDP funds. The average program spending among the 312 recipients was $8,809 (out of a maximum possible $12,000 across years). This figure was higher for four-year college students. Aggregated data from the UW-Milwaukee indicate that the average treated student received $11,155 in TDP funds (almost the full amount of the grant)[26] and saw loan reduction of $6,090 per student. This means the grant both increased students' cash flow and reduced the need to pay back funds, in roughly equal proportions.[27]

While we cannot carry out the same dosage calculations for two-year colleges, the negative correlation between income and two-year college attendance, and the reduced sticker prices in those colleges, leads us to expect that the equivalent average net price figure for community colleges would be much lower. Anderson (2020) reports that 54 percent of

---

[25] We note that these costs omit the expected family contribution (EFC), which increases with family income and reduces the grant aid that students are eligible for. For many Milwaukee students, the EFC is zero and, so, the above calculations provide a clear picture of their net costs. For students not living at home with zero EFC, the majority would be able to use the full TDP amount if they stayed in college through graduation to cover the remaining costs of $8,764-$17,500 per year. The groups that might not be able to use the full amount are those with positive EFC and/or those attending two-year colleges while living at home.

[26] The $8,809 figure is based on the total spending on grant aid by the program funder. UW-Milwaukee was the only college able to provide additional, and more detailed, data on aid packages and these were aggregated to the control and treatment group level. The UW-Milwaukee figures include 87 percent of the students who we observed as being enrolled at UW-Milwaukee in the NSC data (13 percent did not have aid packages on record). It also includes only the first three years post-high school; however, we expect that students received very little TDP funding in the fourth and later years. Indeed, the fact that they received $11,155 on average means there were almost no TDP funds remaining in later years even if students were still enrolled.

[27] This pattern of increased grants and reduced loans is expected, given students' loan aversion and the way that the financial system operates. A similar ratio of grant-to-loan reduction was found in another financial aid experiment in Wisconsin that involved many of the same colleges as in TDP (Goldrick-Rab, Harris, Kelchen, & Benson, 2016).

Wisconsin technical college students were Pell eligible; and, for the average Pell recipient, the Pell grant would have fully covered tuition.

We see three key takeaways from this discussion. First, the vast majority of students in the TDP sample, control and treatment, were eligible for substantial aid outside of TDP. Second, the complexity of the U.S. college financial aid system means that students might not have been very aware of these other aid programs or the real impact of TDP on their financial situations. Third, the vast majority of TDP students meeting the eligibility requirements could use all, or nearly all, of the TDP funds, though they might not have been very sure about this until after they started college. These facts are important to the subsequent interpretation of results.

IV. Results

Below, we present results for the various high school, college, and life outcomes using four econometric models based on equation (1). We begin with just the treatment and vector of pair indicators (Model 1), then add various combinations of individual- and/or school-level pre-treatment dependent variables (Models 2 and 3, respectively). Finally, we add a vector of student demographics and a school closure indicator (Model 4). All the models include a vector of missing indicators so that none of the observations are dropped due to missing covariate data. The results are generally robust across specifications and we have greater confidence in estimates for which this is the case (the Model 3 results are sometimes outliers). The *p*-values from randomization inference are noted in brackets. Standard errors using the stratified bootstrap can be found in Appendix E and generally yield similar inferences.

IV.A. Effects on High School Merit Metrics

Table 3 reports average treatment effects for the first-year academic measures that are included in the merit requirements. All but three of the 44 coefficients are precisely estimated null effects. We see little evidence that students responded to the performance incentives by increasing merit-related effort. The three academic merit requirements—GPA, attendance, or high school graduation—are unaffected. This might seem to contradict Scott-Clayton (2011) and Barrow et al. (2014) who both find positive incentive effects of merit requirements. However, there are several key differences between those studies and this one. Both prior studies focused on performance requirements for *maintaining existing aid* during college. Such incentives are more powerful because they leverage loss aversion in ways that TDP-like merit requirements, focused on the *prospect of future aid*, do not. Also, the other two studies find their largest effects on credit hours, an outcome that requires effort but, unlike high school GPA, almost no knowledge of the production function.

We do see some evidence that the scholarship increased a non-merit outcome, math scores. This effect attenuates somewhat when controlling for baseline differences (recall the baseline imbalance on test scores in Table 2).[28]

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

The effects of TDP are more positive when we turn to the survey measures of low-cost effort. Table 4 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.

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 costs in the communication from TDP program administrators may have raised more attention to costs and created concerns that students did not originally have.

In a 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). Taking this, together with Tables 3 and 4, it appears that students may have responded to the merit requirements in some ways that did not show up in GPA or test scores, perhaps due to limited knowledge of the production function (Fryer, 2010) or because those few efforts that did increase were low-cost and did not require sustained changes in effort or habits. Overall, we view the effects on high school behavior and perceptions as modest, which, as we will see, carry over to equally modest effects on college and life outcomes.

[28] Table 3 also shows some evidence of an increase in transfer to other MPS high schools. Students may have tried to move to high schools for which 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.

IV.C. Effects on College Entry and College Type

Table 5 shows precise null effects on overall college entry. The point estimates for two-year college attendance are only as high as one percentage point (for two-year college enrollment). There is some suggestive evidence of substitution from four-to-two-year colleges, though the differences between the coefficients is very imprecise. On the one hand, we would not expect this type of substitution because TDP provided identical nominal funds for either two- or four-year college, and because the above dosage analysis suggested that two-year college students were able to use a *smaller* share of the TDP funds compared with four-year college students. (In fact, students attending two-year colleges likely could not use all the funds, due to the lower cost of attendance.) On the other hand, the free two-year college framing might have made two-year college seem more attractive. While the analysis was not powered to test this hypothesis, it is worth noting the pattern because of policy interest in free college.

We also see an (imprecisely estimated) pattern that TDP induced students to shift from out-state/ineligible to in-state/eligible colleges and to colleges with lower graduation rates. This is consistent with prior research (Long, Goldhaber, & Gratz, 2021) and could entail reductions in college quality (Cohodes & Goodman, 2014).

IV.D. Effects on College Completion and Life Outcomes

Table 6 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 two-year college (*p*-value: 0.175).[29] The point estimates suggest that the program increased two-year college completion by a half percentage point, compared with a baseline mean of 1.2 percentage points. Note, too, that the point estimates for four-year college graduation are almost exactly zero.

The coefficients and bootstrap standard errors (Appendix E) allow us to rule out completion effects larger than one percentage point.[30] As we show later, even an estimate of this seemingly small size would still be insufficient to make TDP as cost-effective as the typical financial aid program. This reinforces that the analysis is sufficiently powered to conduct

[29] With bootstrap standard errors (Appendix Table E4), the estimates for two-year college completion are significant in Model 1, slightly short of significant in Model 2, and imprecise in Model 3.
[30] RI yields only p-values but since the significance levels are similar with the bootstrap standard errors, we identify the detectable effects from the bootstrap method.

informative policy analysis. (We note, too, that other older studies in this literature have not used randomization inference, so those studies are probably less well powered than they appear.)

The bottom of Table 6 also reports effects on employment, earnings, incarceration, and teen pregnancy observed up to four years after on-time high school graduation (roughly age 22).[31] The point estimates generally suggest positive, but very small, effects on employment and earnings, but these are predictably less precise than the other outcomes. Given the positive effects on college completion, and the short window of post-high school time we can observe, it may also be too soon for effects on labor market outcomes to emerge.

IV.E. Treatment on Treated (TOT)

In most RCTs, participants assigned to treatment often do not experience it in any meaningful sense. For example, when students are assigned to have additional college counseling, they might not be aware of this or might not show up for counselor meetings. In these cases, assignment to treatment is a valid instrument for estimating the complier local average treatment effect (LATE), i.e., treatment on treated (TOT) (Angrist, Imbens, & Rubin, 1996).

With TDP, however, *all* students assigned to treatment were sent communications about the scholarship throughout high school and the scholarship offer could itself affect student behavior and outcomes (e.g., raising college expectations and increasing FAFSA completion rates), potentially violating the exclusion restriction. (See the discussion of disappointment effects in the next section that offers a specific violation.) At the same time, the assumption may be plausible given the limited high school effects we saw in Tables 3 and 4.

Under the assumption that the exclusion restriction does hold, we carry out the usual TOT estimation and show the results in Appendix F. Since only 21 percent of the assigned students met the requirements, the treatment effect is, predictably, a bit less than five times larger for treated students relative to the ITT estimates in Table 6. The TOT effects on two-year college graduation, in particular, average roughly three percentage points.[32]

---

[31] While we cannot identify the age of pregnancy, we did limit to pregnancies occurring during high school when people are typically in their teenage years. We also see no effects on pregnancy in post-high school years either.

[32] 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. See Appendix F and Quan and Harris (2020). However, this method suffers from the same problem, and provides less information, compared with the IV-TOT. The potential high school effects from the offer that call into question the exclusion restriction in the IV-TOT also raise the

IV.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.[33] When a cluster is dropped for any combination of these reasons, the whole pair has to be dropped in this analytic 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.

Still, these are informative exploratory analyses. This is partly because, with our unusually rich data, we can test for and reduce the influence of the above factors. To ensure that the results are most comparable across subgroups, we identify and focus on the subset of schools where all of the subgroup estimates are feasible and re-estimate the average treatment effects to ensure comparability to the whole sample (see Tables 3-6) and show that the results are similar. We also conduct baseline tests for each subgroup analogous to those summarized in Table 2 (available upon request). In the discussion of each type of effect heterogeneity, all instances of baseline nonequivalence are noted in the text. Since there are naturally more baseline non-equivalence issues in the subgroup analyses, we focus on regression Model 4 with the full set of covariate adjustments.

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

We are interested in effects by GPA and test scores in part because they allow us to better understand the role of merit requirements and the incentive for effort during high school. To identify hypotheses about the effort response function, we developed a simple utility maximization problem in which students choose effort and whether to attend college (Appendix A). Under the most realistic assumption (risk aversion and uncertain returns to effort and risk aversion), we can expect the effort response function to be bell-shaped with the peak near the threshold. For example, the effects on attendance should be largest for students who are near a 2.5 GPA at baseline.

---

possibility of 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. Moreover, the RD estimate is just a LATE at a particular point: the performance threshold, while IV-TOT provides estimates across the GPA distribution, so we do not see the RD as very informative, but report them in Appendix F.

[33] For these reasons, when estimating models (3) and (4), we re-estimated the school averages for the covariates using only the values in the subgroup.

But we see only weak evidence of this pattern in our data. Appendix Figure H3 reports piecewise local linear regression estimates.[34] The only figure that looks remotely like the bell curve is the one with attendance as the outcome, which, along with the null average treatment effects, suggests limited evidence of incentive effects.

A second potential reason for using merit incentives is to target aid to students who respond most to it, which would suggest that treatment effects are increasing in the baseline GPA.[35] We see no evidence of this either. Appendix Figure 1B shows no clear pattern of effects across baseline GPA levels. These patterns may be an artifact of imprecision, but this still provides suggestive evidence that typical merit requirements do not effectively target aid.

We also test for targeting effects more directly in Appendix Tables G1-G3, 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., that merit requirements target aid to those who are less responsive to it.

From a behavioral economic standpoint, there is also a possibility that students with low academic endowments who were offered the funds might be disappointed by not meeting the requirements, which could actually reduce their outcomes. Indeed, we do see some evidence of this. Appendix G.2 describes a test that calculates the change in GPA between the freshmen and sophomore years for students who were initially barely eligible but who became ineligible at the end of the freshmen year, and compares this to otherwise identical control group students. In their sophomore years, the treatment group students who dropped out of eligibility saw larger drops in sophomore year GPA than the equivalent control students.

IV.F.2. Effects by Age/Grade

Effects by age/grade are important as evidence of the potential of early commitment to induce additional effort well before college. If students still wait until the end of high school to respond, then this would blunt the benefit of early commitment. Discounting is one reason to expect a small effect in early grades. When students learned of TDP in ninth grade, they might

[34] 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.

[35] Attendance was a requirement, in addition to GPA, 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.

have realized that they needed to increase effort, but that it would be optimal to delay that effort as far in the future as possible, but still before the determination of aid. This temptation toward delay is reinforced by behavioral economic concept of salience—effort directed at increasing aid for college is more salient near the time that students are making decisions about college. On the other hand, the scholarship was also especially salient at the time it was announced, in ninth grade. We might therefore also expect larger effects in both ninth and twelfth grade and smaller effects in tenth and eleventh grades. This would yield a U-shape pattern by age/grade.

To test this, Appendix Figure H2 shows GPA effects 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. Consistent with the theory outlined in Appendix A, we see some evidence that the effects are larger in twelfth grade, as well as perhaps ninth grade.

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

Prior researchers have discussed the role of communications in financial programs, including communication frequency and personalization (e.g., Bloom, Hill, & Riccio, 2003; Benhassine et al., 2013).[36] One prior experiment found that college students generally forgot about their scholarship eligibility (Angrist et al., 2010). Others have found effects from providing information about existing aid (Oreopoulos & Dunn, 2013; Dynarski et al., 2021). The source of communication may also matter; hearing something from an unknown entity, such as Great Lakes, is likely to generate a different response than a message from a specific person (Valant & Newark, 2020) and this might be especially true with messages delivered in-person by a trusted adult, such as a school counselor. TDP provided both types of communications to make the program implementation as realistic as possible.

Earlier we documented substantial variation in the number of communications students reported receiving. Appendix D provides evidence that the effects were larger for students who received more communications. This relationship is likely not causal, however, because qualitative evidence suggests that counselors deliberately communicated more with students who were succeeding in meeting the requirements (Rifelj & Kuttner, 2020). We carried out IV

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

analysis to address this with the average *school-level* number of communications (with leave-one-out) as an instrument for student-level communications. The exclusion restriction may not hold here either, but, even in that case, the positive correlation between communication and outcomes is still noteworthy as it underscores the mediating role played by high school counselors (Mulhern, 2019).

Treatment effect estimates by race, gender, and income can be found in the appendix, but we find no noteworthy patterns.

V. Explaining Variation Across Studies

This analysis raises a key question, why were the effects not larger in TDP, especially given the consistent finding that aid increases college outcomes (Deming & Dynarski, 2009; Nguyen et al., 2019)? In fact, in many respects, theory would lead us to expect that the TDP effects would be *larger* than those of the average program. TDP targeted students from low-income families who seem to be more responsive to aid (Dynarski, Page & Scott-Clayton, 2022); made early and specific commitments of aid ($12,000 in 9th grade) that gave students time to better prepare for college; implemented a multi-faceted communication effort, which resulted in broad awareness of individual eligibility and eligibility rules; and was designed to make two-year college free, reducing uncertainty and addressing loan aversion. On the other hand, the above analysis shows how poorly targeted merit requirements reduce the receipt of funds and possibly create disappointment effects.[37] Merit requirements also undermine the potential benefits of free college, which otherwise has the potential to reduce uncertainty, complexity, and loan aversion effects. A larger point, as the theoretical model in Appendix A shows more clearly, is that the effects of some aid design elements depend on the other design elements.

The above explanations regarding TDP also inform our understanding of why results vary so across programs and studies. Below, we show that both policy and research design are likely key explanations for this variation.

V.A. Empirical Comparisons Across Studies

We can help explain the small effects of TDP, and general variation across studies, by

[37] The predicted effect of merit aid on the TOT is different than the ITT. Even if the incentive effect is small, merit requirements could increase the TOT effect if merit aid targets funds to those who benefit most. Also, from a behavioral economics standpoint, merit requirements may create disappointment effects for some students who expected, but failed, to meet the requirements, thus reducing the ITT estimates.

comparing costs, effects, and program design elements across studies. The prior literature has generally expressed the effects of financial aid as degrees/$1,000 in aid or as aid per degree (Deming & Dynarski, 2009). We start with this approach in Table 7, which lists some of the most prominent, recent, well-identified studies. The middle set of columns shows the fiscal cost per student and the effects on two- and four-degrees, respectively. The effects per $1,000 calculation is based on the simple sum of the two- and four-year degree effects.

While the effects of TDP are toward the bottom and seem small, they are actually slightly larger than the Buffet Scholars program (Angrist et al., 2021).[38] This seems surprising at first because the effects of Buffet Scholars are generally considered to be "large." This seeming disjoint partially reflects the fact that the Buffet Scholars program provided much more generous funding, i.e., higher program costs. This result is also partly an artifact of counting two-year degrees in the TDP program the same as four-year degrees in Buffet Scholars, despite the latter's larger returns. This highlights a limitation of the degrees/$1,000 standard that we address in the cost-benefit analysis in the next section.

All but one of the programs we considered (Tennessee Knox), fall below the averages reported in prior reviews (Deming & Dynarski, 2009; Nguyen, et al., 2019). The table, and comparison of design elements, offers some potential explanations. In particular, many merit-based programs are not included in the prior reviews, yet most of the recent studies include merit requirements, reinforcing the possibility that merit requirements reduce, or at least do not improve, aid effects.[39] Another possible reason that the more recent effects seem smaller is that the proliferation of aid programs has blunted the impact of newer aid programs, i.e., diminishing returns. As noted in section Table 4, a majority of the TDP *control group* that planned to go to college during high school also planned to use *non-TDP* scholarships, which may have further reduced the effects due to diminishing returns.

We also see a pattern in the research designs—RCTs have smaller effects. This could be because of upward bias in quasi-experimental designs but also external validity concerns with RCTs. By their nature, RCTs involve temporary, small-scale programs where even seemingly

[38] We use the TOT estimate for TDP for comparability, as the costs only apply to those who received the funds, but recall that the exclusion restriction is a bit more tenuous with the TOT, even with treatment randomization (see section IV.E).

[39] Nguyen et al. (2019) conduct a meta-analysis in which they regress the estimated impacts on program and research design elements. This study shows no effect (precise zeros) for merit aid relative to other forms of aid. However, they do not distinguish different types of merit aid, which, as we have shown, may be important.

extensive communication like that in TDP is more limited than in scaled-up programs, which can be promoted on billboards and on television news. Also, small-scale RCTs limit the potential of positive spillover effects across students, e.g., the cost of each student's effort may be declining in the effort level of classmates. In addition, 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 that are only justified when they will apply to many future cohorts of students.[40]

Reinforcing this theory, 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 in qualitative studies from the larger TDP project (Rifelj & Kuttner, 2020; Kim & Rifelj, 2021). More generally, the effects of any program that would produce positive spillovers in scale-up will be underestimated in RCTs of small, temporary programs.

V.B. Comparative Cost-Benefit Analysis

Cost-benefit analysis is useful in this setting for several reasons. The first is that it allows us to weight the differing value of two- and four-year degrees based on their present discounted value of future earnings. This addresses the above concern by comparing TDP and Buffet Scholars on a degree/$1,000 basis.

The second reason is the more standard interest in social welfare effects. Harris (2013) and Bartik et al. (2016) have previously conducted cost-benefit analysis on these programs. As Angrist et al. (2021) point out, however, these approaches focus on fiscal costs and do not yield much information about social welfare because financial aid is mainly a transfer payment. To address this, Angrist et al. (2021) 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, it omits the marginal cost of funds (MCF), which arise because of the distortionary effects of taxation required to cover fiscal costs. Also, the use of COA over-states costs by including room and board, costs that arise even when

[40] The Buffet Scholars evaluation is also an RCT, which may partially explain why its effects are also smaller on a degrees/$1,000 basis. But the Buffet Scholars program is a different than TDP because the former is a scaled-up RCT that applies to all students in the state, every year, but we would still expect the sign-up plus randomization in this setting to reduce the spillover effects.

students do not attend college.

We address these concerns and compare the social welfare gains of different program designs. The cost of a two-year (four-year) degree are assumed to be $25,925 ($114,425) and that the present discounted value of future earnings from each degree are $150,000 ($436,350).[41] We also use a baseline discount rate of 3.5 percent and take the MCF to be 1.25 (Ballard, Shoven, and Whalley, 1985). These assumptions are explained and justified in Appendix I. We then apply this framework to TDP as well as other rigorously studied city/county-based programs and to the average effects found in two literature reviews (Deming and Dynarski, 2009; Nguyen et al., 2019).

The results are summarized in the last set of columns in Table 7. As in Harris (2013), all of the programs pass a cost-benefit test and, with the base assumptions, the benefit-cost ratios (BCRs) are in the range of 1.502-2.581, which suggests that financial aid consistently increases social welfare.[42] In the robustness checks, which vary the assumed parameter values, only two of the 21 BCRs are less than unity; these two exceptions are the TDP estimates with lower returns to education or a higher MCF. It seems clear that financial aid programs, even those with smaller effects than prior studies, improve social welfare.

The BCRs are remarkably similar across studies. With the exception of TDP, the range is 2.381-2.581. Buffet Scholars now looks much better than TDP because the PDV of the four-year degrees from Buffet Scholars is so much higher than two-year degrees. This raises the question, why did TDP perform so poorly, and Buffet Scholars so well, with respect to four-year degrees? Program design, again, is a likely explanation. Students had to apply even to be eligible for Buffet. As a result, and with high-level academic requirements, the Buffet Scholars sample had very high average control group GPAs (3.44) and college-going rates (90-96 percent). For this high-achieving, low/middle-income group, financial aid was likely to be the main impediment to four-year college.[43] The TDP control group had even lower average incomes and only a 2.2 GPA; and only 44 percent of high school graduates in the pre-treatment cohort attended college. So, even the marginal students in the Angrist et al. (2021) study have much stronger academic backgrounds than the average student in the TDP sample. Looking at Appendix Figure G2 we

[41] Bartik et al. (2016) find that the equivalent return to education figures are similar at $133,800 ($314,800).
[42] Bartik et al., (2016) also find that the Kalamazoo program passed a cost-benefit test.
[43] Two-thirds of the Buffet Scholars sample also had at least one parent who had attended college, suggesting that they also had considerable support at home.

can see evidence that TDP produced the same pattern as Buffet Scholars: the small number of TDP students with high baseline GPAs saw increases in four-year college-going. But the vast majority of the TDP students evidently faced so many other binding constraints that addressing only the financial constraint was insufficient. Interviews with TDP students reinforce that they faced many barriers to attending college, beyond the financial (Rifelj & Kuttner, 2020).

This comparison of TDP and Buffet Scholars also highlights a key point about the social welfare analysis: the BCRs reflect the welfare gains *per dollar*, but equally relevant is the *total* social welfare benefits, which depend on both the BCR and on how many students are eligible for the program. Looked at in this way, it seems likely that the Knox and Kalamazoo Promise programs would deliver the largest social welfare gains if taken to scale. They have high BCRs per recipient and are broadly available to students with few eligibility requirements. The Buffet Scholars approach of giving aid to low-income, academically high-achieving, self-selected students effectively targets aid, but this group is so small that it limits the total benefits.

This analysis is not without caveats, as Appendix I explains in more detail, but it also addresses some limitations of prior welfare analyses of college financial aid, by comparing across programs, accounting for the MCF and real marginal costs of college attendance, and distinguishing the social welfare implication per student from the overall social welfare effects. It also helps us see the importance of program design.

VI. Conclusion

With consistently rising student debt, and large gaps in college access by income, college affordability is a significant issue of modern economic policy. This study offers a large number of empirical and policy insights about financial aid as a potential solution. A rare RCT among studies of financial aid, we are able to identify causal effects on a wide range of high school, college, and life outcomes; use our rich data to understand how and why the effects were seemingly small; draw lessons from both within the analysis and comparison across studies to draw lessons about policy design and implementation; and examine the social welfare implications of TDP and similar programs.

We find that the effects of TDP were quite limited. The program had no measurable effects on the majority of the outcomes that it was designed to target: GPA, attendance, high school graduation, or college entry. In addition, we found no evidence of effects on employment, teen pregnancy, or incarceration in students' late teens or early 20s. The main exceptions are the effects from exploratory analyses for college expectations, low-cost high school behaviors (participating in college access programs and filling out the FAFSA), and perhaps two-year college completion.

Why were the effects so limited? First, the experimental design necessitated a focus on a single cohort of students, which decreased positive spillover effects on other students and educators that seem to arise in scaled-up programs (Miron et al., 2008, 2009). We looked for spillovers through our interviews and, not surprisingly, found no evidence they occurred. Second, the size of the scholarship was more modest than many other more effective aid programs. Table 7 shows that spending per recipient was less than half as in Buffet Scholars (Deming & Dynarski 2009; Nguyen, et al., 2019). Also, our surveys showed that the control group also expected to use scholarship funds, and the average level of grant funds among four-year TDP control students, for example, was $4,541. Third, the TDP merit requirements severely restricted the share of the treatment group that ultimately received any money, to just 12 percent of the overall sample; and the merit requirements may have created some disappointment among those who lost eligibility. Fourth, the program targeted a population that faces many other, non-financial barriers to college access. In short, the TDP design served a much smaller share of students, each of whom received less-than-usual additional aid and faced more-than-usual non-financial barriers, all while benefiting from fewer positive spillovers during high school.

Our analysis informs the optimal design and evaluation of financial aid. "Merit" requirements are often treated with a broad brush in the literature and our analysis helps distinguish the potential impact of different types. Common merit criteria pertain to initial eligibility based on high school outcomes, such as GPA and test scores. Improving these outcomes not only requires sustained increases in effort but also knowledge of the education production function. In contrast, requiring students to meet credit hour constraints *during college* to maintain aid eligibility requires minimal information about the production function and entail potential *losses* of existing aid, from which behavioral economics predicts larger effects. Merit criteria also tend to target funds to those who are already in a strong position to attend college, undermining equity.

As the paper's title highlights, this study also informs the potential of free college. Our results provide some, albeit limited, support for the arguments made by free college advocates. The suggestive evidence of substitution from four-year to two-year colleges in TDP may reflect a response to the free college design/framing that arose only in the two-year sector. On the other hand, the possibility that free college and other forms of early commitments might make students better prepared seems limited because students put off effort until the end of high school when it is often too late.

In addition to informing some of the individual elements of financial aid policy design, our analysis shows that these decisions are interconnected. The potential mechanisms that might make free college effective—reduced uncertainty about college prices and reduced decision complexity—are all undermined in the presence of merit-based criteria. Administrative barriers also increase with the number of eligibility criteria. Milwaukee students still had to fill out the FAFSA and TDP's program administrators had to go to great lengths to make sure students understood the criteria.

Financial aid does generally seem to improve social welfare and this reinforces its potential. While it is perhaps not the most efficient mechanism for improving college attainment when we consider alternatives, it has the advantage of being easily scaled and with minimal encroachment on student and college autonomy (Harris, 2013). We will learn more over the next decade as evidence emerges regarding the plethora of free college and other aid programs being adopted in states and cities throughout the country. This study, rooted in a rare, long-term randomized control trial, provides a framework within which to design and interpret these studies.

References

Anderson, D.M. (2020). When financial aid is scarce: The challenge of allocating college aid where it is needed most. *Journal of Public Economics* 190, 104253.

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.

Anderson, D.M. & Goldrick-Rab, S. (2018) Aid after enrollment: Impacts of a statewide grant program at public two-year colleges. *Economics of Education Review* 67: 148-157.

Angrist, J.D., Imbens, G., & Rubin, D.B. (1996) Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91(434): 444-455.

Angrist, J.D., 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.D. & Pischke, J-S. (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Angrist, J.D., 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.D., Autor, D., & Pallais, A. (2021) Marginal Effects of Merit Aid for Low-Income Students. *Quarterly Journal of Economics*.

Athey, S. & Imbens, G.W. (2017). The Econometrics of Randomized Experiments. In Abhijit V. Banerjee and Esther Duflo (Eds.), *Handbook of Field Experiments* (Volume 1, pp.75-135). Elsevier.

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.

Bai, Y., Romano, J.P. & Shaikh, A.M. (2022). Inference in experiments with matched pairs. *Journal of the American Statistical Association* 117(540): 1726-1737.

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., 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, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. (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.P. (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.P., 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.

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 (2013). *Trends in Student Aid*. New York: College Board.

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.

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.

Dynarksi, 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.

Dynarski, S, Libassi, C.J., Michelmore, K., & Owen, S. (2021). Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students. *American Economic Review*, 111 (6): 1721-56.

Dynarski, S., Page, L. & Scott-Clayton, J. (2022a). College costs, financial aid, and student decisions. NBER Working Paper 30275. National Bureau of Economic Research.

Dynarski, S., Nurshatayeva, A., Page, L. C., & Scott-Clayton, J. (2022). Addressing Non-Financial Barriers to College Access and Success: Evidence and Policy Implications (No. w30054). 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.

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.

Herd, P. & Moynihan, D. (2018) *Administrative Burden: Policymaking by Other Means*. New York: Russell Sage Foundation.

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.

Kim, Debbi & Kelly Rifelj (2021). Packaging the Promise: Money, Messaging, and Misalignment. *Teacher College Record* 123(6): 1-38.

Larsen, M. (2020). Does Closing Schools Close Doors? The Effect of High School Closures on Achievement and Attainment, *Economics of Education Review*, 76.

Lawrence, E. C. (1991). Poverty and the rate of time preference: Evidence from panel data. *Journal of Political Economy*, 99, 54–77.

Liang, K., and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* 73, 13-22.

Long, M., Goldhaber, D., & Gratz, T. (2021). Washington's College Bound Scholarship Program and its Effect on College Entry, Persistence, and Completion. *Education Finance and Policy* 16 (4): 690–715.

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 Pittsburg 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.

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), 95I-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.

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.

Table 1: Data and Variables

Outcomes	*Sources*	*Years*	*Attrition/ Missing Data Rates*
High School (Academic) *Variables tied to TDP Req*: GPA, attendance, grad *Other variables*: 10th grade test scores, transfer	MPS (admin)	2012-2016 (freshmen to one year post-on-time-HS-grad)	0-6% (see notes)
High School (Other) FAFSA completion, college access programs, college expectations	MPS (climate survey, senior exit survey)	2012-2015 (freshmen thru senior year)	53% overall response rate (control-treatment differential: 3-5 pp)
College College entry, persistence, graduation, institution/sector	NSC (through MPS)	2009, 2015-2019 (four years post-on-time-HS-grad)	~3.5% (Dynarski et al., 2013; see notes)
Life Employment, earnings, crime, pregnancy	UW-MSPF	2015-2017 (two years post-on-time-HS-grad)	Cannot be calculated

Notes: The years in Table 1 pertain to the spring year (e.g., 2012 means 2011-2012). As noted in the text, the missing data rates for the high school surveys range from 54-86 percent. This is not uncommon with student surveys; see the explanations in the text. The NSC missing data rate of 3.5 percent is one hundred minus the "coverage rate" reported by Dynarski et al. (2013) for the state of Wisconsin. No missing data rate can be calculated for life outcomes (see text for explanation).

Table 2: Baseline Equivalence

		Full Sample			Senior Exit Survey Sample		
	N	Control Mean	Treatment Diff	Diff w/ Pars	Control Mean	Treatment Diff	Diff w/ Pars
Panel A: Dependent Variables (at Baseline)							
Attendance	4989	0.886	0.003	0.010	0.951	0.001	0.006
			(0.027)	(0.008)		(0.010)	(0.003)
90% Threshold	4989	0.704	-0.001	0.019	0.886	-0.010	0.000
			(0.078)	(0.023)		(0.041)	(0.015)
GPA	3199	2.164	-0.034	0.060	2.573	-0.019	0.120**
			(0.242)	(0.041)		(0.227)	(0.036)
2.5 Threshold	3199	0.421	-0.017	0.027	0.587	-0.022	0.042*
			(0.106)	(0.021)		(0.107)	(0.018)
Math-MAP	4460	0.009	-0.016	0.093***	0.382	-0.100	0.057**
			(0.233)	(0.021)		(0.218)	(0.020)
Read-MAP	4401	0.029	-0.056	0.050	0.391	-0.160	-0.035
			(0.219)	(0.048)		(0.186)	(0.044)
Panel B: Independent Variables (at Baseline)							
Female	4995	0.501	-0.019	-0.024	0.577	-0.036	-0.044
			(0.037)	(0.021)		(0.037)	(0.028)
Age	4995	15.147	0.005	0.001	14.929	0.034	0.014
			(0.092)	(0.027)		(0.050)	(0.019)
Free/Red. Pr. Lunch	4995	0.803	0.055	0.036*	0.740	0.083	0.039*
			(0.053)	(0.013)		(0.066)	(0.018)
Special Education	4995	0.220	-0.018	-0.023	0.122	-0.010	-0.017
			(0.033)	(0.013)		(0.023)	(0.012)
ELL	4995	0.083	0.052	0.034	0.072	0.073	0.056*
			(0.051)	(0.035)		(0.042)	(0.025)
School Closed	5038	0.049	-0.013	-0.011	0.047	-0.027	-0.023
			(0.047)	(0.038)		(0.043)	(0.036)
English at Home	4995	0.897	-0.078	-0.065	0.892	-0.104*	-0.097**
			(0.053)	(0.039)		(0.050)	(0.032)
Black	4995	0.669	-0.087	-0.088	0.626	-0.075	-0.099
			(0.105)	(0.073)		(0.115)	(0.083)
Asian	4995	0.048	0.021	0.025	0.078	0.039	0.051
			(0.028)	(0.022)		(0.051)	(0.045)
Hispanic	4995	0.153	0.088	0.076	0.135	0.090	0.082
			(0.077)	(0.062)		(0.080)	(0.077)
Native American	4995	0.009	-0.001	-0.001	0.006	-0.001	-0.002
			(0.004)	(0.003)		(0.004)	(0.003)
White	4995	0.121	-0.022	-0.012	0.154	-0.053	-0.032
			(0.035)	(0.019)		(0.042)	(0.032)
College-Going Prob.	4818	0.297	-0.034	-0.022	0.423	-0.039	-0.006
			(0.079)	(0.013)		(0.087)	(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 3: Average Treatment Effects on High School Academic Outcomes (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Attendance pct	5033	0.813	0.004	-0.003	-0.004	-0.002
			[0.766]	[0.537]	[0.300]	[0.660]
90% or above	5033	0.469	0.002	-0.011	-0.010	-0.009
			[0.932]	[0.429]	[0.554]	[0.516]
GPA	4948	1.801	-0.006	-0.010	-0.009	0.000
			[0.901]	[0.825]	[0.868]	[0.997]
2.5 or above	4948	0.263	-0.009	-0.010	-0.008	-0.008
			[0.573]	[0.533]	[0.678]	[0.653]
Meets Both DP reqs.	4948	0.228	-0.007	-0.014	-0.006	-0.010
			[0.652]	[0.448]	[0.779]	[0.554]
Math MAP	4761	-0.057	0.085	0.039**	0.002	0.041
			[0.060]	[0.047]	[0.967]	[0.076]
Read MAP	4753	-0.021	0.026	0.012	-0.014	0.016
			[0.632]	[0.680]	[0.795]	[0.657]
Transferred schools	5038	0.407	0.057	0.058	0.108	0.063
			[0.324]	[0.276]	[0.122]	[0.222]
Missing in 2014-15 data	5038	0.261	0.001	0.001	-0.004	-0.002
			[0.963]	[0.978]	[0.925]	[0.921]
Grad On Time w/ Reg Diploma	5038	0.505	0.006	-0.001	-0.018	0.001
			[0.809]	[0.943]	[0.365]	[0.953]
Grad w/ Any Credential, Anytime	5038	0.550	0.019	0.013	-0.001	0.015
			[0.465]	[0.508]	[0.966]	[0.437]
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: The table reports average treatment effects based on equation (1) based on treatment assignment from Ordinary Least Squares (OLS) estimation. All models include randomization pair indicators. Each coefficient is from a separate regression. We report in brackets the *p*-values from randomization inference (RI) using 1,000 replications. (See bootstrap standard errors in Appendix E.) The last three columns estimate with covariates and, in these cases, we include a vector of missing indicators for all the covariates, so that the sample size is constant across the columns within each row even where there is missing covariate data. Given this, the N in the second column reflects missingness only on the dependent variable. "Baseline Performance (i)" reflects cubic specifications 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. There are 5,050 students in the TDP sample. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 4: 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	2146	0.680	0.041*	0.041*	0.023	0.042*
			[0.044]	[0.037]	[0.399]	[0.047]
Planning on 4-year college	2146	0.465	0.025	0.022	0.008	0.021
			[0.420]	[0.479]	[0.874]	[0.494]
Planning on 2-year college	2146	0.216	0.016	0.019	0.015	0.021
			[0.580]	[0.506]	[0.701]	[0.456]
Planning full-time college	2144	0.565	0.055	0.054	0.037	0.055
			[0.062]	[0.064]	[0.358]	[0.055]
Steps to College						
# college support prog participated in	2146	0.463	0.148*	0.149*	0.056	0.130
			[0.038]	[0.042]	[0.599]	[0.063]
# colleges applied to	2146	2.354	0.309*	0.304*	0.262	0.286*
			[0.036]	[0.033]	[0.213]	[0.044]
First choice college is highest prestige	1920	0.140	-0.029	-0.029	-0.027	-0.025
			[0.392]	[0.395]	[0.402]	[0.489]
Financial aid and college cost (senior year)						
Completed FAFSA	2129	0.519	0.044*	0.043*	0.029	0.044*
			[0.032]	[0.029]	[0.406]	[0.033]
Applied for scholarships	1461	0.596	0.079	0.074	0.046	0.070
			[0.090]	[0.101]	[0.456]	[0.136]
Awarded scholarships	808	0.393	0.066**	0.061*	0.077	0.059*
			[0.003]	[0.017]	[0.236]	[0.029]
Intends to use scholarships	1461	0.589	0.095*	0.086**	0.081	0.077*
			[0.010]	[0.005]	[0.065]	[0.014]
Biggest roadblock to college is cost	2135	0.557	-0.001	-0.002	-0.006	-0.007
			[0.952]	[0.928]	[0.889]	[0.731]
Summer Melt						
Planned on college, but did not enroll	2146	0.528	0.047	0.046	0.047	0.047
			[0.098]	[0.114]	[0.179]	[0.104]
Planned on 4-year college, but enrolled in 2-year college	2146	0.135	0.021	0.021	0.044	0.024
			[0.434]	[0.428]	[0.131]	[0.370]
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: The table reports average treatment effects based on equation (1) based on treatment assignment from Ordinary Least Squares (OLS) estimation. All models include randomization pair indicators. Each coefficient is from a separate regression. We report in brackets the *p*-values from randomization inference (RI) using 1,000 replications. The number of observations drops considerably relative to Table 3 because of the survey response rate (see text discussion). See additional notes to Table 3. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 5: Average Treatment Effects on Initial College Outcomes (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
Any College Enrollment	5038	0.334	0.004	-0.003	-0.015	-0.003
			[0.862]	[0.922]	[0.630]	[0.922]
2-year college	5038	0.198	0.012	0.007	0.001	0.007
			[0.406]	[0.512]	[0.952]	[0.512]
4-year college	5038	0.188	-0.004	-0.009	-0.017	-0.009
			[0.882]	[0.883]	[0.676]	[0.883]
Part-time status	5038	0.214	-0.002	-0.008	-0.020	-0.008
			[0.945]	[0.721]	[0.254]	[0.721]
Full-time Status	5038	0.224	0.009	0.003	-0.011	0.003
			[0.718]	[0.901]	[0.748]	[0.901]
In-state college	5038	0.290	0.008	0.001	-0.014	0.001
			[0.658]	[0.946]	[0.486]	[0.946]
Out-of-state college	5038	0.074	-0.012	-0.013	-0.013	-0.013
			[0.697]	[0.694]	[0.630]	[0.694]
TDP eligible college	1594	0.828	0.018	0.011	-0.032	0.011
			[0.674]	[0.793]	[0.492]	[0.793]
Coll. Competitiveness	935	0.929	0.015	0.004	-0.025	0.004
			[0.812]	[0.949]	[0.817]	[0.949]
Coll. grad. Rate	1566	0.440	-0.011	-0.014	-0.010	-0.014
			[0.664]	[0.713]	[0.608]	[0.713]
Undermatched college entry	803	0.079	0.017	0.024	-0.001	0.024
			[0.630]	[0.690]	[0.977]	[0.690]
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: The table reports average treatment effects based on equation (1) based on treatment assignment from Ordinary Least Squares (OLS) estimation. All models include randomization pair indicators. Each coefficient is from a separate regression. We report in brackets the *p*-values from randomization inference (RI) using 1,000 replications. The college competitiveness and college graduation rate measures are institution-level and only available for students who attended a four-year college, which reduces the number of observations for those variables. The number of observations for undermatched college entry is small because this requires the college expectations data from the surveys, which have some non-response. See additional notes to Table 3. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 6: Average Treatment Effects on College Graduation and Life Outcomes (ITT)

	N	Control Mean	Model 1	Model 2	Model 3	Model 4
College graduation						
Ever grad. from college	5037	0.020	0.005	0.004	0.003	0.004
			[0.249]	[0.304]	[0.622]	[0.384]
Ever grad. from 2-year college	5038	0.012	0.005	0.005	0.003	0.005
			[0.174]	[0.175]	[0.279]	[0.175]
Ever grad. from 4-year college	5038	0.007	0.000	-0.001	0.000	-0.001
			[0.939]	[0.863]	[0.970]	[0.863]
Employment outcomes						
Ever employed	5037	0.602	0.011	0.007	0.016	0.005
			[0.622]	[0.749]	[0.553]	[0.822]
Ever employed or college	5037	0.719	0.007	0.001	0.004	0.001
			[0.608]	[0.956]	[0.855]	[0.951]
Employed or college in 2017	5037	0.639	0.005	-0.001	0.003	-0.002
			[0.700]	[0.953]	[0.894]	[0.877]
Earnings						
Full sample	5037	5.061	0.141	0.095	0.167	0.066
			[0.625]	[0.702]	[0.466]	[0.866]
Excluding 2017 college enrollees	4342	5.122	0.112	0.075	0.164	0.058
			[0.486]	[0.629]	[0.450]	[0.727]
Ever incarcerated	5037	0.033	0.005	0.007	0.014	0.008
			[0.414]	[0.332]	[0.092]	[0.192]
Ever had child during High School	5037	0.018	-0.002	-0.002	0.003	-0.003
			[0.752]	[0.715]	[0.775]	[0.625]
Lagged Dependent Variable (i)				X		X
Lagged Dependent Variable (j)					X	
Student Covariates (i)						X

Notes: The table reports average treatment effects based on equation (1) based on treatment assignment from Ordinary Least Squares (OLS) estimation. All models include randomization pair indicators. Each coefficient is from a separate regression. We report in brackets the p-values from randomization inference (RI) using 1,000 replications. See additional notes to Table 3. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 7: Benefit-Cost Analysis Results

Program	Study/Program Design	Fiscal Cost/ Student	Effects on 2y degree	Effects on 4y degree	Effect/ $1,000	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	0.74	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*	1.47	2.381	1.427	2.141	2.381
TN-Knox (Carruthers & Fox, 2016)	DD; No Merit; Free 2y; Last $	$971	4.00*	-1.00	3.09	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*	1.19	2.399	1.420	2.219	2.700
Buffet Scholars (Angrist et al., forthcoming)	RCT; Merit Req; Covers 2y/4y; Nearly Free; Last $	$8,200	-3.00	8.40*	0.66	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 estimates 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.