Testing the Promise: A Randomized Trial of a Promise College Scholarship for Urban Public School Students

Douglas N. Harris, Tulane University[1]
Larry L. Orr, Johns Hopkins University[2]

Promise programs seek to increase college access by assuring students early in high school that they will have the financial resources to pay for college. Given dramatic tuition increases recent years, college increasingly seems out of reach for low-income students—and the reality is that few of them achieve college success. By increasing the real and perceived affordability of college, and clearly communicating the path to college, the theory is that these "early commitment" programs improve academic preparation and social capital during high school and thereby increase college entry and completion. Promise programs are widely used and feasible alternatives to traditional "late commitment" grant and loan programs that wait until students are leaving high school—when many are already off track.

The Degree Project is the first U.S.-based randomized trial of a promise program, one that commits college funds to low-income ninth graders who meet specified requirements. This paper describes this test of the effectiveness of the promise program approach. As we detail below, prior research is supportive but mixed. The project described here will provide the first rigorous test of the considerable potential of promise programs to help low-income students enter and succeed in college.

The Policy Problem

Economic success in the United States has long been built on the education, skill, and ingenuity of its workers. But over the last two decades, the world changed. The U.S. no longer has an advantage over competitor countries, as the supply of skilled workers has failed to keep up with the demand (Goldin & Katz, 2008). One reason is that high school students do not see a clear path to college and beyond. In particular, dramatic increases in tuition make college seem out of reach. Direct college costs have quickly outpaced inflation and surpassed increases in state and federal government support to colleges and universities. Federal financial aid to students has risen, but is less helpful than it appears because it increasingly takes the form of loans rather than grants, making college investment a riskier proposition for students and their families. The result, as Goldin and Katz (2008) put it, is that "the combination of the high cost of college, credit market constraints, and student debt aversion leaves many youth from poorer and middle-income families behind in the pursuit of a college education" (p. 349).

College now seems unaffordable and leads many low-income and disadvantaged public school students to inadequately prepare during high school for the academic, social, and financial demands of college. Eighth graders from families in the highest income quartile are nearly 10 times as likely as those from families in the lowest income quartile to receive bachelor's degrees

[1] Douglas N. Harris is Associate Professor of economics and University Endowed Chair in Public Education at Tulane University
[2] Larry L. Orr is Associate in the Department of Health Policy and Management, Institute for Policy Studies, Johns Hopkins University.

(Kirst, 2004; Bailey & Dynarski, 2011). Only 65% of minority students graduate from high school (Heckman & LaFontaine, 2007) and, of those, only a little over half go right on to some type of college (Aud et al., 2011). Among ninth graders in the Milwaukee Public Schools—a highly disadvantaged group and the sample for this study—only 67.8% complete high school on time, and only 44.4% of those high school graduates directly transition to college.

Helping students view college as a viable option has the potential to increase both high school and college success—outcomes that must all be improved if the U.S. is to retake its position as world leader in human capital. Yet, the challenge among urban students is enormous. As researchers with the Consortium on Chicago School Research wrote, "The primary issue in college access is no longer building college aspirations, but building a clear path for students to achieve their goals" (Nagaoka, Roderick, & Coca, 2008, p. 1). Scholars have done a masterful job of documenting the difficulties—students who are motivated but directionless, taking easy courses in high school, attending class and studying rarely, conducting consequential but ill-informed college searches, enrolling in colleges for which they are mismatched academically, and floundering through higher education propelled by a compelling but underspecified goal of getting a "college degree" (Bowen, Chingos, & McPherson, 2009; Roderick, Nagaoka, Coca, & Moeller, 2009; Rosenbaum, 2001; Schneider & Stevenson, 1999). But interventions such as the federal TRIO programs have generally failed to produce large impacts, especially relative to their large costs (Bailey, 2011; Chingos, 2011; Harris & Goldrick-Rab, 2010). This has led many to conclude that the end of the teenage years is simply too late to intervene effectively (Cabrera & La Nasa, 2001; Heller, 2006; Long, 2008; St. John, 2001).

Financial aid is one of the key interventions that comes too late (Dynarski & Scott-Clayton, 2006; Long & Riley, 2007). The average person believes that the costs of college are nearly double the actual costs, and triple the actual costs for 4-year colleges (Ikenberry & Hartle, 1998). The over-estimation of costs is even worse in the population of interest here—African-Americans and people with low incomes (Ikenberry & Hartle, 1998). These groups not only face more *actual* barriers to college, but *perceive* even higher barriers. Financial aid programs could provide a useful corrective to these misperceptions if the commitments and cost information came earlier. We need to persuade middle and high schoolers early on that college is affordable—and making affordability a reality by offering a substantial financial commitment. We also have to educate students about the concrete steps they must take to achieve college success. In short, education leaders and policymakers need to build a road to college—and promise some gas money—with which students can "drive" their strong college aspirations.

Promise programs could do just that. With over 73 promise programs nationally, they are clearly replicable and scalable. The basic element of promise programs—the *early* commitment of aid—could be extended to traditional programs simply by having students apply for aid much earlier in their school careers and setting aside a specific amount well in advance of college. *The Degree Project* (TDP) scholarship experiment will be the first rigorous test of the potential of this simple, yet potentially powerful, idea.

The TDP Intervention

On November 17, 2011, the TDP program promised students attending half of Milwaukee's 36 public ninth-grade schools a total of $12,000 each to pay for college. The 18 TDP schools

were randomly selected by the researchers on behalf of the program funder, the nonprofit Great Lakes Higher Education Corporation. The remaining 18 schools serve as the control group. All ninth graders (one cohort) in each TDP treatment school were given the promise. There was no explicit income requirement in order to be selected for TDP, but the low income level of most MPS students makes the program implicitly need-based. To receive the money, students will have to meet various high school requirements and then attend college, as discussed below.

Students will receive the TDP funds (through their aid package) so long as they graduate from any Milwaukee Public Schools (MPS) high school on time (within 4 years of starting ninth grade), complete a Free Application for Federal Student Aid (FAFSA) senior year and each year of college, and attend an eligible college at least half-time. The MPS specific course work requirements for high school graduation will apply (MPS, n.d., p. 2)[3] and General Educational Development (GED) certificates do not qualify. In addition, TDP scholarships will require students to graduate with at least a 2.5 cumulative grade point average (GPA), equivalent to a C+/B-, and attend class 90% of the time. The GPA and attendance requirements are cumulative across years, so that students who fall behind can catch up.

The TDP scholarship must be used within 4 years of expected high school graduation—specifically, by the spring of 2019. Students will be able to spend up to half the total scholarship per year if they attend full-time ($\geq$ 12 credits) and half this amount if they attend at least half-time (but less than full-time). (Students cannot obtain any type of degree in one year; therefore requiring that the funds be spread out over at least two years may facilitate degree completion.) Students need not start college immediately, but must start within 15 months of on-time high school graduation. For example, students who do not attend college at all in the first year after high school graduation will still have the full scholarship amount to spend, but they would still have to spend the money by spring, 2019. There are no GPA requirements during college.

To receive the funds upon reaching college, students must be first-time enrollees, degree-seeking and have at least $1 of unmet need.[4] (Therefore, while income does not affect initial eligibility it will affect how much funding TDP students receive.) Funds will be disbursed to financial aid offices by the Wisconsin Higher Educational Aids Board (HEAB) following the same process used to disburse state grant aid. HEAB's involvement means that the college must be a nonprofit 2- or 4-year institution in the state of Wisconsin—that is, a college in the University of Wisconsin (UW) System, the Wisconsin Technical College System, the Wisconsin Association of Independent Colleges and Universities, or the tribal college system. There are 64 colleges and universities meeting these criteria.

The TDP scholarship is "last dollar" and will cover up to the cost of attendance. Many of the TDP students will have a zero expected family contribution; for these students, the total TDP scholarship, combined with other forms of aid, will cover the entire cost of attendance for more than 2 years at a public 2-year college. Looking at the full-time tuition and fees of the 2- and 4-year institutions most commonly attended by MPS students—$3,184 annually at Milwaukee

[3] Under the district's standards, MPS ninth graders have to complete 4.0 units of English/language arts and 3.0 units each of mathematics (only courses that include or go beyond Algebra I; remedial courses do not count toward this total), science, and social studies, among other requirements involving physical education, service learning, and standardized test scores.
[4] Unmet need is the cost of attendance minus the expected family contribution and existing grant and scholarship aid (excluding loans and work study).

Area Technical College (MATC) and $8,675 annually at the University of Wisconsin–Milwaukee (UW–Milwaukee)—we see that TDP by itself would cover all tuition and fees for a two-year degree at MATC and more than one full year at UW–Milwaukee. (Half of MPS students who go on to college attend one of these two institutions.) Although tuition and fees are likely to rise before 2015, when TDP recipients first enter college, the point is that $12,000 constitutes a substantial reduction in the direct costs of college, and perhaps more important, will likely seem a large amount of money to a ninth grader. As we discuss below in Promise Programs: The Evidence, a prior randomized controlled trial and several quasi-experiments with similar programs suggest that this amount should be sufficient to observe effects of TDP.

Initial Implementation

The program was implemented almost exactly as planned. Two weeks prior to the November 17 announcement, Great Lakes and the researchers met with most school principals of the 36 TDP schools. The researchers presented the program, explained that half the schools would be selected and that, if selected, the program would apply to only one cohort of ninth graders. The principals greeted the program enthusiastically.

Letters announcing the scholarship offer, addressed to each student by name, were hand-delivered to each school in a sealed envelope two days prior to announcement day. Schools were directed to distribute the envelopes containing the $12,000 scholarship offer and hold an assembly on November 17. If any of the selected students were no longer attending the schools, principals were directed to return those letters to Great Lakes. Only 84 letters out of more than 5,000 were returned and Great Lakes worked with MPS to see whether students had switched to other MPS schools and in those cases delivered the letter to the new school. The day before announcement day, a copy of the award letter was sent home to parents by regular U.S. Mail (translated as appropriate).

At one of the announcement day assemblies—more like a pep rally--attended by the lead researcher, the school principal announced the $12,000 offer in an auditorium decorated with college pennants. Students were handed their personalized award letters and the principal explained the program.

Additional Communications

Information is crucial to the success of almost any program, and especially ones that involve financial incentives with eligibility requirements. As the initial implementation highlights, the program funder (under the advice of the researchers) is carrying out an aggressive communication plan. Prior research suggests that students are ill-informed about the steps they have to take to be successful in college, especially about costs and financial aid (Bowen et al., 2009; Roderick et al., 2009; Rosenbaum, 2001; Schneider & Stevenson, 1999). Even when they are already receiving aid, students often forget about the opportunities available to them (e.g., Fowler, et al., 2009).

Therefore three months after the initial announcement, Great Lakes sent individualized reminder letters that indicated whether each student was “on-track” to meet the requirements.

These on-track letters will be sent approximately three times per year during high school. As with the initial award letters, all on-track letters and subsequent communication will be sent twice, one to school and one home (by U.S. Mail).

These letters will also include information about typical high school course work of successful college students, average college costs and financial aid amounts in Wisconsin, names of colleges recently attended by MPS students, and the process for—and importance of—signing up for the ACT exam.[5] Because one requirement is filling out the FAFSA, the program will provide FAFSA information to students multiple times as they begin their senior year. Research shows that FAFSA completion is a significant impediment to college entry (Bettinger et al., 2009).

Parents will also receive updates through the district's Parent Assist web site that allows each parent to log in and see the academic progress of their children. A growing percentage of students and parents provide their cell phone numbers to the district and these will be used to send text message updates about TDP. Finally, the program has a public web site (www.degreeproject.com) and a telephone hotline to address questions from students, school leaders, community members, and others.

After the program was launched, program staff at Great Lakes participated in a monthly meeting of district counselors to explain the program and provide lists with the names of all TDP recipients in each of their respective schools. Counselors were shown data regarding student awareness of the program and district staff emphasized to counselors the importance of the program to the district's larger college-going focus.

The objective of this extensive communication plan is to keep students informed and aware—in a way that is scalable. We have been in close contact with the administrators of promise programs in Kalamazoo, Pittsburgh, El Dorado (Arkansas), and others. Because most students are eligible in those scaled up versions, those programs are able to carry out more "blanket communications" than we are. To compensate, the discussion above shows that we are using more targeted communications. We believe the overall level of awareness of the TDP will be similar to these scaled up programs.

Comparison with Similar Programs

We are aware of 73 promise programs in the U.S. At least 15 of these programs have some sort of merit or performance requirement, based on GPA (ranging from 2.0 to 3.5), class attendance, SAT/ACT scores, and/or class rank/percentage. In terms of target population, requirements, and scholarship amounts, TDP is most similar to the Pittsburgh Promise as well as New Haven, which have both GPA and attendance requirements. Almost all statewide promise programs have a GPA requirement, though only some of these are targeted to low-income students (Indiana, Oklahoma, and Washington). Programs in Florida and Georgia are similar but the vast majority of funds go to middle- and high-income students (Dynarski, 2000).

[5] Nearly all U.S. colleges require that entering students take either the ACT or SAT exam. Colleges in Wisconsin generally require the ACT.

Promise programs are already in wide use and involve billions of dollars each year. Yet they have received little rigorous evaluation. Cities across the country are considering adding programs of their own, including Akron, OH; Hartford, CT; Janesville, WI; Jacksonville, FL, and a statewide program in Michigan. The TDP program and evaluation will therefore inform large existing and prospective investments. Perhaps more importantly, the project will inform—and perhaps improve—the $177 billion annual investment the nation currently makes in traditional financial aid.

Research Questions

The randomized control trial (RCT) is generally considered the most rigorous method for identifying internally valid estimates of average treatment effects. The potential of RCTs to learn about how and why programs work, however, is often under-estimated. In a recent edited volume about RCTs, lead investigator Harris defines the "comprehensive RCT" as one designed to understand not only average treatment effects, but the mediators and moderators of those effects. He argues that the additional cost and effort that go into a comprehensive RCT is warranted under certain conditions, including: (a) considerable resources are already being devoted to programs; (b) when there has not been a prior RCT in a similar context; and (c) data on mediators and moderators can be collected inexpensively (Harris & Goldrick-Rab, 2012).

All of these conditions hold in this case; therefore we take a comprehensive approach by asking, and attempting to answer, the following questions:

- *Question 1:* What are the average treatment effects of TDP on college outcomes (college applications, entry, type of college, persistence, and graduation)?

- *Question 2:* What are the mechanisms (mediators) of the TDP average treatment effects? How does TDP influence perceived college affordability, academic preparation, and social capital in high school, as well as high school graduation? What do the patterns of effects on each separate college outcome (entry, persistence, etc.) imply about the mechanisms?

- *Question 3:* How do impacts vary by student subgroups such as race, gender, parent education, prior academic ability, family income, and participation in other college access programs? How about students "on the margin" of program eligibility and college success?

- *Question 4:* How is TDP implemented and how does this affect its impacts? What is the "achieved relative strength" of implementation?

Theoretical Model

Theory informs the potential answers to these questions. As shown in Figure 1, we break down the TDP into three components: the reduced price of college, the on-track letters, and other information about the key steps to college success. We hypothesize that these components collectively result in improved affordability (actual and perceived); better academic preparation;

and increased social capital (Nagaoka et al., 2008). While the three program components are interdependent, the primary influences of components are also suggested by the three arrows (at left). For example, the primary effect of the price reduction is to make college more affordable (first arrow); also, by reminding students of their academic success with the on-track letters, we hypothesize that may motivate students to improve their academic preparation (second arrow). Finally, we believe the additional information we provide about college, combined with the school/cohort-randomization, will facilitate social capital formation (third arrow).

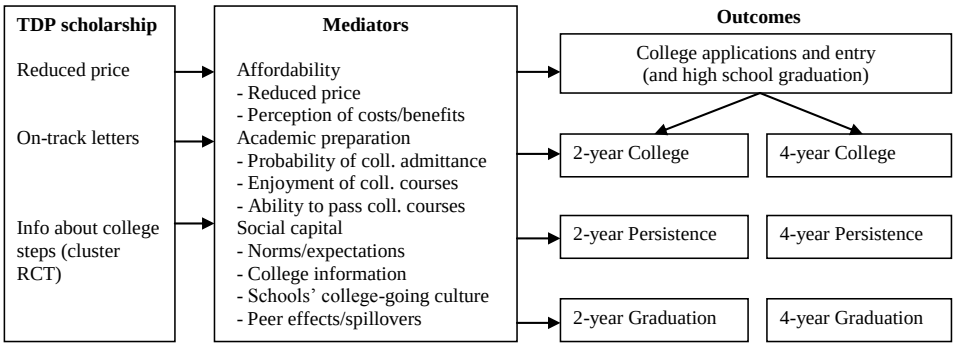

Figure 1. **Theory of change.**

The outcomes are listed on the right side of Figure 1. Research suggests that many of the same general factors affecting college entry also affect persistence and completion with affordability being most important and academic preparation less important in driving dropout (Public Agenda, 2012).[6] In fact, many students appear to drop out because they have chosen colleges that are less competitive than their skills warrant—they are "under-matched" (Bowen, Chingos, & McPherson, 2009). Academically prepared students often never apply to colleges commensurate with their ability (Bowen, Chingos, & McPherson, 2009) partly because of the college-going culture of the school (Roderick, et. al., 2008). Affordability, academic preparation, and social capital may each have different effects on entry versus persistence. For this reason, we list each of the four main college outcomes separately and allow separate paths/mechanisms for each outcome.

Figure 1 oversimplifies matters in at least one important way. We view the decision to go to college—and to acquire academic and social capital—as the result of complex and interrelated processes that are hard to illustrate in a single figure. For example, academic preparation could be influenced by either the increased affordability of college (college is more likely and induces students to work harder to be ready for it) and/or by the GPA and attendance requirements (some students would go to college anyway, but the requirements still induce them to work harder). These complex processes—which are difficult to illustrate with boxes and arrows—become more evident as we elaborate on the theoretical model and incorporate existing evidence about promise programs.

[6] The efficacy of the colleges themselves, especially in engaging students, is also considered an important factor (Tinto, 1993).

Evidence on Promise Programs

Below we discuss the evidence about prior promise programs. Taken as a whole, the existing research is mixed and inconclusive, but tends to favor the theory of change noted above and the idea that promise programs like TDP can help students succeed in high school and in college.

Question 1: What are the average treatment effects of TDP on college outcomes?

There has never been a U.S. randomized trial or rigorous quasi-experiment of a promise program that has reported impacts on college outcomes. We discuss below some rigorous studies of related programs and show that the evidence is mixed.

The best and most positive evidence in our view comes from the Canadian Future to Discover (FTD) program (Fowler et al., 2009). The FTD provides "learning accounts" to low- and moderate-income students in New Brunswick. The program recruited ninth graders into a study in 2004–2005, and when the students reached tenth grade, part of the eligible group—selected by lottery—was promised that, if they met certain minimal requirements, the program would place $8,000 in an account to cover college costs. Students have just recently become college age and in an unpublished working paper, researchers have found large effects on college applications and initial entry (Fowler et al., 2012). The problem with this evidence is that the results vary considerably across subgroups—specifically, there were much larger effects for Francophone versus Anglophone students (and there is no obvious equivalent to the Francophone group in Milwaukee or elsewhere in the U.S.).

Georgia HOPE, like TDP, includes merit requirements and students are aware of the opportunity well before they make college decisions, but a key difference is that Georgia HOPE was offered to a much higher-income student population. All three quasi-experimental studies examining this program have found positive average effects on college-going (Dynarski, 2000; Cornwell, Mustard, Sridhar, 2006; Sjoquist & Winters, 2012). Scott-Clayton (2011) found similar results from a related program in West Virginia.

Three randomized trials have examined the effects of grants and scholarships on students who have *already entered college*, identifying effects on persistence (conditional on entry). MDRC's Opening Doors experiments focus on community college students (Brock & Richburg-Hayes, 2006; Barrow, Richburg-Hayes, Rouse, & Brock, 2010), while a second experiment occurred at a single Canadian university (Angrist, Lang, & Oreopoulos, 2009).[7]

The Wisconsin Scholars Longitudinal Study (WSLS) is the first U.S. randomized trial of a need-based aid program like the Pell grant.[8] In fall, 2008, 1,200 freshmen Pell grant recipients already enrolled in two-year and four-year colleges were randomly assigned to receive grants of up to $1,800 and $3,500 per year, respectively. Like the TDP, the money was provided through

[7] We omit discussion of Gates Millenium Scholars because almost all of the control group went to college, making it impossible to identify effects on recipients (DesJardins & McCall, 2008).
[8] Pell grants are federal aid to low-income students.

the formal financial aid system. The WSLS is particularly relevant here because: (a) like the TDP, the WSLS is a single RCT that involves both the two- and four-year sectors; and (b) the WSLS and TDP evaluations are both designed as comprehensive RCTs.

Two key findings from the WSLS stand out: First, in contrast to Opening Doors and the Canadian experiment, the grant has had no average effect on persistence in either two- or four-year colleges (Goldrick-Rab, Harris, Kelchen, and Benson, 2011). The comprehensive nature of the evaluation allowed us to discover two apparent reasons: (1) students lacked awareness of the grant and its eligibility criteria; and (2) some subgroups lost the grant for reasons unrelated to their behavior (for example, if students' parents saw increased earnings, then this made them ineligible for the Pell grant, which is a pre-requisite for the Wisconsin Scholars Grant, so they lost both grants simultaneously). Both problems are much less likely in the TDP. Great Lakes is doing much more to raise awareness and the TDP does not require Pell eligibility, so students are less likely to lose the grant once they have it. We do not know for certain why the WSLS results were not more positive, or whether our program design will address them.

The larger debate on conditional cash transfers (CCTs) further calls the potential of TDP into question. CCTs include any program that provides financial incentives to people in exchange for specific behaviors and actions. College financial aid is a form of CCT because aid is provided only if students go to college. Perhaps the most relevant evidence on CCTs comes from a series of important and clever experiments conducted by the economist Roland Fryer (2010) to test the impact of financial incentives on student achievement. He compared two types of incentives—those tied to "inputs" such as attendance and reading books and those tied to "outputs" such as course grades and test scores—to see which was more effective. Fryer concluded that "student incentives increase achievement when the rewards are given for inputs to the educational production function [e.g., attendance], but incentives tied to output [grades, test scores, and perhaps college] are not effective" (2010, n.p.).

Overall, on this first research question, we view the evidence as mixed. On the one hand, the FTD experiment as well as two other financial aid experiments on college entrants suggest positive effects. This is also consistent with the conventional wisdom on financial aid in general in various literature reviews (Deming & Dynarski, 2009; Harris & Goldrick-Rab, 2010). On the other hand, the WSLS and Fryer experiments raise some skepticism. It is important to emphasize that there has *never been a randomized trial of a U.S. college aid program designed like the growing number of promise programs around the country, i.e., targeted to students early in their high school years.* As we showed above in the discussion of theory, there are good reasons to think that the early commitments involved in promise programs could be more effective than traditional financial aid programs by getting students on track. This leads to our next question.

Question 2: What are the mechanisms and mediators of the TDP average treatment effects?

Figure 1 posits three interconnected factors serving as potential mechanisms for college access in general and for promise programs in particular: affordability, academic preparation, and social capital. We discuss each below in turn.

Affordability. Basic economic theory suggests that promise programs, as well as other forms of financial aid, increase the likelihood of college success simply by making it less expensive

(Goldrick-Rab, Harris, & Trostel, 2009). While students are, as we show below, responsive to costs, tuition, and financial aid, they do not act like "adolescent econometricians" (Manski, 1993) and do not make education decisions in ways predicted by basic economic theory (Manski, 1993; Beattie, 2002). Part of the problem is that students misperceive the costs and benefits of college. People tend to dramatically overestimate the cost of college—even more so in African-American and low-income populations (Ikenberry & Hartle, 1998). TDP will address affordability by providing substantial funds and by trying to reduce misconceptions about college costs through direct communications with parents as described above. We are aware of no prior evidence about how students' perceptions about college costs are affected by promise programs.

Academic preparation. A substantial literature indicates the importance of academic preparation—specifically, preparation for college-level courses (Adelman, 2006). Some researchers of college readiness have stressed the need to develop basic skills (e.g., reading and writing), content-specific academic skills, and noncognitive skills (e.g., perseverance). This may be why interventions that make financial awards contingent on satisfying academic skill requirements appear to be more effective than those that do not (Deming & Dynarski, 2009; Angrist, Oreopoulos, & Williams, 2010; Scott-Clayton, 2011). TDP could improve academic preparation with its GPA and class attendance requirements and by communicating the connection between high school academics and college goals.

Evidence about effects of U.S. promise programs on high school academic outcomes is also positive, but not very persuasive. For example, the I Have a Dream program, founded in 1981 by Eugene Lang, has been widely viewed as "enormously successful" (Kahne & Bailey, 1999, p. 321) with a "dramatic impact" on high school grades, class attendance, and graduation (as well as on college attendance) (Arete, 2001). Likewise, student achievement increased considerably in Kalamazoo (Michigan) schools after the Kalamazoo Promise was instituted (Bartik, Eberts, & Huang, 2010). For example, of students surveyed in Kalamazoo, 30% said that they had enrolled in more college preparation courses during high school as a result of the promise (Miron, Spybrook, & Evergreen, 2008, p. 7). Also, 58% of the students interviewed and 66% of the school employees believed that students' attitudes about school work had improved (Miron, Jones, & Kelaher Young, 2009, pp. 5–6). *Unfortunately, studies have been hampered by selection bias problems*, due to the ways in which the programs were implemented. The Kalamazoo program evaluators concluded that "it is difficult to determine how much of [the effects] can be attributed to the Promise versus other changes" (Bartik et al., 2010, p. 2). We will not have this problem in the TDP both because of random assignment and the fact that we are closely tracking the other programs students participate in and other changes in school policy.

More recent evidence from Kalamazoo is mixed. Bartik and Lachowska (2011) report a 0.2 standard deviation effect on GPA and one fewer suspension day per year. On the other hand, there were no effects on the number suspensions, suggesting that the effect on suspensions may be due to principals being more lenient in punishing scholarship-eligible students. Also, another recent report suggests that high school graduation rates for minorities have been unchanged since the start of the Kalamazoo Promise.[9] There have not yet been studies of the effects of the Kalamazoo Promise on college outcomes.

[9] There is some evidence on the Florida Bright Futures Scholarship (Harkreader, Hughes, Tozzi, & Vanlandingham, 2008), but this study lacks a comparison group.

Social capital. Social capital can be defined in different ways, but in this case we focus on social norms and information available to students and their capacity to navigate the bureaucratic processes pertaining to college. Limited access to these forms of social capital is likely to be a key factor preventing students from reaching their college potential (Roderick et al., 2009), but the key policy issue is to what degree educational policy can address this mechanism.

Students' peers represent one important social network and have been shown to play an important role in college access (Spielhagen, 2007). PI Harris (2010) has described the complexity of peer interactions in general and surveyed a range of theories to explain them. Building on prior work (Hoxby & Weingarth, 2005; Jencks & Mayer, 1990), he argues that the evidence favors a theory of *group-based contagion*—that is, students tend to emulate the peers with whom they identify most closely (e.g., same race or gender) (Harris, 2010). Thus, if getting a TDP scholarship can directly influence the information and beliefs of individual students, then this could have positive spillovers for their classmates and friends in terms of group information and social norms around college-going.

School teachers and counselors can also be viewed as sources of social capital. Some research shows that students need "structures of support" to help them navigate the college selection, admission, and financial aid processes (Goldrick-Rab & Pfeffer, 2009; Roderick et al., 2009). Educators, along with peers, also help to establish a college-going culture that sets a norm of college entry (Gamoran & Hannigan, 2000). This highlights how social capital and academic preparation are intertwined. A college-going culture, for example, may induce students to take more demanding courses and study harder. Further, if college-going becomes more a norm, then students are likely to gather more college information and share that through their social networks. In this way, peers and counselors can help to offset limited social capital in families.

Question 3: How do impacts vary by student subgroups?

Most studies of subgroup effect (or "effect heterogeneity") focus on demographic characteristics. Recall that in the FTD experiment, there were much larger effects on Francophone versus Anglophone students. PI Harris and Goldrick-Rab (2012) conducted an extensive review of evidence about the variation in financial aid impacts by race, gender, family income/SES, ACT/GPA, and traditional student status.[10] They concluded that effects are generally larger for disadvantaged groups and women, although program designs and contexts vary so widely that it is difficult to separate subgroup differences from program design effects. The WSLS results suggest that the least academically prepared students benefit most, but this is at least partly because these same students were most likely to keep the grant.

A key goal of our evaluation of the Milwaukee TDP program is to both identify and help explain such differences. We argue that the standard demographic dimensions may not be the most useful way for understanding subgroup effects. Instead, we expect the effects of TDP to be largest for students "on the margin" of college success—those who could be nudged over the

[10] The widely cited financial-nexus model also suggests effect heterogeneity by subgroup (Cabrera, Nora, & Castañeda, 1993; St. John, Paulsen, & Carter, 2005).

threshold to college by something like a scholarship. Some minority students are doing well academically and others poorly. Thus, the effects estimated by race and other subgroups, even if they differ from other subgroups, are still hard to interpret.

What it means to be on the margin depends on the specific college outcome. At the one extreme are students who are already planning to go to expensive four-year colleges and can cover these costs through large expected family contributions. These students (rare in our sample) may be more likely to persist and graduate because college will still be cheaper, but no more likely to apply or enter. Then, there are students who would have gone to a four-year college, but are on the margin of choosing a cheaper one. Perhaps the largest effect, given the low-income MPS population, will be on students who are on the margin between attending a two-year college versus a four-year college. Finally, at the other end of the spectrum, are students who would not have even graduated from high school; they will probably be least affected because for them a two-year college degree may already have been within reach, but they would not have gone for other reasons (e.g., they simply do not enjoy classes). (This is why we omitted high school graduation from Figure 1, though we will of course test for such effects.) In short, it is the students *in the middle* who are most likely to benefit.

More generally, we hypothesize that the level of social, academic, and financial advantage of the marginal students will increase as we progress from less to more ambitious college outcomes. Testing this, as we explain in more detail below, requires examining all of these characteristics (race, gender, parent education, etc.) simultaneously. We will also replicate what other studies have done in comparing effects by individual demographic characteristics.

Question 4. How is TDP implemented and how does this affect its impacts?

There is very little evidence about implementation of promise programs except what we gathered ourselves from program administrators in Kalamazoo, Pittsburgh, and elsewhere in the course of designing the TDP program. Program implementation almost always plays a key role (Pressman & Wildavsky, 1984; Sherwood & Doolittle, 2003) and the same is no doubt true here as well. In this case, we believe the key is how well students remember they have the scholarship and how well they understand the rules—which we try to influence through our aggressive communications strategy--as well as what messages students hear about it from their parents and educators in their schools. The role of messages has been found to be in important in other aspects of public policy as well (Bloom, et al., 2003).

Summary of Theory and Evidence. The theory and evidence reviewed above lays the groundwork for our analysis. There are promising signs from prior studies of other types of financial aid, from the FTD experiment in Canada, and from exploratory research of the Kalamazoo Promise. On the other hand, the FTD experiment primarily benefited Francophones and the Kalamazoo research is not very convincing. Also, the results from Fryer's experiments at the K-12 level and the WSLS suggest that promise programs may not be effective. In designing the program with Great Lakes we attempted to draw on the lessons of research to design a program that would generate potentially larger effects than traditional financial aid:

- TDP starts in ninth grade and has more time to work, in contrast to FTD (tenth grade), traditional aid (twelfth grade) and WSLS (after college entry);

- TDP is a cluster-randomized trial (i.e., entire schools were randomized) to facilitate the role of peer effects and college-going culture (Fletcher & Tienda, YEAR; Harris, 2010; Spielhagen, 2007), in contrast to all prior randomized trial of college aid which randomize individuals;

- TDP includes performance requirements, which may increase effects relative to purely need-based programs (Baum et al., 2012; Deming & Dynarski, 2009);

- TDP includes output-oriented performance requirements, but, as suggested by Fryer (2010), also includes school attendance as an input-oriented requirement;

- TDP provides considerable information to students so that they better understand the steps, or inputs, needed for college success (Angrist et al., YEAR; Roderick et al., 2009); and

- TDP does not require eligibility for Pell or any other grant, reducing the possibility that, as in the WSLS, students might lose the TDP for reasons outside their control.

In short, by increasing the real and perceived affordability of college and clearly communicating the path from high school to college, our theory is that early commitment programs improve academic preparation and social capital during high school and thereby increase college success—both entry and completion. The U.S. invests more than $177 billion per year in college financial aid. Making these same financial commitments more concrete and at an earlier stage would be essentially costless and could have a meaningful effect on students.

Implementation of the Experiment

The Study Sample

The TDP will serve primarily low-income students with 78% of the TDP sample being eligible for free or reduced price lunches. As shown in Table 1, the district is racially and ethnically diverse with substantial percentages of African-Americans (63%) and Hispanics (20%) as well as Whites (11%). Test scores, while being well below the national median, are almost identical to the national urban district average in reading (below average in math). Parent income is slightly below the national urban district average (UW-Milwaukee, 2010). Thus, in terms of academic ability, this sample of Milwaukee students is similar to other districts where promise programs have taken hold (e.g., Pittsburgh and Kalamazoo).

Below are the means and distributional information for some key characteristics of students and their families. Table 1 demonstrates both the overall level of disadvantage and a fair degree of diversity, which allows us to examine effect heterogeneity (Research Question 3).

Table 1 ***Sample Characteristics***

Variable	*MPS Data*		
	Average	*10th percentile*	*90th percentile*
African-American	0.63	---	---
Hispanic	0.20	---	---
Free/reduced lunch	0.78	---	---
Math scores (percentiles)	28.8	3	66
Mother's education (years)	12.37	7	16
Family income (2011 $)	29,912	62	75,000

Notes: The race/ethnicity and free/reduced lunch, and test score information come from MPS administrative data and refer to the TDP sample in their 8th grade year. Mother's education and family income come from a separate survey of a similar Milwaukee sample by the UW Survey Center.

Randomization

Random assignment of schools to the TDP program facilitates unbiased estimates of program effects. To ensure that, the research team, on behalf of Great Lakes, carried out the randomization.

Cluster randomized trials generally have low statistical power. To improve precision, we used a *paired randomization* approach. Specifically, we: (a) ranked schools by the college entry rates from recent cohorts, (b) created pairs of schools based on the rankings (e.g., the two schools with the highest prior college enrollment rates would form the first pair), (c) randomized one school in each pair to the treatment, and (d) pooled to obtain the overall control and treatment groups (Bloom, 2010). This approach maintains internal validity and generally increases statistical power considerably (Bloom, Richburg-Hayes, & Black, 2007; Greevy, Lu, Silber, & Rosenbaum, 2004; Imai, King, & Stuart, 2008; Imai, King, & Nall, 2009). We discuss statistical power to a much greater extent below.

Through this randomization process, half (18) of the 36 MPS schools were assigned to treatment and the remainder to the control group. All students enrolled as first-time ninth graders in selected schools on November 1, 2011 were sent a TDP award letter. Students remain in the same treatment condition regardless of whether they switch schools subsequently (though they still have to graduate from an MPS school to receive the money); therefore, this approach does not create an incentive for any MPS student to switch schools to receive the treatment. In all, 2,587 students were assigned to treatment and 2,464 students were assigned to control.

Data Collection

We have already begun to collect data to answer our research questions. The majority of the data come from existing administrative sources and are de-identified (i.e., student names and other identifying information will be omitted). This approach allows for extensive analysis at a very low cost. In addition to extensive administrative and survey data from multiple sources, we plan to interview students, school counselors, and principals. Table 2 below summarizes the main constructs and exactly how each will be measured.

Administrative Data

MPS and State student records. The primary threat to validity in randomized trials is differential nonresponse/missing data. One way to avoid this problem is to obtain nearly complete administrative data. MPS has agreed to provide de-identified student records for all MPS ninth graders who remain in the MPS system. The Wisconsin Department of Public Instruction (DPI) has agreed to provide data on students who depart MPS but remain in public schools in Wisconsin. The MPS and DPI data allow longitudinal linkages of individual student data using unique identifying numbers.

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

National Student Clearinghouse. We will obtain nearly complete data on college attendance and completion from the National Student Clearinghouse (NSC). The NSC collects information from the colleges and universities attended by 92% of U.S. undergraduates. Only 6 of the 64 eligible Wisconsin colleges do not participate in the NSC—all very small colleges that are not among those frequently attended by MPS students. Because NSC is a national data system, we will observe enrollments even for students who attend a college that is not TDP eligible. Also, the fact that NSC is a near-census of college enrollment, means that it is generally reasonable to assume that if a student is not shown as enrolled in the NSC then the student is not enrolled anywhere. (We are currently testing the accuracy of NSC in a separate study using transcript data and are finding that it has very high accuracy. This is unsurprising given that the NSC data originate from the colleges themselves.)

Kommentar [P1]: Do we want to emphasize that we do not need to obtain student permission for this data – incredible efficiency!

The NSC includes reliable data on college enrollment (including 2- and 4-year colleges), persistence, and graduation. Data are available for each individual college and term/semester a student attends college. Enrollment intensity (part-time, full-time, etc.) is also included. Because it is directory information (i.e., does not provide Social Security Number or other sensitive information), student consent is not necessary to use the NSC data for research purposes.

Administrative data from other state agencies. Education is a long-term investment and even a college degree is not really the end goal. Programs like TDP are intended to help students throughout their lives. In studies such as the Perry Preschool Project, some of the largest effects came not from higher test scores but from participants' increased earnings and reduced reliance on government programs such as welfare. We will examine these effects using state data collected for research purposes by the University of Wisconsin Institute for Research on Poverty (UW-IRP) through agreements with state agencies. These records include employment, earnings, incarceration, and participation in a host of government programs. MPS has agreed to send names and birthdates to UW-IRP who will then match the data to their records using a process similar to the National Student Clearinghouse (see above).

Survey Data

MPS administers a variety of online surveys of students and teachers each year. The survey responses include the same student and teacher identifiers as the other MPS records and can therefore be linked to the administrative and other data sources at the individual level.

MPS School Climate Survey (Students). The MPS Student Climate Surveys are based on the well-respected and research-based surveys developed and administered in the Chicago Public Schools in conjunction with the Consortium on Chicago School Research (the CCSR director is a consultant on this project; see Personnel section). One key construct of interest is students' peers and their interactions, as a form of social capital. Regarding academic preparation, we are also interested in how student interact with teachers. Student responses to these questions have been shown to be strong predictors of college-going (e.g., Nagaoka et al., 2008). Because we can obtain these and data on grades and test scores from prior years for the same students, these are important sources of baseline data. (See Table 1 above and baseline equivalence below.)

MPS High School Exit Survey (Students). The most important data elements from the High School Exit Survey are the specific colleges students applied to (up to 8) and the colleges to which students have been accepted. The survey also includes questions about (a) perceptions of the expectations of, and support received from, parents, teachers, and counselors; (b) college plans, including planned course of study and possible delayed entry; (c) participation in other college access programs; and (d) parents' educational background. We are particularly interested in the survey items about another source of social capital: parents. MPS has agreed to allow us to add some questions to their surveys for this project; among these will be questions about parental expectations and our third mediator: perceptions about financial aid and college affordability.

MPS Climate Surveys (Teachers). TDP might influence students, as noted above, by changing the college-going cultures of schools, as reflected in interactions among students and teachers. The teacher climate survey and instructional practice survey will complement the student surveys in gauging these changes. As with the student surveys, MPS links teacher survey responses to teachers' unique identifiers. The pertinent constructs of the instructional practice surveys are (a) teacher demographics, (b) expectations about students' college-going, (c) orientation toward higher order thinking skills and academic rigor, and (d) student engagement.

MPS Youth Risk Behavior Surveillance System (YRBSS). MPS participates in the YRBSS monitoring system. Managed by the Centers for Disease Control (CDC), YRBSS is a national survey that asks students about a variety of health-risk behaviors: sexual behaviors that contribute to unintended pregnancy and sexually transmitted diseases, alcohol and other drug use, tobacco use, unhealthy dietary behaviors, and inadequate physical activity, obesity and asthma among youth and young adults. We will examine effects on these non-academic outcomes as well. Students who are more future-oriented, specifically toward college, may engage in less risky behavior.

Survey Measures and Response Rates. MPS specifies its own survey constructs, each of which is measured through multiple items. We will conduct factor analysis to test the validity of the conceptual map and to construct new variables. We will further develop measures through Rasch analysis so that constructs are comparable over time, are checked for item fit and reliability, and produce standard errors of measurement based on response pattern consistency.

In 2012, Climate Survey response rates were 53-59% for students and 51% among teachers. MPS has agreed to take additional steps to increase these rates. We will be including financial incentives in future years to increase response rates. In another research project conducted in MPS, researchers obtained at least 70% response rates from staff in every school (student surveys were not relevant to this other project) with incentives of just $200 per school. Because student surveys are administered during the school day, student response rates are, according to MPS staff, heavily driven by whether teachers obtain computer lab time for students to fill out the survey. To encourage higher response rates, we will offer school staff incentives, tying them both to student and staff survey response rates. Schools in which student and teacher response rates exceed 70% will receive $500 to be used for any purpose. Also, MPS has agreed to send school principals regular updates of their respective survey response rates and to remind them in weekly communications about the $500 incentive and the importance of the surveys.

Qualitative Interviews

While we have extremely rich administrative and survey data, comprehensive RCTs also benefit from qualitative interviews with participants and program providers in order to understand program implementation, develop alternative hypotheses, explain puzzling findings, and to triangulate unclear patterns of evidence to develop more convincing explanations.

Student interviews. The utility of embedding in-depth interviews in an experimental study is widely recognized. While quantitative approaches are useful for testing predetermined hypotheses, they "may not help discern the full range of explanatory processes that hold in any particular cause-effect relationship" (Yoshikawa, 2008, p. 347). The ability of interviews to uncover additional mechanisms as part of comprehensive experiments has helped improve the interpretation and use of several recent large-scale experiments in social policy, including the PI's own work on the WSLS (discussed in the Prior Research and TDP Design section).

From the overall consenting sample[11], we selected a stratified random sample of 6 schools (three control, three treatment) and, within each of those schools, selected a stratified sub-sample of 4 students for interviews. Given our theory that students near the margin of the scholarship requirements will be affected, we specifically selected two students in each school who were near the 2.5 GPA/90% attendance threshold based on pre-treatment data, one student who was far below these levels, and one student who was far above. We confirmed that this approach also created an overall pool of 24 students that is diverse in terms of race/ethnicity and gender. At this point, we have carried out interviews with some of these selected students, but we have limited funding for this at this point. We plan to interview the 24 students annually, including the period after they leave high school.

Interviewers will have information about students' treatment status, as well as their most recent academic progress and survey responses. Interviews will be audio-recorded, transcribed, and coded. Students will be paid $10 for their time.

Table 2 ***Summary of Constructs and Examples of Measures***

Affordability
- "What do you see as the greatest barrier to college?" (Student Climate Survey; student interviews)
- "Select the two most important reasons why you chose this college" [e.g., "cost" and "financial aid package"] (Exit Survey)
- "How are you planning to pay for college?" (High School Exit Surv.)
- Family income and aid packaging (FAFSA)
Academic Preparation
- Grades, courses, state tests, district tests, ACT (MPS/DPI admin. data)
Social Capital
Information about college
- "How often do you talk to friends, family, and school counselors about college?" (Student Climate Survey)
- "How often you have done the following during this school year: "Talked to a counselor college; Attended a college fair or met with a college representative; Visited a college or university; Reviewed college materials at your school counseling office" (Exit Survey)
Expectations and College-going culture
- "Students at my school focus on learning" (Teacher Climate Survey)
- "Teachers in this school expect most students to go to college." (Teacher and Student Climate Surveys)
- "What do you do to help students applying to college?" (Interviews with counselors)
- "Teachers at my school feel it is part of their job to prepare students to succeed in college." (Student Climate Survey and Exit Survey)
- "Teachers at my school help students plan for college." (Student Climate Survey, and Exit Survey)
- "How much do your teachers and counselors want you to go to college?" (Student Climate Survey)
Peers
- "How many of your friends received the TDP scholarship?" (Student Climate Survey)
- "Students at my school focus on learning" (Student Climate Survey)
- "Most of the students in my school are planning to go to college." (Student Climate Survey)
- "How many of your family and friends have gone or expect to go to college?" (Student Climate Survey)
Family
- "My parents encouraged me to continue my education after high school" (Exit Survey)
- "My parents talked to me about colleges suited to my interests and abilities." (Exit Survey)

[11] We obtained consent by regular mail for the interviews from a small portion of students and schools in October and early November, 2011, prior to the program announcement.

College and Other Long-Term Outcomes
- College entry, college type, persistence, graduation (NSC)
- Employment, income, incarceration, gov't health and welfare benefits (state agencies and UW-IRP)
Implementation
- "Since November of 2011, have you received a letter saying that you received a college scholarship for $12,000?" (Student Climate Survey)
- "Please indicate if you need to fill each of these requirements in order to receive the $12,000 scholarship: 2.5+ GPA, Pass the state test (WKCE), 90% attendance, Attend summer school, and Graduate from high school on time." (Student Climate Survey)

Key topics in the student interviews will be:

- How their parents, teachers, and peers have affected their thinking about college
- Beliefs about the importance of college and steps required to matriculate
- Perceptions about college costs and understanding of college financial aid
- Knowledge of their own TDP status and that of their friends and classmates
- Understanding and impressions of TDP: program eligibility requirements, etc.
- Number and type of communication and messages about TDP from school staff
- How TDP affected their beliefs, attitudes, and behaviors

Interviews with counselors and principals. We will attempt to interview all 36 principals during the students' junior year and all the counselors during their senior year. Some schools have more than one counselor (assigned by grade) and we will identify the one assigned to the TDP cohort. These interviews will focus on:

- Counselor caseload and general roles and responsibilities
- Counselor role in college application process
- College-going culture and messages communicated to students
- Awareness and actions taken pertaining to the degree project

Consent, FAFSA, Aid Packages, and Long-Term Follow-Up

In their senior years, we will obtain consent from a subset of students for additional long-term data collection—in particular, income information from the FAFSA and other information from students' actual financial aid packages. This consent process (and brief survey) will be administered by the University of Wisconsin Survey Center (UWSC). Because this population is relatively hard to reach, we have developed an aggressive consent process with advance phone calls, $5 pre-incentives, and $10 post-incentives, postcard reminders, and three full mailing of the consent and survey spread over several months.

FAFSA completion is a requirement for receiving TDP funds; therefore it is important to know what portion of students' completed the form. (Students will be reminded to do so by Great Lakes during their senior year.) Also, FAFSA data include important information about students' family income. Prior research suggests that the effects of financial aid vary by family income (Harris & Goldrick-Rab, 2012). We will estimate impacts by income subgroup among consenters using (de-identified) data from the FAFSA, provided by the Wisconsin Higher Educational Aids Board (HEAB), the state agency helping to administer TDP. These data will be much more detailed than those available from the state agencies through UW-IRP.

Data Analysis

We organize our discussion of data analysis according to our four research questions. Since all of the analyses will rely on baseline equivalence or "balance" between the control and treatment groups, we start by reporting baseline equivalence of the control and treatment groups.

Baseline Equivalence

For baseline equivalence, we focus on 8th grade (pre-treatment) values of the dependent variables because these are the most crucial from a balance standpoint. The first three variables in Table 3 (attendance, GPA, and math scores) are for the TDP sample. Because college success is the real goal, we also report the prior college-going rate in the same schools for the 2005 cohort of ninth graders (the pairing variable). Looking across all the tests, the pairing process worked well and the control and treatment groups are well balanced.

Table 3 Baseline Equivalence on Lagged Dependent Variables

Variable	*Control*	*Treatment*	*Difference*	*p-value for Diff.*
Attendance (percentage)	90.9	90.4	-0.510	0.758
GPA	2.250	2.190	-0.060	0.816
Math scores (s.d. units)	0.014	0.002	-0.012	0.712
Prior college-going (%)	41.6	41.9	- 0.300	0.552
F-test	--	--	--	0.564

Notes: Number of observations vary across rows, but are all in the range of 3079-4048.

Given that the groups are balanced on measureable characteristics at baseline, we can turn to how we will analyze the four general research questions. In all of these, we will focus primarily on intent-to-treat (ITT) analyses because anyone who becomes aware of their treatment status is in some sense "treated" with this type of intervention.

General Analytic Approach

Generally, the unit of data analysis in randomized trials is the unit of randomization—in this case, schools. Further, if simple randomization had been used, we would begin with simple differences in school means and, to obtain greater statistical power, follow with the usual covariate-adjusted models. As noted in the Sample Selection section, however, this is a clustered design with paired randomization and this requires adjusting the usual analysis.

This implies the following model:

$$Y_{st} = \beta_0 + \beta_1 Y_{s,8th} + \beta_2 X_s + \beta_3 T_s + \kappa_p + \varepsilon_{st} \quad (1)$$

where outcome measure Y in school s at time t is a function of the lagged value of the dependent variable ("pretest") from eighth grade (where feasible), X_s is a vector of other time-invariant covariates (e.g., student demographics) and lagged values of *other* dependent variables (e.g., lagged reading scores when math scores are the dependent variable), and κ_p is a vector of pair

indicators[12] with p=1,2,…,18 (i.e., indicating the randomization pair). Including lagged values as covariates improves precision considerably (Raudenbush, Martinez, & Spybrook, 2007).

The variable T_s represents the treatment status of the schools/students so that β_3 is the average treatment effect. This is similar to the approach used by, for example, Kane and Staiger (2008) who randomized pairs of teachers and then estimated a model akin to (1) with many covariates and indicators for each pair of teachers.

For all of these variables, the school-level means will be calculated from the underlying student-level data, allowing us to keep each student in the original school of assignment for analysis purposes. This is important because, unlike most cluster randomized trials (e.g., whole school reform), individual students maintain their treatment status even if they leave the "treated school." Mobility across MPS schools is common and important to account for in this way. (We will test whether impacts are smaller for students who switch to control group schools where peer spillovers will differ, recognizing that mobility is endogenous.)

The above equation and approach applies to essentially all the outcomes, except that with outcomes such as college entry and persistence, there are no lagged values.

Answering the Research Questions

Below we discuss how we will try to answer each of the four questions posed earlier.

Question 1: How does TDP influence college outcomes? We can answer this by estimating logit versions of equation (1) using NSC measures of college entry, persistence, and graduation. We will also examine the likelihood students attend two-year versus four-year colleges and their likelihood of attending in-state (a requirement of the program).

The NSC college data will include nearly all students in the original sample no matter where they go to college—public or private, in state or out of state. Moreover, any errors in the NSC data are likely to be unrelated to treatment status (see Goldrick-Rab & Harris, 2010) and therefore do not pose a significant threat to internal validity. With randomization and almost no missing data, these estimates will therefore have extremely high internal validity. These models will be estimated with covariate adjustments to increase statistical power.

Question 2: What are the mechanisms and mediators of the TDP impacts on high school graduation and college-going? One straightforward aspect of mediator analysis is to estimate equation (1) using the measures in Table 3 and other measures of the mediating variables: affordability, academic preparation, and social capital (Raudenbush, 2011). These estimates are unbiased, but the effects of the mediators on outcomes like college entry are generally biased (e.g., students are not randomly assigned to their academic trajectories). Some have proposed

[12] The pair effects are necessary because of the paired randomization and can be modeled as fixed or random depending on the desired inference. If we are drawing conclusions only about these particular pairs, then fixed effects are appropriate; if we think of the pairs as a sample from a population of pairs, then this implies random effects (Hedges, 2009). We are interested in both inferences and will therefore use both approaches.

using instrumental variables to solve this problem, but this approach also requires restrictive assumptions (Gennetian et al., 2005).

We believe the best way to go about mediator analysis is to triangulate across different types of data and analysis. First, the patterns of effects on different college outcomes tell us something about the mechanisms. For example, if the promise scholarship has no influence on grades and test scores, then this would tend to suggest that academic preparation is not a mediator—and this interpretation would be reinforced if the TDP has no influence whether students are accepted to four-year colleges. We will also examine the possibility that the mediators are highly correlated, which would make it more difficult to disentangle their roles.

Qualitative interview data also lend themselves well to mediator analysis and therefore we briefly discuss our approach to these data. Interview data will be analyzed by (a) organizing transcripts chronologically by individual students and across cohort years; (b) manually coding respondents' references within the overall topical areas covered in the interview questions; (c) examining trends to discover key categories that emerged from the transcripts; (d) re-coding transcripts within these key categories, and (e) compiling data within these coding categories into a tabular format to formulate themes. This process will be used within and across the student, counselor, and principal interviews.

We can also incorporate the interviews with counselors and principals to carry out quasi-experimental methods to examine the role of program implementation. It may turn out that some schools were more aggressive than others in communicating about the TDP. In that case, we could compare the measured impacts on TDP knowledge (from the student surveys) from schools that were less active in promoting TDP. This is of course non-experimental and might be driven by unobserved differences in the schools (and the school-level randomization limits power). We could also complement this approach with propensity score matching as an alternative quasi-experimental method (Rosenbaum & Rubin, 1983).

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

Question 3. How do impacts vary by student subgroup? We begin to answer this question by re-estimating equation (1) with interaction terms for each subgroup. In addition to the demographic and academic measures, we have student-level data on participation in other college access programs allowing us to explore program interactions. Table 1 shows that, even in a single large district, there is considerable heterogeneity in student characteristics.[13] Before

[13] While the cluster sizes fall in the subgroup analysis, statistical power is much more sensitive to the number of clusters than to the number of students. Therefore this analysis will also be fairly powerful.

conducting any such tests, we will carry out the same tests of baseline equivalence shown in Table 3 for the respective subgroups.[14]

As noted above, however, looking at each characteristic separately might not be very informative. A student from a high-income family might have a low likelihood of college-going if the student has very low academic skills. Conversely, a low-income student with strong academic skills and a parent who went to college might have better college prospects. This suggests a need to look at the characteristics simultaneously to identify students on the margin of college success. We will calculate effects based on the predicted outcomes (e.g., probability of going to college) as follows: (1) regress the probability of college entry on the above factors (gender, race, eighth grade academic performance, etc.) using prior cohorts of students; (2) use this model to obtain the predicted college entry probability for each student in the TDP sample (control and treatment); and (3) divide students into two or more groups based on these predictions and test the impacts separately by group. We are conducting similar analyses in the WSLS (Goldrick-Rab, Harris, Benson, & Kelchen, 2011).

Question 4. How is TDP implemented and how does this affect its impacts? Our analysis on this question is based on a series of papers by David Cordray (e.g., Cordray, 2005; Hullemna & Cordray, 2009). He defines the *achieved relative strength* (ARS) of an intervention as the difference in actual implementation experienced by the treatment and control groups while the *expected relative strength* (ERS) is the intended implementation for each group. The first step is to identify the key components of the intervention and what constitutes fidelity of implementation for each. In this case, the intervention is fairly simple: a performance-based promise scholarship to attend college. Thus, fidelity of implementation is best measured by student awareness of the program and the incentive structure.

If students were perfectly informed, then 100% of students would accurately report their treatment status and the program requirements. We measure ARS using two survey questions on the recently completed 2012 MPS Student Climate Survey. Of those students who responded to these climate survey questions, 65% of the actual treatment group accurately reported being a scholarship recipient. We also asked students about the specific scholarship requirements; 45% of those who correctly reported being scholarship recipients also got all 5 requirement questions correct, far more than the 3% expected by chance.

There are different ways to calculate the ARS index depending on the nature of intervention and the data available (Cordray, 2005). In this case, we assign individual students who correctly reported their treatment condition *and* all the requirements an index value of 100. At the other extreme, students who mis-reported their treatment condition are assigned a value of 0 regardless of whether they got the requirements correct. In between are students who correctly report their treatment condition and got one program requirement wrong (i.e., 83.3% correct), two wrong (67% correct), and so on. This yields an ARS index of 53% (0-100 scale). The same survey questions will be asked each year and the ARS index will be cumulative across years.[15] In the

[14] Some researchers also suggest carrying out heterogeneity analyses by balancing each subgroup on measureable characteristics using propensity score matching (PSM) (Brand & Xie, in press) and related methods (Djebbari & Smith, 2008).

[15] In analysis of post-entry outcomes, we will also consider the share of funds received as loan reduction as part of ARS, though loan reduction is an intended aspect of implementation.

impact analysis, we will estimate impacts on subgroups defined by varying levels of the ARS index.

Power Analysis

The statistical power of an analysis is the probability of rejecting the null hypothesis (in this case, no TDP effect) when it is indeed false. Researchers typically require power of at least 0.80, meaning that there is an 80% chance of rejecting the null hypothesis when it is false. While statistical power is often a concern in cluster randomized trials, we believe that statistical power is strong in this study. Recall (a) the large number of lagged covariates and (b) the fact that the paired randomization is based on a lagged value of a key dependent variable (college entry in prior cohorts). As Bloom has shown, the "breakeven" R^2 for the pairing variable (that is, the R^2 necessary to make paired randomization worthwhile) only has to exceed 0.06 to improve precision with 36 clusters (Bloom, 2005). Because we are pairing on a lagged dependent variable, the R^2 is much higher. Using the correlation among Milwaukee public schools, we estimate a correlation of 0.72 between 2010 college entry rates (by school) and college entry rates averaged across the 2005 and 2006 cohorts. This implies an R^2 of 0.52. The effective R^2 rises further when we add additional lagged dependent variables.

We used 7th and 8th grade (baseline) data for the TDP sample to estimate the intraclass correlations (ICCs) and R^2 (percent of variance explained by lagged dependent variables) for each key dependent variable. The R^2 are from the regression in equation (1), again using the prior cohort data. We apply the ICCs and R^2 calculated from the MPS data using software by Spybrook, Raudenbush, Congdon, and Martinez (2009). We can think of this as a RCT with 18 blocks with 2 clusters per block. Table 4 summarizes the calculations. We focus on the last two rows because college outcomes are most important. The results suggest we have an 80 percent chance of identifying an effect of 8 percentage points on college entry and completion. These are much smaller than many of the point estimates found in the FTD experiment. Specifically, they found average effects of 10-13 percentage points on college entry; the Francophone subgroup saw effects of 18 percentage points on college entry. Thus, the power here is ample so long as the effects in Milwaukee are at least half of the largest subgroup effects in Canada. While this is a different population, we view this as very plausible given that school-level randomization (and associated peer effects etc.) should actually generate *larger* effects from TDP (FTD was individual randomization). In any case, if an intervention as expensive as TDP does not increase college entry or graduation by 8 percentage points, it is probably not of interest for policy.

Table 4 Power Analysis Summary

Dependent Variable	*ICC*	R^2	*Missing Data Proportion*	*N (clusters/size)*	*MDE*	
					Effect Size	*Perc. Points*
Attendance	0.145	0.564	0.20	36/112	0.26	---
GPA	0.266	0.581	0.22	36/109	0.33	---
Math Scores (MAP)	0.199	0.849	0.22	36/109	0.19	---
College Entry/Graduation	0.150	0.610	0.00	36/144	---	0.08

Test scores are typically reported in effect size (standard deviation) so the 0.19 MDE in Table 4 can be readily interpreted. The standard deviation on attendance is 11 percentage points implying that we are powered to identify effects of 0.26*11 or 2.9 percentage points per year. For GPA, the MDEs translate to 0.34 GPA points (4.0 scale). The FTD researchers do not report effects on any of these outcomes.

In analyses with multiple outcomes and multiple subgroups (heterogeneity), standard analyses yield high probabilities of making at least one Type I error. To address this multiple comparisons problem, we will follow recommendations of Bloom and Michalopoulos (2010) and Schochet (2008)—specifically, (a) outlining prior to the analysis a set of "confirmatory" tests based on prior evidence and theory and (b) adjusting the *p*-values using approaches such as the Benjamini-Hochberg method. We will report both the unadjusted and the adjusted *p*-values.

Bibliography and References Cited

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

Allensworth, E. (2006). Update to: *From high school to the future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from four-year colleges*. Chicago, IL: Consortium on Chicago Schools Research.

Allensworth, E. & Easton, J. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago, IL: Consortium on Chicago Schools Research.

Allensworth, E. & Easton, J. (2007). *What matters for staying on-track and graduating in Chicago Public School*s. Chicago, IL: Consortium on Chicago Schools Research.

Angrist, J. & Lavy, V. (2002). The effect of high school matriculation awards: Evidence from randomized trials (NBER Working Paper No. 9389). Retrieved from National Bureau of Economic Research web site: http://www.nber.org/papers/w9389

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics, 1*(1), 1–28.

Angrist, J., Oreopoulos, P., & Williams, T. (2010). *When opportunity knocks, who answers? New evidence on college achievement awards* (NBER Working Paper No. 16643). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w16643

Arete Corporation. (2001). *"I Have a Dream": The impacts*. Retrieved from http://www.ihaveadreamfoundation.org/images/downloads/AreteSummary_2003.pdf

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The condition of education 2011* (NCES 2011-033). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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.

Bailey, T. (2011). *Can community colleges achieve ambitious graduation goals?* Washington, DC: American Enterprise Institute.

Barnett, S. W. (1992). Benefits of compensatory education. *Journal of Human Resources, 37*(2), p.279-312.

Bartik, T. J., Eberts, R., & Huang, W.-J. (2010, June). *The Kalamazoo promise, and enrollment and achievement trends in Kalamazoo Public Schools.* Paper presented at the PromiseNet 2010 Conference, Kalamazoo, MI.

Bartik, T. & Lachowska, M. (2011) *The Short-Term Effects of the Kalamazoo Promise Scholarship on Student Outcomes*. Paper presented at the 2011 annual meeting of the Association for Public Policy and Management, Washington, DC.

Baum, S., Breneman, D., Chingos, M., Ehrenberg, R., Fowler, P., Hayek, J.,…Whitehurst, G. (2012). *Beyond Need and Merit: Strengthening State Grant Programs*. Washington, DC: Brookings Institution.

Beattie, I. (2002). Are all 'adolescent econometricians' created equal? Racial, class and gender differences in college enrollment. *Sociology of Education, 75*(1), 19–43.

Bettinger, E. (2004). How financial aid affects persistence. In C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 207–238). Chicago, IL: University of Chicago Press.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2009). *The role of simplification and information in college decisions: Results from the H&R Block FAFSA experiment* (NBER Working Paper No. 15361). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w15361

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). New York, NY: Russell Sage.

Bloom, H. S., (2010). *Sample design for group-randomized trials*. Presentation prepared for the IES/NCER Summer Research Training Institute held at Northwestern University on July 27, 2010.

Bloom, H. S., Hill, C. J., & Riccio, J.A. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management. 22*(4) 551–575.

Bloom, H. S., & Michalopoulos, C. (2010). *When is the story in the subgroups?* Strategies for interpreting and reporting intervention effects on subgroups. MDRC Working Papers on Research Methodology. New York, NY: MDRC.

Bloom, H. S., Richburg-Hayes , L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Bowen, W. G., Chingos, M. M., & McPherson, M. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Brand, J. & Xie, Y. (in press). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*.

Brock, T. & Richburg-Hayes, L. (2006). *Paying for persistence: Early results of a Louisiana scholarship program for low-income parents attending community college*. New York, NY: Manpower Development Research Corp.

Cabrera, A. F., & La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education, 42*(2), 119–150.

Cabrera, A. F., Nora, A., & Castañeda, M. B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *Journal of Higher Education, 64*, 123–139.

Chetty, R., Friedman, J.N., & Rockoff, J. (2011). *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood*. NBER Working Paper 17699. Cambridge, MA: National Bureau of Economic Research.

Chingos, M. (2011). *Graduation rates at America's universities: What we know and what we need to know*. Washington, DC: American Enterprise Institute.

Cordray, D. S. (2005). *Assessing Implementation Fidelity and Achieved Relative Strength in RCTs.* Presentation for Researchers Without Borders. June 9, 2010.

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. Bootzin & P. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103–124). Washington, DC: American Psychological Association.

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.

Djebbari, H. & Smith, J. (2008). *Heterogeneous impacts in PROGRESA*. Discussion paper 3362. Bonn, Germany: Institute for the Study.

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

DesJardins, S. L., & McCall, B. P. (2008). *The impact of the Gates Millennium Scholars program on the retention, college finance- and work-related choices, and future educational aspirations of low-income minority students*. Unpublished Working Paper. Ann Arbor, MI: University of Michigan.

DiNardo, J., McCrary, J., & Sanbonmatsu, L. (2006). *Constructive proposals for dealing with attrition: An empirical example*. Retrieved from http://www.personal.umich.edu/~jdinardo/DMS_v9.pdf

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.

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.

Fletcher, J. M., & Tienda, M. (2009). High school classmates and college success. *Sociology of Education, 82*(4), 287–314.

Flick, U., von Kardoff, E., & Steinke, I. (Eds.). (2004). *A companion to qualitative research*. London, UK: Sage.

Fowler, H. S., Currie, S., Hébert, S., Kwakye, I., Ford, R., Hutchison, J., & Dobrer, S. (2009). *Future to discover: Interim impacts report*. Montreal, Quebec, Canada: Social Research and Demonstration Corporation. Retrieved from http://www.srdc.org/en_publication_details.asp?id=234

Frenette, M., Kwakye, I., Nicholson, C., Ford, R. , & Fowler, H. (2012). *Do Career Education and Financial Aid Early in High School Increase Access to Postsecondary Education?* Unpublished working paper. Ottawa, Canada: Social Research Development Corporation.

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

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students of diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis, 22*(3), 241–254.

Gennetian, L., Morris, P., Bos, J., & Bloom, H. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches*. New York, NY: Russell Sage Foundation.

Goldin, C., & Katz, L. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.

Goldrick-Rab, S., & Harris, D. (2010). *Memo to interested researchers using National Student Clearinghouse data.* Chicago, IL: Spencer Foundation. Retrieved from http://www.spencer.org/resources/content/3/documents/NSC-Dear-colleagues-letter.pdf

Goldrick-Rab, S., Harris, D., Benson, J., & Kelchen, R. (2011, March). *Can need-based financial aid increase college persistence among low-income university students? Evidence from a randomized trial*. Paper presented at the annual meeting of the Association for Education Finance and Policy, Seattle, WA.

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., & Pfeffer, F. (2009). Beyond access: Explaining social class differences in college transfer. *Sociology of Education, 82*(2), 101–125.

Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics, 5*, 263–275.

Harkreader, S., Hughes, J., Tozzi, M. H., & Vanlandingham, G. (2008). The impact of Florida's Bright Futures Scholarship Program on high school performance and college enrollment. *NASFAA Journal of Student Financial Aid, 38*(1), 1–16.

Harris, D. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record, 112*(4), 1163–1197.

Harris, D., & Goldrick-Rab, S. (2012a). *The (un)productivity of American higher education: From "cost disease" to cost-effectiveness* (WISCAPE Working Paper). Retrieved from Wisconsin Center for the Advancement of Postsecondary Education website: http://wiscape.wisc.edu/uploads/media/34a516f3-8ffe-49c1-8a4e-9e82fe7ff594.pdf.

Harris, D. & Goldrick-Rab, S. (2012b). 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., & LaFontaine, P. A. (2007). *The American high school graduation rate: Trends and levels* (NBER Working Paper No. 13670). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w13670

Hedges, L. V. (2009). *Problems with the design and implementation of randomized experiments*. Presentation at the 2008 Institute for Education Sciences Conference. Washington, DC.

Heller, D. (2006, October 13). To make college more affordable, award Pell grants earlier. *Chronicle of Higher Education*. p. B16.

Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects.* Unpublished manuscript, Harvard University, Cambridge, MA.

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. and Stuart, E. A. (2008). Misunderstandings among experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, 171*, 481–502.

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.

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. E. Lynn, Jr., & M. G. H. McGeary (Eds.), *Inner city poverty in the United States* (pp. 111–186). Washington, DC: National Academy Press.

Kahne, J. & Bailey, K. (1999). The role of social capital in youth development: The case of "I Have a Dream" programs. *Educational Evaluation and Policy Analysis, 21*(3), 321–343.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. Working Paper 14607. Cambridge, MA; National Bureau of Economic Research.

Kirst, M. W. (2004). The high school/college disconnect. *Educational Leadership, 62*(3). p.51–55.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Long, B. T. (2008). *The effectiveness of financial aid in improving college enrollment: Lessons for policy*. Unpublished manuscript, Harvard University, Cambridge, MA.

Long, B. T., & Riley, E. (2007). Financial aid: A broken bridge to college access? *Harvard Educational Review, 77*(1), 39–43.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review, 80*(2), 319–323.

Manski, C. (1993). "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" In C. Clotfelter and Rothschild, *Studies of Supply and Demand in Higher Education* (pp.43-57). University of Chicago Press.

Milwaukee Public Schools. (n.d.). *Administrative policy 7.37: Graduation and promotion requirements and learning targets by grade level.* Retrieved from Milwaukee Public Schools website: http://www2.milwaukee.k12.wi.us/governance/pages/policyManual/adminPolicies/CH07/7_37.pdf

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). Retrieved from the Western Michigan University website: http://www.wmich.edu/kpromise/documents/aspirations-6.pdf

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). Retrieved from the Western Michigan University website: http://www.wmich.edu/kpromise/documents/studentsurvey-3.pdf

Nagaoka, J., Roderick, M., & Coca, V. (2008). *Barriers to college attainment: Lessons from Chicago*. Chicago, IL: Consortium for Chicago Schools Research.

Orr, L., Feins, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., & Kling, J. R. (2003). *Moving to opportunity for Fair Housing Demonstration Program: Interim impacts evaluation*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Pressman, J. & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland; or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes*. University of California Press.

Public Agenda (2012). *With Their Whole Lives Ahead of Them*. New York.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials (NCEE 2009-0049). Retrieved from National Center for Education Evaluation and Regional Assistance website: http://ies.ed.gov/ncee/pdf/20090049.pdf

Raudenbush, S. (2011). *Modeling mediation: Causes, markers, and mechanisms*. Presentation to the Society for Research on Educational Effectiveness, Washington, DC, March 3, 2011.

Raudenbush, S., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29*(1), 5–29.

Richburg-Hayes, L., Brock, T., LeBlanc, A., Paxson, C., Rouse, C. E., & Barrow, L. (2009). *Rewarding persistence: Effects of a performance-based scholarship program for low-income parents.* New York, NY: MDRC.

Roderick, M., Nagaoka, J., & Allensworth (2006). *From high school to the future: A first look at Chicago public school graduates' college enrollment, college preparation, and graduation from four-year colleges*. Chicago, IL: Consortium for Chicago Schools Research.

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2009). *From high school to the future: Making hard work pay off: The road to college for students in CPS's academically advanced programs*. Chicago, IL: Consortium for Chicago Schools Research.

Rosenbaum, J. (2001). *Beyond college for all: Career paths for the forgotten half*. New York, NY: Russell Sage Foundation.

Schneider, B., & Stevenson, D. (1999). *The ambitious generation: America's teenagers, motivated but directionless*. New Haven, CT: Yale University Press.

Schochet, P.Z. (2008). *Guidelines for multiple testing in experimental evaluations of educational interventions*. Princeton, NJ: Mathematica Policy Research.

Sherwood, & Doolittle, F. (2003).What lies behind the impacts? Implementation research in the context of net impact studies. In *Policy into Action: Implementation Research and Welfare Reforms*, T. Corbett & M. C. Lennon (Eds.). Washington, DC: Urban Institute Press.

Sjoquist, D. L. & Winters, J. V. (2012). Building the stock of college-educated labor revisited. *The Journal of Human Resources, 47*(1), 270–285.

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.

Spielhagen, F. (2007). *Intersecting inequities: Examining the impact of early high school experiences on college admissions through the Texas Higher Education Opportunity Project.* (Working Paper). Retrieved from Higher Education Opportunity Project website: http://theop.princeton.edu/publications/wp.asp

Spybrook, J., Raudenbush, S., Congdon, R., & Martinez, A. (2009) *Optimal Design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Retrieved from http://www.wtgrantfoundation.org/resources/overview/research_tools

St. John, E. P. (2001). The impact of aid packages on educational choices: High tuition/high loan and educational opportunity. *Journal of Student Financial Aid, 31*(2), 35–54.

St. John, E., Paulsen, M., & Carter, D. (2005). Diversity, college costs, and postsecondary opportunity: An examination of the financial nexus between college choice and persistence for African Americans and Whites. *Journal of Higher Education, 76*(5), 545–569.

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed.). Chicago: University of Chicago Press.

U.S. Department of Education, National Center for Education Statistics. (2000). *National Education Longitudinal Study of 1988 (NELS:88/2000): Fourth follow-up, 2000*. Retrieved from http://nces.ed.gov/programs/coe/2003/section3/tables/t22_1a.asp

University of Wisconsin–Milwaukee, Center for Economic Development. (2010). *The economic state of Milwaukee, 1990–2008*. Retrieved from http://www4.uwm.edu/ced/publications/milwecon_2010.pdf

Yoshikawa, H., Kalil, A., Weisner, T., & Way, N. (2008). Mixing qualitative and quantitative research in developmental science: Uses and methodological choices. *Developmental Psychology*, *44*(2), 344–354.