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

An Improving Education Systems Proposal, Goal Three (Efficacy)

Project Narrative

Promise programs seek to increase college access by assuring students by 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 from high school to college, the theory is that these "early commitment" programs improve academic preparation and social capital during high school and thereby increase college success. 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 they are already off track.

This proposed Goal Three project will be 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. Specifically, we propose to study the Great Lakes in Milwaukee Promise Scholarship (GLIMPS)—a program externally supported with $10 million—to test the effectiveness of the promise program approach. The Milwaukee program seeks to provide students with a "GLIMPS" of their future so that they persevere through high school and into college. The funds requested here will allow us to study the program's effects on students' college affordability (real and perceived), academic preparation, and social capital through their high school years, as well as initial college entry.

Promise programs meet the IES Goal Three criterion of a fully developed intervention already in wide use (we count at least 73 currently in operation in the U.S.), and as we detail below, prior research suggests that they have a positive impact. The project proposed here will provide a rigorous test of the considerable potential of promise programs to help low-income students succeed in high school and college.

Significance

The Problem

Economic success in the United States has long been built on the education, skill, and ingenuity of its workers. Over the last two centuries, the U.S. gained an advantage over other countries by helping its citizens attain formal education, thereby developing an able workforce and promoting technological advancement. Technology, in turn, made skilled workers more productive, driving up their wages and leading even more people to obtain education (Goldin & Katz, 2008). Public policies fueled this growth in educational attainment by setting high school graduation as a universal goal and helping to make postsecondary education affordable for students and their families.

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). This can be traced to at least two interconnected factors. First, high school graduation rates have not increased for almost half a century and are now declining (Heckman & LaFontaine, 2007). Second, high school students do not see a clear path beyond to college. 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), shifting the burden of college costs to students and their families and making college investment a riskier proposition. 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). The larger consequence is that the U.S. is seeing more income inequality and falling behind the rest of the world in "the race between education and technology" (Goldin & Katz, 2008).

College unaffordability—perceived and real—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 (Kirst, 2004). 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 high school academic outcomes 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). The problem is that we have not found cost-effective ways of doing this (Bailey, 2011; Chingos, 2011; Harris & Goldrick-Rab, 2010). 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 have generally failed to produce large impacts, especially relative to their costs (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). This is a particular problem because the average person believes that the costs of college are nearly double the actual costs, and triple the 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 believe the barriers are even higher. Financial aid programs could provide a useful corrective to these misperceptions if the commitments and cost information came earlier.

In short, education leaders and policymakers need to build a road to college on which students can "drive" their strong college aspirations. Getting students onto that road requires persuading high schoolers early on that college is affordable—and making affordability a reality

by offering a substantial financial commitment. It also requires educating students about the concrete steps they must take to achieve college success.

This is the vision underlying promise programs. The expansion of such programs demonstrates that they are replicable and scalable: As noted, 73 programs are now in operation around the country, and more appear each year. The basic element of promise programs—the *early* commitment—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.[1] Moreover, promise programs have the advantage of being relatively inexpensive in comparison with the earlier generation of pre-college programs—the well-known TRIO programs, Upward Bound and Talent Search. A recent cost-effectiveness analysis conducted by PI Harris and colleague Goldrick-Rab (2010) showed that these and other commonly discussed avenues to increasing college access are inefficient and do little to help disadvantaged students. Promise programs, as we show later, are potentially better alternatives.

The GLIMPS Intervention

Overview

In October 2011, the GLIMPS program will promise students attending half of Milwaukee's 36 public ninth-grade schools $10,000 to pay for college.[2] The 18 GLIMPS schools will be randomly selected by the researchers on behalf of the program funder, the nonprofit Great Lakes Higher Education Corporation. The remaining 18 schools will serve as the control group. All ninth graders (one cohort) in each GLIMPS treatment school will be given the promise.[3] There is no explicit income requirement, but the low income level of the vast majority of 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 GLIMPS funds (through their aid package) so long as they graduate from any MPS high school on time (within 4 years of starting ninth grade), complete a Free Application for Federal Student Aid (FAFSA) each year of college, and attend an eligible college at least half-time. MPS has specific course work requirements for high school graduation (MPS, n.d., p. 2);[4] GEDs do not qualify. In addition, GLIMPS scholarships will require students to graduate with at least a 2.5 cumulative GPA (C+/B-) and attend class regularly.[5] The GPA and

---

[1] The financial aid application process itself is also a problem (Bettinger et al., 2009; Dynarski 2000, 2002; Cornwell, Mustard, and Sridhar 2006), but this is a somewhat separate issue.

[2] There are more than 36 schools in the city of Milwaukee, including charter and private schools. We are focusing on the 36 MPS and MPS-sponsored charters schools from which the district collects all the data discussed later and which serve ninth through twelfth grades. (Some of the 36 schools serve additional grades.)

[3] We also considered advising the program funder to randomize middle schools. However, we decided against this partly because the positive spillovers will be larger with high school randomization; there are no clear "feeder patterns" in MPS; eight grade recipients would have been quickly mixed across treatment conditions when they moved to high school.

[4] Under the district's standards, MPS ninth graders have to complete 4.0 units of English/ language arts and 3.0 units 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.

[5] We are working with MPS to identify an appropriate class attendance threshold. The GPA is similar to some other promise programs and refers to the overall GPA (not specific courses).

attendance requirements are cumulative across years, so that students who fall behind can catch up in later years.

The GLIMPS 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 (≥ 12 credits) and half this amount if they attend at least half-time (but less than full-time). Students need not start college immediately; 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 in the remaining 3 years. There are no GPA requirements during the college years.

Eligible students will be degree-seeking and have at least $1 of unmet need (this excludes loans and work-study and simply means that they must not already have their attendance cost met by the expected family contribution and existing grant and scholarship aid). 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 (see Table A2 in Appendix A for an illustration of how GLIMPS funding will be incorporated into students' financial aid packages). 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 statewide meeting these criteria.

The GLIMPS scholarship is "last dollar" and will cover up to the cost of attendance. Many of the GLIMPS students will have a zero expected family contribution (according to federal aid rules); for these students, the total GLIMPS 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—$2,886 annually at Milwaukee Area Technical College (MATC) and $8,152 annually at the University of Wisconsin–Milwaukee (UW–Milwaukee)—we see that GLIMPS by itself would cover all tuition and fees for a degree at MATC and more than one full year at UW–Milwaukee. As a point of comparison, the PI's own research shows that the average financial aid package among Wisconsin first-time college freshmen who are Pell Grant recipients is about $11,000—half in grants and half in loans and work-study (Goldrick-Rab, Harris, Benson, & Kelchen, 2011). Although tuition and fees are likely to rise before 2015, when GLIMPS recipients first enter college, the point we want to make is that $10,000 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 in the section, 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 GLIMPS.

Communication Plan

As an efficacy trial, this project seeks to demonstrate that the GLIMPS program *can work*, and as information is crucial, the program funder (under our advice) has put in place an

---

GLIMPS does not require students to take courses beyond the MPS graduation requirements. This is partly because course offerings beyond the required courses are uneven across MPS schools. As in other promise programs with GPA requirements, it is possible that that these requirements will lead to grade inflation or students taking easier courses. We will examine this possibility in our data analyses.

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).

To ensure that recipients are aware of the GLIMPS scholarship, Great Lakes and MPS will give letters about GLIMPS awards and program requirements (e.g., the 2.5 GPA) directly to students and also send them to the home addresses in the MPS data system. Parents will 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 GLIMPS. Schools will hold meetings or pep rallies with recipient students; high school counselors will be provided with lists of recipients and encouraged to remind promise recipients of their eligibility. Students will be directed to a public website where they can gather more information and ask questions. The district will also have a hotline for questions from students, school leaders, community members, and others.

Written reminders will be given to students and sent to parents three times per year, every year through their potential college years. Reminder letters will include individualized information about each student's progress in meeting GLIMPS class attendance and GPA requirements. These "on-track update" letters will also include typical high school course work of successful college students, average college costs and financial aid amounts in Wisconsin, colleges attended by MPS students in recent years, and the process for—and importance of—signing up for the ACT. Because one requirement is filling out the FAFSA, the program will also provide FAFSA information to students multiple times, especially as they begin their senior year. Research shows that FAFSA completion is a significant impediment to college entry (Bettinger et al., 2009). MPS collects information about language spoken at students' homes and translates and targets written communications accordingly (e.g., into Spanish); Great Lakes and MPS will also translate GLIMPS materials.

The program funder and operator, Great Lakes Higher Education Corporation, has decades of experience with college access programs. Over the last 6 years alone, Great Lakes has provided more than $58 million in funding and other support to programs seeking to increase college access for disadvantaged groups in Wisconsin, and especially MPS. Great Lakes has recently operated one-on-one financial aid advising, SAT/ACT test preparation, and FAFSA completion programs; it also provides funds for long-term personal mentoring; college tours; after-school tutoring; and college access partnerships with colleges and local, state, and federal governments. This experience will be invaluable in the successful administration of GLIMPS.

Comparison with Similar Programs

As noted, we have documented 73 promise programs currently operating across the country, with more launched each year.[6] At least 15 of the 73 have some sort of merit or performance requirement, based on (a) GPA (ranging from 2.0 to 3.5); (b) class attendance; (c) SAT/ACT scores; and (d) class rank/percentage. In terms of target population, requirements, and scholarship amounts, GLIMPS is most similar to district programs in Denver, Kalamazoo (Michigan), New Haven, and Pittsburgh, and statewide programs in Indiana, Oklahoma, and

[6] We thank Beth Vaade of the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE) for providing this information (see also Vaade, 2009).

Washington. Programs in Florida and Georgia are similar but the vast majority of funds go to middle- and high-income students. Promise programs are already in wide use and involve billions of dollars each year. The GLIMPS program and evaluation will therefore inform these existing large investments—and perhaps improve them over the long run.

Research Questions

Our research questions target four topics, which in turn inform our analytic framework and guide our data collection and analysis.

- *Topic A*. How does GLIMPS influence a range of high school outcomes, as well as initial college entry?
  - *A1:* How does GLIMPS influence perceived college affordability, academic preparation (e.g., test scores, attendance, grades, and graduation), and social capital in high school?
  - *A2:* How does GLIMPS influence college entry?
- *Topic B*. What are the mechanisms and mediators of the GLIMPS impacts?
  - *B1*: How do impacts (if any) on end-of-school academic outcomes arise? Do changes in perceived affordability, academic preparation, and social capital seem to play a role?
  - *B2*: Do changes in high school outcomes seem to increase college entry? What other factors emerge as potentially important?
- *Topic C.* How do impacts vary by student race, gender, parent income/education, prior academic ability, and availability/participation in other college access programs?
- *Topic D*. How is GLIMPS implemented and how does this affect its impacts?
  - What is the achieved relative strength of the intervention?
  - How do the impacts vary by school-level implementation?

Theory and Analytic Framework

In this section, we show how our research questions are rooted in prior theory and research. Figure 1 illustrates our analytic framework and illustrates (in bold) that the key focus is on Topic A: the impact of GLIMPS on high school outcomes. The remainder of the section discusses our basic theories in more detail.

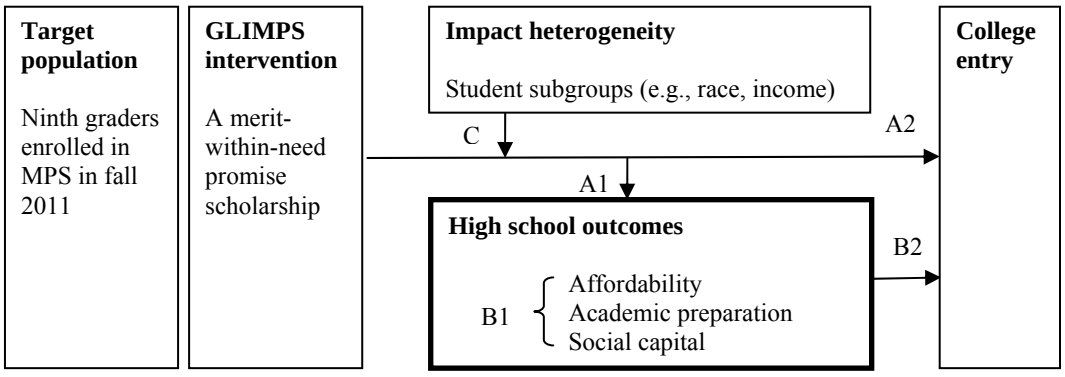

Figure 1. **Analytic framework.**

Basic economic theory suggests that promise programs, as well as other forms of financial aid, increase the likelihood of college-going 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). If they did,

students would already be going to college in greater numbers, motivated by the large and well-documented private returns to education (Ashenfelter, Harmon, & Oosterbeek, 1999; Card, 1999). The reality is that people do not make education decisions in ways predicted by basic economic theory (Manski, 1993; Beattie, 2002). Three problems have been identified as limiting college entry and completion by low-income students: (a) actual and perceived unaffordability, (b) weak academic preparation; and (c) low social capital (Nagaoka et al., 2008). We hypothesize that promise programs have the potential to address all three:

1. *Actual and perceived unaffordability.* The perception that college is unaffordable is a significant barrier to college attainment. As noted earlier, people tend to dramatically over-estimate the cost of college—even more so in African-American and low-income populations (Ikenberry & Hartle, 1998). The *financial nexus model* developed by Paulsen and St. John (1997) integrates perception of cost into students' decision making about college while also acknowledging interaction with other factors, including engagement and finances. In particular, the model posits a connection between students' financial reasons for choosing to attend a college and the ways they respond to net prices. The model is useful partly because, as a differential price response model, it recognizes variation in responses to money based on family background, for which there is much empirical support (Alon, 2007; Cabrera, Nora, & Castañeda, 1993; Dynarski, 2002; Goldrick-Rab et al., 2009; Grodsky & Jones, 2007; Harris & Goldrick-Rab, in press; St. John, Paulsen, & Carter, 2005). GLIMPS will address affordability by providing substantial funds and, through related communications, informing students about the realities of financial aid and college costs.
2. *Weak academic preparation.* A substantial literature indicates the importance of academic capital—specifically, preparation for college trajectories (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), and to meet prevailing norms of academic performance (Conley, 2008; Gamoran & Hannigan, 2000). 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, in press). GLIMPS will improve academic preparation with its GPA and class attendance requirements and by communicating the connection between high school academics and college goals.
3. *Low social capital.* Limited access to social capital (specifically, norms, information, and clear structures of support) is a key factor preventing students from reaching their college potential (Roderick et al., 2009). Social capital is closely intertwined with academic preparation and affordability. Students with more social capital are better informed and better able to navigate the many steps of the college application and entry process (Goldrick-Rab & Pfeffer, 2009). GLIMPS will communicate with principals, teachers, and counselors about establishing a college-going school culture so that students make use of scholarships.

We view the decision to go to college—and acquire academic and social capital—as the result of complex and interrelated processes, processes that can be influenced by promise programs. For example, one of our main hypotheses is that students' peers play an important role in college access (Spielhagen, 2007). PI Harris (2010) has described the complexity of these peer interactions and surveyed a range of theories to explain them. Building on prior work (Hoxby & Weingarth, 2005; Jencks & Mayer, 1990), he has argued that the evidence favors a theory of *group-based contagion*, according to which the most important student peers are those with whom students identify. Having peers with better outcomes and behavior in a group is generally

beneficial to the entire group (Fletcher & Tienda, 2009; Harris, 2010). Thus, if getting a "GLIMPS of their future" can directly influence individual students, then this could have large positive spillovers and feedback loops for peers.

We also hypothesize that the students most likely to benefit from GLIMPS are those on the margin of college access. A student who already has strong academic preparation, considerable college knowledge, and real and perceived financial means is likely to attend college with or without a promise scholarship. The opposite is true for students at the other end of the spectrum: because they have so far to go to be ready for college, no program will be sufficient by itself. It is the students *in the middle* who are most likely to benefit. We return to this idea later as it is central to our discussion of impact heterogeneity (see Topic C in Data Analysis section).

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 access.

Promise Programs: The Evidence

Promise programs have existed for at least three decades. Below, we discuss prior research pertaining to Topics A–C in Figure 1. Taken as a whole, the existing research suggests that promise programs like GLIMPS can indeed help students achieve in high school and in college.

Topic A1: Impact on High School Outcomes

At the center of Figure 1—and the center of our analysis—are the effects of GLIMPS on a range of high school outcomes. We discuss evidence about these below.

Lack of rigor in widely cited evidence. Most reports on promise programs have argued that they have great potential. 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). Teachers are increasingly using the Kalamazoo Promise as a long-range incentive in classroom interactions with students, motivating students to make cognitive and behavioral adjustments to achieve the identified goal of postsecondary education (Miron et al., 2009, p. ii). 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 increases in student achievement] can be attributed to the Promise versus other changes" (Bartik et al., 2010, p. 2).

The Florida Bright Futures Scholarship provides a somewhat more rigorous source of evidence. Research suggests that Bright Futures increased the percentage of students taking a college preparatory curriculum by almost 50%, and the largest effects were on racial minorities (Harkreader, Hughes, Tozzi, & Vanlandingham, 2008). Although a significant limitation of these analyses is that data are not available for the cohorts just *prior* to Bright Futures initial implementation, it is likely that the subsequent upward trend in rigorous high school preparation requirements was due at least in part to the program. However, Bright Futures is not need-based,

and the vast majority of benefits went to middle- and high-income students. Whether the program generated similar impacts on low-income students is unclear.

Most rigorous evidence: Future to Discover program. The Canadian Future to Discover (FTD) program (Fowler et al., 2009) 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, $8,000 would be placed in an account to cover college costs.[7] Researchers identified positive effects of the learning accounts during students' high school years on a wide range of outcomes, including future orientation, aspiring to earn more advanced postsecondary credentials, and being on track to finish high school on time.[8] The effects were particularly large for students whose parents lacked college degrees. Sufficient time has not yet passed to observe college enrollment.

Promise programs as conditional cash transfers. There has been great interest among researchers and policymakers in recent years in conditional cash transfers (CCTs)—that is, programs that provide financial incentives to people in exchange for specific behaviors and actions. College financial aid can be thought of as a form of CCT, in that the 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. He 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 and test scores] are not effective" (2010, n.p.). As we discuss in the Prior Research and GLIMPS Design section, we have used the Fryer results to inform and improve the GLIMPS design.

Topic A2: Impact on College Outcomes

There has never been a randomized trial or rigorous quasi-experiment of a promise program that has reported impacts on college outcomes (FTD researchers will have such results soon). However, there are rigorous studies of related programs that we believe reinforce our theory about promise programs and college access. Some programs start before college but lack income requirements; others target aid to low-income students but are not announced until after the students enter college. Nevertheless, the theory of action is similar in these related programs therefore we believe it is important that the evidence is consistently positive.

---

[7] The FTD really involved two programs and experiments, the learning accounts and a separate program that enhanced career planning and information services for some students. Here, we discuss only comparisons between the control group and those students who received *only the learning accounts* in order to more closely match the GLIMPS design. Students could meet the learning accounts requirements "by attending a New Brunswick high school and successfully completing each consecutive school year until graduation, and by successfully enrolling in a post-secondary education program" (Fowler et al., 2009, p. 4).

[8] The effects listed are statistically significant, except for on-time high school graduation. The weak statistical significance in this case appears to be due to both the very small samples and the analytic approach: the researchers reported results separately for groups, each with only 500 students; no average treatment effect was reported. Also, they did not report results with covariate adjustments, which would likely increase precision considerably.

Georgia HOPE is the only *promise-like* program with rigorous evidence on college outcomes.[9] Georgia HOPE is like GLIMPS in that it includes merit requirements and students are aware of it well before they have to make college decisions; it differs from GLIMPS in that it lacks an income requirement. Two prominent quasi-experimental studies have compared the outcomes of cohorts just prior to HOPE with those of the initial HOPE-eligible cohorts (Dynarski, 2000; Cornwell, Mustard, Sridhar, 2006). Although there were some differences in the types of students who benefited most, both studies found positive effects on college-going.

In MDRC's financial aid experiments, called Opening Doors (Brock & Richburg-Hayes, (2006; Richburg-Hayes et al., 2009; Barrow, Richburg-Hayes, Rouse, & Brock, 2010), participants were not made aware of the program until after they had started at a community college. We can therefore think of the results of these experiments as estimates of the effect of financial aid on students who have already made the decision to attend college. The analysis, across several sites and slight variations in program design, found positive effects on student persistence, credits, and grades. These findings are important because they suggest that GLIMPS might continue to benefit students even after they enter college.

The story changes somewhat when we turn to the Gates Millennium Scholars (GMS) Program, which, like GLIMPS, targeted low-income minorities (including African Americans). Participants in GMS, like those in Opening Doors and traditional aid programs, did not know whether they would receive the scholarship until the end of high school or later. Researchers used a quasi-experiment (specifically, a "fuzzy" regression discontinuity design) to determine program effects, but in contrast to Opening Doors, GMS did not show an impact on persistence (DesJardins & McCall, 2008). The likely reason is that GMS was limited to students who had a 3.3 GPA and the awareness and wherewithal to apply for the scholarship. That is, the entire sample of students was already on track for college attendance and success and therefore not on the margin where program benefits are likely to arise. As a result, 98% of non-selected GMS applicants (the control group) persisted from the freshman to the sophomore year—a much higher rate than that found recently among MPS graduates. It is nearly impossible to identify positive effects on groups that are already extremely successful to start with.

Two reviews of this and other evidence on financial aid, including but going beyond the evidence reviewed above,[10] concur that the effects of merit-based aid are positive and at least as large as those of purely need-based aid (Deming & Dynarski, 2009; Harris & Goldrick-Rab, 2010). Overall, it appears that merit-based financial aid can improve college outcomes, especially for disadvantaged students.

Topic B1: Mediating Mechanisms for High School Outcomes

Although prior research provides relatively little guidance on the mechanisms through which aid operates (Goldrick-Rab et al., 2009), our analytic framework can provide a starting point for thinking about the question. As we showed in Figure 1, promise programs might induce students to perceive greater affordability, prepare academically, and gain more of the social capital that can help them navigate the college process. These three factors are interconnected; for example, perceptions of college affordability might influence whether students identify with the college-going crowd which in turn might influence how they interact with teachers and who they choose as friends (an aspect of social capital). The FTD study suggests that promise scholarships might

[9] As noted in the discussion of Topic A1, the FTD has not yet considered college outcomes.

[10] We excluded from this review studies with quite different program designs and/or serving different populations, e.g., Angrist, Lang, & Oreopoulos (2009) and Scott-Clayton (in press).

induce students to be more oriented toward the future, which may translate into high school graduation, though they do not explicitly examine the issue (Fowler et al., 2009). We will explore the mediators of high school outcomes more explicitly.

Topic B2: Mediating Mechanisms for College Entry

Turning to the mediators of college outcomes, we can compare evidence on the effects of aid on high school outcomes (Topic A1) with the effects on college outcomes (Topic A2). If there are no impacts on a given high school outcome (e.g., social capital) then it cannot be a mediator. But, even if a high school outcome is affected by the intervention, this does not mean it mediates effects on college outcomes. For example, prior research suggests positive effects on high school outcomes (e.g., FTD and Florida Bright Futures), and these *appear* to translate into effects on college outcomes (e.g., Georgia HOPE). But we cannot say this is the real path without also taking a second step and showing that the mediator affects college outcomes. This second step is challenging even in an RCT because the participants are randomized to the intervention, not the mediator (Raudenbush, 2011). In general, trying to take on this challenge and identify mechanisms has not been a focus of prior financial aid research (Goldrick-Rab, Harris, & Trostel, 2009). We will attempt to address this void in our analysis.

Topic C: Impact Heterogeneity (by Student Subgroups)

As noted, we expect the effects of GLIMPS to be largest for students on the margin of college access. Most of the studies discussed above report financial aid effects by some type of subgroup. PI Harris and Goldrick-Rab (in press) conducted an extensive review of the variation in aid impacts by race, gender, family income/SES, ACT/GPA, and traditional/nontraditional student status. 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 be sure about this. No study has examined all of the listed dimensions of heterogeneity simultaneously, which is something we hope to rectify with GLIMPS.

Prior Research and GLIMPS Design

The evidence reviewed above lays the groundwork for the efficacy trial proposed here. Great Lakes, the program funder, committed $10 million in program funding for a scholarship where the primary requirement was college attendance. They sought our advice regarding details and, drawing on prior research, we advised (and they agreed to) the following:

- GLIMPS will start in ninth grade instead of tenth grade (FTD program) or twelfth grade and later (traditional aid, GMS, etc.), allowing more time to influence students;
- GLIMPS will select entire schools rather than individual students (FTD), yielding potentially positive spillovers across student peers;
- GLIMPS will include merit requirements, which may increase effects relative to purely need-based programs (Deming & Dynarski, 2009; Harris & Goldrick-Rab, 2010); and
- GLIMPS will include the type of input-oriented eligibility requirement suggested by Fryer (2010)—school attendance—and provide considerable information to students so that they better understand the steps, or inputs, needed for college success.

The design and evaluation of GLIMPS also reflects lessons learned from prior research conducted by PI Harris and colleague Goldrick-Rab (Goldrick-Rab, Harris, Benson, & Kelchen, 2011). The Wisconsin Scholars Longitudinal Study (WSLS) experiment involves a program that provides financial aid at random to low-income college students (similar to Opening Doors but without the merit requirement). Just as we propose to do with GLIMPS, WSLS studies short-term (proximal) and long-term (distal) outcomes, the relationship between these (mediating

mechanisms), and variation in impact across student types. Harris's experience with WSLS informed our recommendations to Great Lakes in two ways:

- Preliminary WSLS analyses found no average treatment impacts on academic outcomes such as grades, credits, or enrollment in a 4-year college.[11] Given the consistently positive average effects in other related programs (see above), this result supports our hypothesis that aid announced to students after they have started college comes too late—and thus reinforced the decision to have GLIMPS target ninth graders.[12]
- The finding of no average treatment effects reinforced our commitment to examine mediators and moderators and incorporate a qualitative component in the GLIMPS research. In WSLS surveys and interviews, we learned that (a) students lacked awareness of their eligibility for the grant and of financial aid in general; (b) many of the students gave grant money to their families, potentially diluting the academic benefits; and (c) the grant seemed to work only for students on the margin of college success.
- The lack of awareness about aid among WSLS students reinforced the need to pursue an aggressive communications strategy.

The PI's experience in carrying out the WSLS randomized trial also made it possible to design a GLIMPS evaluation that is nearly as extensive as WSLS at much lower research cost.[13]

Research Plan

Study Setting

We consider GLIMPS a need-based program partly in view of Milwaukee's extreme level of poverty and low educational outcomes. The rate of child poverty is 35%, and higher still among the city's large African American population (UW–Milwaukee Center for Economic Development, 2010); 81% of MPS students are eligible for free or reduced-price lunch.

The GLIMPS study population is similar to that targeted by the only prior randomized trial of a similar promise program. In FTD, 75% of participants had family incomes less than $40,000 (Fowler et al., 2009), similar to the $37,879 average income in Milwaukee in 2010.

Milwaukee's low economic status is linked to its weak educational outcomes. As noted earlier, only 67.8% of the 2002 cohort of ninth graders graduated from an MPS high school on time in 2006. Among those who did graduate on time, only 44.4% immediately went on to college—a much lower percentage than the national average (Aud et al., 2011).

Sample Selection

Random assignment of schools to the GLIMPS program will facilitate unbiased estimates of program effects. To ensure that such balance occurs, the researchers, on behalf of Great Lakes, will carry out the randomization.

Cluster randomized trials generally have low statistical power. To improve precision, we have recommended to the program funder a *paired randomization* approach: (a) rank schools by the percentage of college enrollment from recent cohorts, (b) create pairs of schools based on the rankings, (c) randomize one school in each pair to the treatment, and (d) pool to obtain the

[11] The aid did seem to get more students on track for college graduation, by taking more credits, though this was not enough to make the average treatment effects statistically significant.

[12] Opening Doors also found positive effects from a late commitment aid program, but this was targeted to a non-traditional college population. Other research suggests that non-traditional students benefit more from financial aid (Seftor & Turner, 2002).

[13] We have raised $3.3 million for the WSLS covering about 4 years of study. In contrast, we have budgeted just $1.9 million in this proposal to cover the same length of time.

overall control and treatment groups (Bloom, 2010). This approach will generally increase statistical power considerably, while maintaining internal validity (Greevy, Lu, Silber, & Rosenbaum, 2004; Imai, King, & Stuart, 2008; Imai, King, & Nall, 2009). As Bloom (2010) indicated, the "breakeven" R^2 for the pairing variable (that is, the R^2 necessary to make paired randomization worthwhile) only has to exceed 0.07 to improve precision with 36 clusters. In this case, because we are using a lagged value of a key dependent variable, the R^2 will be much higher than 0.07. In MPS, we estimate a correlation of 0.72 between 2010 college entry rates (by school) and college entry rates averaged across the 2005 and 2006 ninth grade cohorts. This implies an R^2 of 0.49. The broader statistical literature also supports paired randomization (Greevy, Lu, Silber, & Rosenbaum, 2004; Imai, King, & Stuart, 2008; Imai, King, & Nall, 2009).

Through this randomization process, half (18) of the 36 MPS schools will be assigned to treatment and the remainder to the control group.[14] All students enrolled as ninth graders in selected schools on the official "count day" (the third Friday of September) will be eligible for GLIMPS. Students will remain in the same treatment condition regardless of whether they switch schools after count day (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. Based on current middle school enrollments and recent eighth-to-ninth-grade transitions, we estimate that 4,590 ninth graders will constitute the study sample.[15]

Data Collection

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

Administrative Data

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

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

[14] We will check for baseline equivalence.

[15] While we are studying only one cohort of ninth graders, we continue referring to "school-level" randomization.

data system for the GLIMPS treatment status of each student, and we will cross-check this indicator against attendance data.

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

The MPS climate surveys are based on the well-respected and research-based surveys developed and administered in the Chicago Public Schools in conjunction with the Consortium on Chicago School Research (the CCSR senior director is a consultant on this project; see Personnel section). One key construct of interest is students' peers and their interactions, as a form of social capital. The student climate survey includes items for "Students at my school focus on learning" and "Most of the students in my school are planning to go to college." Regarding academic preparation, we are also interested in how student interactions with teachers; sample items for this include "At school I am expected to do my best all the time" and "Teachers at my school expect most students to go to college." Student responses to these questions have been shown to be strong predictors of college-going (e.g., Nagaoka et al., 2008).

The high school exit survey includes questions about (a) perceptions of the expectations of, and support received from, parents, teachers, and counselors; (b) college plans, including planned course of study; (c) participation in other college access programs; and (d) parents' educational background. We are particularly interested in the items about another source of social capital: parents. Sample items include "my parents encouraged me to continue my education after high school" and "my parents talked to me about colleges/schools suited to my interests and abilities." MPS has agreed to allow us to add some questions to their surveys for this project; among these will be questions about parental expectations and our third mediating construct: perceptions about financial aid and college affordability.

MPS teacher surveys (high school social capital). One way in which GLIMPS might influence students, as noted above, is by changing interactions among students and teachers. The teacher climate survey and instructional practice survey will complement the student surveys in gauging these changes. MPS links teacher survey responses to teachers' unique identifiers.

The pertinent constructs of the instructional practice surveys are (a) teacher demographics, (b) expectations for students, (c) orientation toward higher order thinking skills, and (d) student engagement. Key items include "Required students to interpret, analyze & evaluate information in their work" and "How much emphasis do you give [to] . . . tracking student progress toward expected outcomes?" The teacher climate survey includes measures of academic rigor and expectations for students' college education. Sample items include "Students at my school focus on learning" and "Teachers in this school expect most students to go to college."

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

With minimal district encouragement and no formal incentives, survey response rates in recent years have been 40% among students and 51% among teachers. MPS has agreed to take additional steps to increase these rates. In another research project conducted in MPS, researchers obtained at least 70% response rates from staff in every school (student surveys were

not relevant to this other project) with incentives of just $200 per school. Because student surveys are administered during the school day, student response rates are, according to MPS staff, heavily driven by whether teachers obtain computer lab time for students to fill out the survey. Therefore, to encourage high response rates, we will offer school staff incentives, tying them both to student and staff survey response rates. Schools in which student and teacher response rates exceed 70% will receive $500 to be used for any purpose. Also, MPS has agreed to send school principals regular updates of their respective survey response rates and to remind them in weekly communications about the $500 incentive and the importance of the surveys.

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

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

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

Other Data

The above data sources are de-identified and rely on unique identifying numbers to track individual students over time; therefore, they do not require student or parental consent. Other forms of data collection will require consent from roughly 900 students prior to treatment randomization. In August 2011, we will send consent forms by regular mail to a subsample of 1,800 students, prior to announcing the treatment group schools. Specifically, we will solicit consent from 50 students per school (on average) with an expected consent rate of 50%, or 25 students per school. In a cluster randomized design such as ours, the statistical return to a larger cluster size is small, making this is an efficient sampling design. It is also one with sufficient statistical power for analyzing financial aid packages. The treatment here is a scholarship, so the impact on students' aid packages will clearly be large (Goldrick-Rab, Harris, Benson, &

Kelchen, 2011). Obtaining consent prior to randomization will help ensure equivalence of the control and treatment groups.

The consent form will be accompanied by a short survey requesting contact and other information, which will allow us to track students over time and supplement the MPS surveys (we will add some questions to the MPS-administered surveys but this additional survey extends our capabilities). A $5 bill will be included as an incentive to complete the survey. Small pre-incentives are highly effective in encouraging participation, especially among low-income respondents (Berlin et al., 1992; Church, 1993; Ryu, Couper, & Marans, 2006; Hopkins & Gullikson, 1992; James & Bolstein, 1992; Trussell & Lavrakas, 2004). The consent forms and surveys will be sent directly to students' homes using addresses in the MPS system.

Student interviews (social capital and program implementation). 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 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 GLIMPS Design section)

From the overall consenting sample, we will take a stratified random sample of 6 schools (three control, three treatment) and, within those, select a stratified sub-sample of students for additional qualitative data collection. We will block-randomize on gender and individual predicted probability of college entry, then select four students from each school to be interviewed annually (a total of 24 students and 120 interviews through the end of the requested research grant period). The goal is to collect interview data that can serve to both validate and explain findings based on the administrative and survey data.

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.

Key topics in the interviews will be:

- How their parents, teachers, and peers have affected their thinking about college
- Beliefs about the importance of college and required steps
- Perceptions about college costs and understanding of college financial aid
- Knowledge of their own GLIMPS status and that of their friends and classmates
- Understanding and impressions of GLIMPS
- Initial reactions to GLIMPS
- Knowledge of school communication and messages about GLIMPS
- Beliefs about whether and how GLIMPS affected their beliefs, attitudes, and behaviors

FAFSA and state data on college financial aid packages. FAFSA completion is a requirement for receiving GLIMPS funds, so one obvious issue is whether the program increases completion. 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, in press). We will estimate impacts by income subgroup using (de-identified) data from the FAFSA, provided by the Wisconsin Higher Educational Aids Board, the state agency helping to administer GLIMPS.

To understand the effects of the scholarship after students enter college, it will be essential to know exactly how students received it. The PI, in his role at co-director of the WSLS, has obtained financial aid data from the University of Wisconsin System Administration (UWSA)

and the Wisconsin Technical College System (WTCS) for consenting students. The vast majority of MPS students who go on to college are included in one of these two data systems.

Table A2 in Appendix A shows some illustrative aid packages. These highlight the fact that students will receive some of the money as loan reduction and some as cash. Together with the student interviews, the financial aid package data will contribute a great deal to our understanding of program implementation. First, they will provide a sense of the achieved relative strength (Cordray & Pion, 2006) of the intervention—that is, the degree to which the program was implemented as intended—and a better understanding of what communication students received and how they interpreted it.

Other long-term data collection. While beyond the scope of this proposal, we plan to collect more detailed data on students' academic outcomes from college transcripts and longer term data about students' workforce participation and other outcomes. The initial consent of 900 students (see above) will enable us to obtain data up through the summer in which high school graduation is expected (2015). Thus, for most students who attend college, the period covered by the initial consent will include their college financial aid packages, which are usually sent out at the end of high school and following summer, but exclude college transcripts or other subsequent college data. We will obtain additional funding to re-consent students at the very end of high school and continue long-term data collection efforts.

Data Analysis

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

Topic A1: Impact on High School Outcomes

To address our first research question, we will analyze MPS/DPI student record and survey data. While differences in means between control and treatment provide unbiased estimates of impacts, we have the advantage of longitudinal data and can therefore include lagged values of the dependent variables as explanatory variables in most analyses. In general, including lagged values as covariates improves precision (Raudenbush, Martinez, & Spybrook, 2007). For this reason, the minimum detectable effects (MDEs) for these analyses range from 0.14 (standard deviations) for ninth-grade student academic outcomes to 0.17 for the high school exit surveys. Power is naturally lessened in teacher surveys, with an estimated MDE of 0.28. For high school graduation and college entry, we can identify (dichotomous) effects of 5 and 7 percentage points, respectively. Overall, the study has strong statistical power (see Power Analysis section).

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

Level 1: $Y_{ist} = \alpha_0 + \alpha_1 Y_{is,8th} + \alpha_2 X_{it} + \varphi_s + \varepsilon_{ist}$ (1)

where high school outcome measure Y of student i in school s at time t is a function of the same lagged dependent variable ("pretest") from eighth grade, X_{it} is a vector of other covariates (e.g., student demographics), and φ_s represents a vector of school effects.[16] The school effects from Level 1 then enter as the dependent variable in Level 2:

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

where κ_p is a vector of pair indicators[17] with p=1,2,…,18 (i.e., indicating the pair used in the randomization). The variable T_s represents the treatment status of the schools so that the scalar β_3 is the average treatment effect. (As this is the intent-to-treat (ITT) analysis, students will be assigned to their original school in the coding of T_s.) This is similar to the approach used by Kane and Staiger (2008) who randomized pairs of teachers and then estimated a reduced form version of equations (1) and (2) with many covariates and indicators for each pair of teachers.

The above approach will be applied to most of the data sources for high school outcomes discussed earlier, including student surveys, student academic records, and financial aid packages. With slight adjustments for the unit of analysis, this also extends to teacher surveys.

Differential nonresponse is the primary threat to validity in the analysis of high school outcomes. In particular, if GLIMPS influences the high school dropout rate—which is one of the key objectives and a focus of our analysis—then this will create differential missing data between control and treatment groups, a potential serious threat to internal validity. We will therefore test whether the difference between exited and retained students (using eighth grade and other prior data) is the same in the control and treatment groups. In the remaining discussion, we assume the worst case scenario that such differentials will exist, requiring additional steps.

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

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

[16] Ordinary least squares estimation of Equation (1) is potentially biased with a lagged dependent variable. Therefore, we will also estimate these models with the lagged value on the left-hand side under the assumption that $\beta = 1$. We can also estimate (1) via instrumental variables estimates using further lags and/or other related survey measures.

[17] 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). The same logic applies to the school effects. We are interested in both inferences and will therefore use both approaches.

also the Personnel section.) These data will provide a valid test of whether the differential response rates introduced bias into the impact estimates, and the direction and size of any bias.

Topic A2: Impact on College Outcomes

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

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

Topics B1 and B2: Mediating Mechanisms for High School and College Outcomes

Students are randomly assigned to the treatment but not to the mediators. Identifying the causal effects of these mechanisms therefore requires strong assumptions (Raudenbush, 2011). Some have proposed using instrumental variables to solve this problem, but this approach also requires restrictive assumptions (Gennetian et al., 2005). Consequently, we view this aspect of our work as exploratory. We will start by applying a variety of quasi-experimental methods.

To make this concrete, recall Figure 1, which highlights our hypothesis that real and perceived affordability, academic preparation, and social capital are interconnected. With longitudinal data collection, we can examine time patterns in the data. For example, students might begin taking more rigorous courses in their freshman year, then in the sophomore year we might see a change in student reports about their peer interactions, suggesting the new courses influenced who students interacted with.

We will also use quasi-experimental methods to examine mediating mechanisms. For example, it may turn out that some schools were more aggressive than others in advertising GLIMPS. In that case, we could compare the measured impacts on GLIMPS knowledge (from the student surveys) from schools that were less active in promoting GLIMPS. This is of course non-experimental and might be driven by unobserved differences in the schools (and the school-level randomization severely limits power). We will therefore also consider propensity score matching as an alternative quasi-experimental method (Rosenbaum & Rubin, 1983).

We emphasize that all of these approaches are quasi- or non-experimental, but such exploratory analysis are worthwhile. If we are going to obtain highly valid estimates of impacts, then it is incumbent on us to try and understand how those effects arose (Harris & Goldrick-Rab, in press). Combining the above quantitative analyses with the student surveys will allow us to triangulate the analysis, and perhaps develop plausible theories about what happened and why.

Topic C: Impact Heterogeneity (by Student Subgroups)

Based on prior evidence (Harris & Goldrick-Rab, in press), there is some reason to expect heterogeneous impacts in gender, race, parent income, and students' prior academic ability. The most straightforward way to test for these is adding interaction terms such that the interaction between the treatment and the given subgroup characteristic provides a direct test of heterogeneity. It is possible, however, that heterogeneity will be based not on one particular measureable characteristic, but on a collection of them—or, really, the overall likelihood of

college entry. We will therefore regress the probability of college entry on the above factors (gender, race, eighth grade academic performance, etc.) to obtain each student's expected college entry probability. Next, we will divide students into two or more groups based on these estimated probabilities and test the impacts separately by group.[18]

In preliminary analyses in the WSLS, we have found that students in the middle third on these college persistence propensities were the ones most likely to benefit from the grant. The groups benefiting most in GLIMPS may or may not be the same, given the different population and program design, and our limited focus on college entry rather than persistence.

Topic D: Program Implementation and Strength of Intervention

We will begin the analysis of program implementation using the student interviews (see above bulleted list of specific topics). As interview transcripts become available, the coding team will develop schemes that reflect both core research questions and emergent constructs. They will first read through a sample of interviews to decide on coding categories and then jointly code a second sample to ensure that the categories are comprehensive and valid. Each interview will be coded by two coders, with a third assigned if there is disagreement. At the end of the coding cycle, the team will meet to conduct spot-checks on a sample of interviews. This cross-sectional coding will allow for preparation of the data set for analysis and ensure that team members have immersed themselves in the sample of interviews and identified sets of themes and questions that need to be explored more thoroughly (Flick, von Kardoff, & Steinke, 2004). Codes will be entered into either NVivo or Dedoose software along with the survey and administrative data. Integration of all data into a single interactive database will facilitate an ongoing and iterative mixed-methods analysis that will ensure a constant and informative feedback loop between the quantitative and qualitative data.

In the surveys, we will complement the student interviews with questions about whether they have received a "promise of a scholarship," whether they have heard of GLIMPS, and if so, whether they are aware of the program's requirements. These measures will provide descriptive data on students' level of awareness and will allow us to test whether awareness is greater in schools with more GLIMPS-related school activities. Interviewers will have the individual student survey responses with them at the interviews to help guide the conversation.

To further improve our understanding of implementation, the district will collect information about the number of special events schools hosted, the frequency of school announcements, and other GLIMPS-related school activities designed to generate greater awareness among students.

Threats to Validity

Internal Validity

We addressed threats to internal validity at length in the discussion of impacts on high school outcomes in the Data Analysis section above.

External Validity

Our sample is one Midwestern school district and the results may or may not generalize to other districts, though we believe that the results will be of considerable interest to promise

[18] 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). While randomization yields unbiased impact estimates, this balancing approach can help improve precision especially in heterogeneity analyses where subgroups are typically small (and, in cases with missing data, also help ensure internal validity).

program leaders in places like Pittsburgh and Kalamazoo that have similar populations and have already adopted promise programs. Also, when promise programs are adopted at scale, all students in the given unit (e.g., a whole district) are eligible for the money. This creates the possibility of positive spillovers, e.g., peer effects and raised expectations among teachers. One of the advantages of the GLIMPS demonstration over the Canadian FTD is the use of school-level randomization, more closely mimicking the potential spillovers of a scaled up program.

Power Analysis

The statistical power of an analysis is the probability of rejecting the null hypothesis (in this case, no GLIMPS scholarship 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. Below, we apply this value to power analysis implemented with software by Spybrook, Raudenbush, Congdon, and Martinez (2009). Table 1 summarizes these calculations.

Hedges and Hedberg (2007) reported intraclass correlation coefficients (ICCs) of 0.05 for ninth grade achievement and that lagged (pretest) values for dependent variables can explain up to 75% of the variation in the dependent variable. While this might seem optimistic, recall that (a) the large number of lagged covariates and (b) the paired randomization is based on a lagged value of a key dependent variable (college entry in prior cohorts), which will also improve precision considerably (especially for college entry, but also other outcome analyses). We assume an R^2 of 0.75 for all the surveys and academic outcomes such as test scores (most of which are also continuous variables).

Table 1

Power Analysis Summary

Sample	*N* (# clusters/cluster size)	MDE	
		Continuous var. (effect size)	Dichotomous var. (perc. points)
Student acad. outcomes (9th grade)	4,590 (36/128)	0.14	—
Student climate surveys (9th grade)	3,213 (36/89)	0.15	—
Student HS exit survey (12th grade)	2,184 (36/61)	0.17	—
Teacher surveys (per HS grade)	500 (36/14)	0.28	—
HS graduation	4,590 (36/128)	—	5
College enrollment	4,590 (36/128)	—	7

Note: HS = high school. MDE = minimum detectable effect. Var. = variable. Acad. = academic. The first two rows focus only on ninth grade for simplicity, though we will have data for Grades 10–12 also.

Based on our expected survey response rate of 70%, our student sample will include 3,213 students in ninth grade. For the high school exit survey, we multiply the ninth-grade figure by the share of MPS ninth graders who graduate on time (0.68), yielding 2,184 student observations.

Unlike the analysis of high school survey and academic measures, the analysis of high school graduation and college enrollment cannot rely on covariate adjustments using lagged dependent variables, as there cannot be any "prior" values for graduation and enrollment variables. Thus, we will increase power, first, using paired randomization based on college outcomes of recent cohorts in the same schools (Bloom, 2010; Greevy, Lu, Silber, & Rosenbaum, 2004; Imai, King, & Stuart, 2008; Imai, King, & Nall, 2009). While we are primarily interested in high school outcomes for this proposal, the longer term goal is to understand college success; therefore, we believe using the college entry of prior cohorts as the pairing variable is the most sensible approach. As discussed in the Sample Selection section, college entry averaged across the 2005 2006 ninth grade cohorts was correlated with 2010 college entry at 0.72. Covariate adjustments

in both the Level 1 and Level 2 equations further improve precision of high school graduation and college entry. MPS research staff provided us with statistics indicating that a regression of initial college entry on eighth grade test scores, course grades, attendance, race/ethnicity, and free or reduced price lunch eligibility yield on R^2 of about 0.25. The combination of paired randomization and covariate adjustment complicates the power analysis, but we estimate MDEs of 5 percentage points for high school graduation and 7 percentage points for college entry.

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 that are justified based on prior evidence and theory and (b) adjusting the *p*-values using approaches such as the Bonferroni correction (though we note that these tests are generally too conservative and overcorrect for the problem). We will report both the unadjusted and the adjusted *p*-values. If the results are affected by the adjustments, we will note this and draw conclusions based on the more conservative tests to avoid the possibility of concluding that the program worked when in fact it did not.

Timeline and Other Funding

GLIMPS will begin in October 2011, one year before this IES grant would start. This early launch is a strength of this proposal because the program will likely already have been implemented before this proposal is considered. Randomized trials often never get off the ground because, while superintendents often sign letters of support, they may not be able to follow through when it comes time to randomize. In contrast, GLIMPS offers negligible risk of failure.

Table 2
Timeline

				IES Funding											
	2011–12 (freshman)			**2012–13 (sophomore)**			**2013–14 (junior)**			**2014–15 (senior)**			**2015–16 (college)**		
Research activities	**F**	**Sp**	**Su**	**F**	**Sp**	**Su**	**F**	**Sp**	**Su**	**F**	**Sp**	**Su**	**F**	**Sp**	**Su**
Obtain consent/assent	■														
Assign schools to treatment	■														
Survey data															
Collect, clean, & code older data			■												
MPS administers climate survey		■			■			■			■			■	
MPS administers IPS survey		■			■			■			■			■	
MPS administers h.s. exit survey											■				
MPS student records															
Collect, clean, & code older data			■												
Obtain new records		■			■			■			■			■	
Conduct student interviews		■			■			■			■			■	
Collect NSC & financial aid data													■		■

Note. F = fall semester. Sp = spring semester. Su = summer. IP = instructional practice. NSC = National Student Clearinghouse.

We have secured a small amount of funding from a foundation (that wishes to remain anonymous) to cover costs of the first-year consent process, student interviews, and some data cleaning and coding. The IES funds will enable us to report impacts on all high school outcomes

and college entry. If it is effective in improving college access, the program evaluation will continue many years beyond the end of the grant period Table 2 summarizes the project timeline.

Products and Dissemination

We anticipate producing at least six journal articles and a book. Findings will also be disseminated through presentations at national, state, and local levels, including at the annual meetings of the American Economic Association, the Association for Education Finance and Policy, the American Educational Research Association, and the National Bureau of Economic Research (PI Harris has presented previously at each of these). When papers are ready for publication, they will be submitted to *American Economic Review*, the *Journal of Human Resources*, and *Educational Evaluation and Policy Analysis*, among other journals.

Harris is already involved in various networks of promise program administrators and researchers. He has organized a panel of researchers, including the lead FTD experiment researcher and two scholars who have studied the Kalamazoo Promise for the annual meeting of the Association for Public Policy Analysis and Management in November 2011. Also, an informal group of promise program operators and researchers, known as PromiseNet, meet each year to discuss their work. The next conference will be held in October 2011 in Pittsburgh, and a team from GLIMPS (including Harris) will attend.

We plan to publish not only research papers, but also working papers and shorter policy briefs that summarize the implications of the research for the media and policymakers. Both policy briefs and working papers will be distributed electronically to key audiences through various UW research centers. In our prior work, results have been reported by national education and other media, such as *Education Week*, *Inside Higher Ed*, and the *New York Times*.

Personnel

Douglas N. Harris (30% time each year)**,** associate professor of educational policy and public affairs at the University of Wisconsin–Madison (UW–Madison), is the principal investigator of this project and will assume overall responsibility for the evaluation. He studies the efficiency and equity of K–12 and higher education programs, using experimental, quasi-experimental, and mixed methods. Since 2008, he has served as co-director (with Sara Goldrick-Rab) of the Wisconsin Scholars Longitudinal Study (WSLS), a randomized trial of financial aid that has received more than $3 million in support, operates at 42 colleges, and is similar to GLIMPS in research design and scale. Harris has also published a number of studies on the cost-effectiveness of K–12 and higher education programs that have informed the design and rationale of GLIMPS. He has experience working in schools and school districts, which will be important for the first four years of GLIMPS and interactions with MPS. In addition to being a former school board member, he has conducted quantitative and qualitative research in schools. His study of teacher effectiveness, funded by IES, included methodological advancements and applications of value-added statistical techniques (results published in the *Journal of Public Economics*), as well as interviews with school principals (results published in a range of educational journals). Harris is the author of the recent book *Value-Added Measures in Education* (Harvard Education Press), and his other research has been published in books and in journals such as *Brookings Papers on Education Policy*, *Education Finance and Policy*, and *Journal of Policy Analysis and Management*. He has been PI or co-PI on research grants totaling more than $6 million from eight different funders.

Elaine Allensworth (5% time over four years) is the Senior Director and Chief Research Officer of the Consortium on Chicago Schools Research (CCSR). Dr. Allensworth is nationally recognized for her work on early indicators of high school graduation and the transition from

middle to high school, and has published widely on the structural factors that affect school improvement. She is an expert in methodology concerning studying graduation and dropout rates, and using longitudinal data systems to measure student growth and school improvement. The Chicago Public Schools (CPS) has been very active in implementing new college access programs in recent years—Elaine and other researchers at the CCSR have been lead evaluators. She has conducted research about on-track indicators and factors for high school graduation (Allensworth & Easton, 2005, 2007) and on the college success of former CPS students (Allensworth, 2006; Roderick, Nagaoka, & Allensworth, 2006). She is frequently invited to discuss her research with national, local, and school district audiences and has been PI on grants from IES, the National Science Foundation, Gates Foundation, Carnegie Corporation of New York, and other foundations. Dr. Allensworth received her Ph.D. in Sociology from Michigan State University, broadening the disciplinary base of the research team. In GLIMPS, Elaine will focus particularly on formulating the roll out and study of the initial implementation of the program, as well as participating in some data analysis and report writing.

Larry Orr (5% time over four years) will serve as lead consultant on research methods and advise the project on the research design and analysis. He has 40 years' experience in the analysis of public programs and policy, specializing in the design and implementation of large-scale field projects involving random assignment. He has served as project director or PI on projects such as the National Job Training Partnership Act Study, the Lifelong Learning Demonstration, and the evaluation of the Moving to Opportunity (MTO) experiment. As a leading expert on rigorous evaluation methods, Orr has served on expert advisory panels for several evaluations in education, including an evaluation of charter schools, the evaluation of DC Choice, a project to design a rigorous evaluation of interventions to turn around chronically low-performing schools, and the Congressionally mandated evaluation of IES. He has authored, coauthored, or edited six books, including *Social Experiments: Evaluating Public Programs with Experimental Methods*. He was a member of the editorial board of *Evaluation and Program Planning* and served on the National Academy of Sciences Committee on Postsecondary Education and Training. He currently teaches program evaluation at Johns Hopkins University.

Bradley Carl is a sociologist and researcher at UW-Madison's Wisconsin Center for Education Research. He serves as an embedded researcher for the Milwaukee Public Schools (MPS), conducting program evaluations and research involving key MPS initiatives and district improvement efforts. Examples include evaluating MPS high schools, developing "early warning" systems system to identify students at risk of failing to graduate from high school and/or graduating with low levels of college and workforce readiness, developing a postsecondary tracking system for MPS graduates. Dr. Carl holds a Ph.D. in Sociology and Urban Studies from Michigan State University. He worked previously for the American Institutes for Research in Washington, D.C. and the Office of Educational Accountability at the Wisconsin Department of Public Instruction.

The project also has two advisory boards (see Table A1 in Appendix A). The Research Advisory Board includes a distinguished group of external scholars who will provide expert counsel on the analyses: Rodney Andrews, Joshua Angrist, Larry Hedges, and Bridget Long. The Program Advisory Board includes Wisconsin K–12 and higher education leaders with expertise in college access and knowledge of the local context. The PI will also consult with the leaders of the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE), including its well known respected senior scholar, Sara Goldrick-Rab.

Resources

The GLIMPS program is a partnership, developed over two years, between Great Lakes and MPS, under the advice of the PI (Harris) and others. Because of this prior work, the program is carefully planned, roles and responsibilities are clearly defined, and the working partnerships are well established. The researchers will be responsible for the randomization and evaluation. Great Lakes and MPS are responsible for program administration, including communications. MPS will provide data to the researchers and facilitate collection of additional data that the district does not already have. Each partner is well situated to carry its responsibilities.

The research will be housed with the UW–Madison Wisconsin Center for Education Research, one of the nation's oldest and largest university-based education research and development centers with annual outside funding exceeding $25 million. Great Lakes is one of the most experienced organizations in the nation in administering college access programs; Great Lakes staff understand the college access problem for low-income students and know Milwaukee well. Finally, MPS has an excellent data system and experienced staff accustomed to engaging with, and providing data for, research projects. District leaders have made college access a high priority and view GLIMPS as fitting well within the district's strategic plan. Support for the project within these organizations runs deep, involving not just a few individuals, but whole teams of upper and middle management. (See letters of support in Appendix C).

In addition to developing strong partnerships, we have taken care to follow the IES Request for Applications (RFA) recommendations for Goal Three proposals that there will provide "reasonable safeguards … to ensure the objectivity and integrity of the evaluation" (p.56). Specifically, (1) "The procedure for assignment of units to condition is conducted by an individual (or team) who is independent of the developer" (p.56). Consistent with this, the researchers are conducting the randomization on behalf of the program funder. (2) "Analysis of data is conducted by individuals who are not involved with the development or distribution of the intervention and have no financial interest in the outcomes" (p.56). The researchers will be the sole evaluators and are seeking third-party funds to support the evaluation work.[19]

As a final note, we also considered submitting this proposal under the Postsecondary section, but, in consultation with IES staff, decided that Improving Education Systems was a better fit because the overwhelming majority of the outcomes we analyze will be at the high school level. The only college outcome we study is initial college entry, so we will have to seek additional funds in the future to study college credits, grades, and graduation. Early commitment programs are more likely to work because they start early, but this also means studying them across K-12 and higher education sectors over a long period of time—probably at least a decade in this case—to learn whether and how they really work. The longer-term analysis will be better suited to the IES Postsecondary topic.

[19] The RFA also recommends "Collection and coding of outcome data should be under the supervision of someone other than those who were or are involved in the development or distribution of the intervention." We are relying primarily on administrative data, which are collected by MPS. But we view this as a small sacrifice given the quality of the instruments and the very low cost of obtaining these data for research purposes.