The Impact of the Gates Millennium Scholars Program on Selected Outcomes of

Low-Income Minority Students: A Regression Discontinuity Analysis

Stephen L. DesJardins

Center for the Study of Higher and Postsecondary Education

University of Michigan

and

Brian P. McCall

Carlson School of Management

University of Minnesota

October 2006

Abstract

The Gates Millennium Scholars (GMS) program, funded by the Bill & Melinda Gates Foundation, was established in 1999 to improve access to and success in higher education for low-income and high-achieving minority students by providing them with full tuition scholarships and other types of support. The effects of programs such as the GMS are often difficult to assess, however, because of the non-random nature of selection into the program. This article uses regression discontinuity to ascertain the causal effect of the scholarship on a number of important educational outcomes.

Keywords: Regression Discontinuity, Local Polynomial Smoothing, College Retention, Loan Debt Accumulation

I. Introduction

Social scientists often evaluate programs into which students are not randomly assigned, and the process by which students become program participants may make valid inferences about program effects difficult. One such program is the Gates Millennium Scholars (henceforth GMS) program, funded by the Bill & Melinda Gates Foundation and administered by the United Negro College Fund. Established in 1999, the goal of the GMS is to establish a cadre of future leaders by improving access to and success in higher education for low-income, high-achieving minority students in the United States by providing them with scholarships and other forms of support. Students are eligible for the GMS scholarship if they meet pre-specified and multiple criteria set by the program. Given the selection mechanism, GMS program participants may be systematically different than non-participants, making it difficult to determine whether subsequent educational outcomes are related to program participation or are a function of systematic observed and unobserved differences in the students.

In our analysis we estimate the impact of GMS program participation on retention, student loan debt accumulation, and hours worked while enrolled in college. Since our data on GMS participants only tracks students through the junior year of college, we explore retention rather than college graduation. One of the stated goals of the GMS program is to "develop a diversified cadre of future leaders for America by facilitating successful completion of bachelors, masters, and doctorate degrees." [1] In an effort to determine whether the program is likely to be successful in reaching this goal,

---

[1] See http://www.gmsp.org/about.aspx.

we assess whether GMS participation impacts factors related to college completion and graduate school attendance.

Debt upon college completion has been found by some (Millett, 2003) to influence whether or not an individual attends graduate school and may also influence choice of career. Working while in college has been found to increase college dropout and the time until degree completion among those who persist (Ehrenberg and Sherman, 1987). Thus, we also explore the impact of GMS program participation on student loan debt and working behavior in college.

Specifically, below we describe the results obtained using regression discontinuity methods to estimate the effect of GMS program participation on student retention in the freshman and junior year of college, the amount of loan debt accumulated during the same two points of the student's academic tenure, and the amount the student worked during their first and third years enrolled in college. We employ both parametric instrumental variable methods (which account for the "fuzzy" nature of the design in our case) as well as local polynomial smoothing techniques to identify the causal effect of the Gates scholarship on these aforementioned outcomes. Estimates are provided for each of the minority groups covered by the scholarship (African Americans, Asian Americans, and Latino/a students). We find causal evidence that the GMS program improves a number of important student outcomes for low income, high ability, minority students served by the program. Generally, retention is higher and loan debt and work hours during college are lower for GMS recipients, though the results vary by entering cohort and racial/ethnic group.

This paper is organized in the following way: In the next section we discuss the structure of the selection mechanism by which students are chosen for the GMS program. In Section III we discuss the estimation strategies used. Section IV details the results of the analysis conducted and Section V concludes the article.

II. The Gates Millennium Scholars Program

The Gates Millennium Scholars (GMS) program is a $1 billion, 20-year project designed to promote academic excellence by providing higher education opportunities for low-income, high-achieving minority students. Graduate and undergraduate [2] students apply for the program and have to meet a number of eligibility criteria before being accepted. Cognitive assessment measures are used to judge the academic potential of applicants (e.g., academic rigor of their high school course work, high school grades), but non-cognitive measures are also used in the selection process. Applicants must provide evidence that their high school grade point average is at least 3.33 (on a 4.00 scale). Regarding the non-cognitive component of selection into the program, students applying for admission are required to write short answers to a series of questions developed to measure an applicant's non-cognitive abilities (for information on the development and use of the non-cognitive measures see Sedlacek, 1998, 2003, 2004). The answers to each of these questions are graded by trained raters and a total non-cognitive test score is assigned to each applicant. Thresholds on these non-cognitive tests are established and they vary by racial/ethnic group and by matriculating cohort. These thresholds or "cut scores" are used as another program selection mechanism. In keeping with the goal of

---

[2] In this paper we examine the GMS effects on undergraduates only.

the program to fund needy students, applicants also have to demonstrate financial need by documenting that they are eligible for the federal Pell grant program. Finally, applicants need to be citizens or legal residents of the United States and have to complete all the required application materials to be eligible for the scholarship.

Of the 4,000 or so undergraduates who apply for the program in a given year, about one-half typically make it through the reader selection process, and about 1,000 of them are eventually selected for the program. Once in the program the students receive a scholarship that is a "last dollar" award meaning that it covers the unmet need remaining after the Pell and any other scholarships or grants are awarded. The GMS scholarship is portable to any institution of higher education of the student's choice in the United States and can be used to pay tuition and fees, books, and living expenses. The average award to freshman is about $8,000 and about $10,000-$11,000 for upper division (juniors and seniors) students. The average award also differs by institution type, with students attending public institutions of higher education receiving about $8,000 and private school attendees receiving slightly more than $11,000 in financial support. As undergraduates, students are eligible for the financial support for up to five years and they can apply for additional support if they decide to attend graduate school in engineering, mathematics, science, education or library science.

In the spring of their freshman year in college all program participants and a random sample of the non-participants are surveyed by the National Opinion Research Center (NORC) at the University of Chicago. In this "baseline" survey students are asked to respond to questions that provide information about their backgrounds, enrollment status, academic and community engagement, college finances and work, self-

concept and attitudes, and future plans. These students are also resurveyed in the late spring of their junior year in college, constituting the first "follow-up" survey.

The sample used in the analyses described below was constructed by matching data from a number of sources including the baseline and follow-up surveys mentioned above, a file containing the non-cognitive scores of applicants, and a data set containing the reasons why students were eligible or not. Two cohorts of entering undergraduate students were combined (the fall 2001 matriculants, known as Cohort II and the fall 2002 entering matriculants, known as Cohort III). After removing a few (less than 20) inaccurate cases, the effective sample used in the analysis contains about 3,000 students, nearly evenly divided between GMS participants and non-participants. There are observable differences in the overall sample including more (fewer) Latino/a (Asian American) students receiving (not receiving) scholarships than in the non-recipient group. Not surprisingly given the selection criteria, the parents of GMS scholars tend to have lower incomes and lower levels of education compared to their non-recipient counterparts. The SAT scores and percent of students who have less than four years of mathematics in high school are roughly equivalent between program participants and non-participants. In the Cohort II and III sample used in this study nearly all the freshman are retained to the fall of their sophomore year. The average retention rate is about 98 percent (compared to about 74 percent nationally), with GMS participants slightly more likely to be retained in college in their freshman year (about 99 percent) compared to their non-GMS counterparts (about 98 percent; this difference is significant at p=.004). Through their junior year of college we observe very little additional attrition, with about 97 percent of all students being retained overall, and similar

differences in the GMS/non-GMS retention rates as in the freshman year (98 vs. 97 percent, respectively; p=.03).

The dollar amount of loans borrowed in the freshman year is about $2,140 for the full sample. Not surprisingly, GMS participants borrow much less than their non-GMS colleagues, the former borrowing about $1,000 in their freshman year compared to about $3,200 for non-participants. Using National Postsecondary Student Aid Study (NPSAS) 1999-2000 data, we calculated freshman loan levels for high ability, Pell eligible students and the average was slightly lower (at about $2,800) than the overall average in Cohorts II and III. Average cumulative loan levels though the junior year of college for the full sample are about $6,800, with GMS students borrowing about $3,400 and their non-recipient counterparts borrowing about $10,200. NPSAS data indicates cumulative borrowing for similar students (high ability, low income) to be about $6,100 on average.

The NPSAS data also contains information on hours worked while students are enrolled in college. In 1999-2000, high-ability, low income students worked about 19 hours in their freshman and 19.5 hours per week in their junior year of college. The average number of hours worked in the Cohort II and III sample during the freshman year was substantially smaller (at slightly less than 13 hours) than national averages during the freshman year. GMS participants worked about 11 hours during an average academic year work-week, whereas the control group reported working 13.4 hours (significant at p=.0000). During their junior year, students in the Gates sample reported increasing their work effort to about 16 hours, with the difference between GMS recipients (15 hours) and their non-recipient colleagues (19 hours per week) being about four hours (significant at p=0000). The reduction in hours worked during the freshman year may be

beneficial if this extra time is used for additional studying and/or students become more engaged in the academic and social fabric of the institution. Reasons for the relative increases in hours worked during the junior year deserve more investigation.

III. The Estimation Strategy

In the early 1960s Thistlewaite and Campbell (1960) used the regression discontinuity (RD) technique to study the effects of the National Merit Scholarship program on career choice. Since then the method has also been used to examine the effects of compensatory education programs, especially Title I programs (Trochim, 1984) and in recent years RD has been used to examine school district and housing prices (Black, 1999), the effect of school funding on pupil performance (Guryan, 2000), how student financial aid affects student enrollment behavior (van der Klauuw, 2002; Kane, 2003), how teacher training impacts student achievement (Jacob and Lefgren, 2002), and the relationship between failing the high school exit exam and graduation from high school and/or subsequent postsecondary education outcomes (Martorell, 2004).

RD is a quasi-experimental, pre-test/post-test design (see Cook and Campbell, 1979) where subjects are assigned to the treatment (e.g., GMS participation) and control groups (e.g., GMS non-participants) based on a score on some pre-specified criterion (or criteria). As noted above, to participate in the GMS program students first need to have scores on the non-cognitive essay test score or "running variable" threshold which varies by race/ethnicity and freshman entering cohort. Students who meet this condition then must also meet the criteria for the federal Pell financial aid program in order to receive the scholarship granted to GMS participants.

Given the selection mechanism operating we expect that students are distributed quite randomly above and below the cut point. If this is the case then the observed and unobserved characteristics of students around the cut score are very similar, akin to a randomized experiment around the cut point. Under these circumstances an evaluation of the effect of the program at this point has strong causal implications. If the program has a positive (or negative) effect on a particular educational outcome we expect to observe a discontinuity at or near the cut score. This discontinuity helps to identify the causal effect of the GMS program, defined as the vertical distance between the regression intercepts on each side of the cut point.

Figures 1 and 2 also provide descriptions of the non-cognitive score densities by race/ethnicity for Cohorts II and III respectively. The distributions appear more variable for the Cohort II sample than for Cohort III matriculates. The Cohort III sample distributions also appear to be slightly skewed to the right.

We examine the distributions of demographic and high school performance characteristics of students for the full sample and within one and two-point intervals on either side of the cut score. We find no evidence of statistically ascertainable differences on these measures around the cut point (see Table 1). We also estimate predictive models for our three dependent variables which included several predictor variables other than our total non-cognitive score running variable.[3] If the are non-random differences in these explanatory variables around the cut point then we would expect to see jumps in the predicted values of these dependent variables at the cut point (See Card, Chetty, and Weber, 2006). Figures 3 through 5 present plots of the average predicted a value by total

non-cognitive score as well as non-parametric regression estimates using Lowess (see

Härdle, 1990) of the predicted value on total non-cognitive score. As theses figures show,

there appear to be no discernable jumps at the cut points. This further bolsters our

confidence that, approximately, individuals are randomly distributed around the cut

point.

A. The Regression Discontinuity Approach

Given the mechanism by which the Gates Millennium scholarships are awarded,

we believe it makes sense to analyze the impact of scholarships on student retention,

debt, and work behavior while in college using regression discontinuity inference. In this

section we briefly outline our approach.

Suppose that an outcome variable (y) depends on a regressor (x), and whether a

treatment is received (or not) is represented by an indicator variable (D). According to

(1) $y = m(x) + D\alpha + \varepsilon$

where m(x) is a continuous function of x, α measures the impact of the treatment (D) on

the E(y) and ε is a zero mean random error with $E(\varepsilon \mid x) = 0$ and $\text{var}(\varepsilon) = \sigma_x^2$. In a

"sharp" regression discontinuity design there is a variable, z, such that D = 1 if $z \geq \tilde{z}$,

where the value $\tilde{z}$ is the threshold or cut point, and D equals zero otherwise. Taking

expectations of both sides of (1) with respect to z yields

(2) $E(y \mid z) = E(m(x) \mid z) + \alpha + E(\varepsilon \mid z)$

when $z \geq \tilde{z}$ and

---

[3] Models are estimated separately by ethnic group and control for type of high school (public, private, religious), composite SAT score, number of years of science, number of years of math, family size, whether the family owns a home, parents' education, and immigrant status.

(3) $E(y\,|\,z) = E(m(x)\,|\,z) + E(\varepsilon\,|\,z)$.

when $z < \tilde{z}$. The regression discontinuity design assumes that

the $\lim_{z\downarrow\tilde{z}} E(\varepsilon\,|\,z) = \lim_{z\uparrow\tilde{z}} E(\varepsilon\,|\,z)$. So, taking limits of (2) and (3) as $z\downarrow\tilde{z}$ and $z\uparrow\tilde{z}$,

respectively results in

(4) $\lim_{z\downarrow\tilde{z}} E(y\,|\,z) = \lim_{z\downarrow\tilde{z}} E(m(x)\,|\,z) + \lim_{z\downarrow\tilde{z}} E(\varepsilon\,|\,z) + \alpha = E(m(x)\,|\,\tilde{z}) + E(\varepsilon\,|\,\tilde{z}) + \alpha$

(5) $\lim_{z\uparrow\tilde{z}} E(y\,|\,z) = \lim_{z\uparrow\tilde{z}} E(m(x)\,|\,z) + \lim_{z\uparrow\tilde{z}} E(\varepsilon\,|\,z) = E(m(x)\,|\,\tilde{z}) + E(\varepsilon\,|\,\tilde{z})$.

Subtracting (5) from (4) yields

(6) $\lim_{z\downarrow\tilde{z}} E(y\,|\,z) - \lim_{z\uparrow\tilde{z}} E(y\,|\,z) = \alpha$.

Estimates of $\lim_{z\downarrow\tilde{z}} E(y\,|\,z)$ and $\lim_{z\uparrow\tilde{z}} E(y\,|\,z)$ can then be used to estimate α.

In a "fuzzy" regression discontinuity design we replace the assumption that D = 1

when $z \geq \tilde{z}$ and equals zero otherwise with

(7) $\lim_{z\downarrow\tilde{z}} E(D\,|\,z) - \lim_{z\uparrow\tilde{z}} E(D\,|\,z) = \delta > 0$.

In this situation, as z crosses the threshold $\tilde{z}$ there is a discontinuous jump in the

probability of treatment (in this case GMS participation). Again taking conditional

expectations of both sides of (1) with respect to z, and taking limits as $z\downarrow\tilde{z}$ and

$z\uparrow\tilde{z}$ produces

(8) $\lim_{z\downarrow\tilde{z}} E(y\,|\,z) = \lim_{z\downarrow\tilde{z}} E(m(x)\,|\,z) + \lim_{z\downarrow\tilde{z}} E(\varepsilon\,|\,z) + \lim_{z\downarrow\tilde{z}} E(D\,|\,z)\alpha = E(m(x)\,|\,\tilde{z}) + \lim_{z\downarrow\tilde{z}} E(D\,|\,\tilde{z})\alpha + E(\varepsilon\,|\,\tilde{z})$

(9) $\lim_{z\uparrow\tilde{z}} E(y\,|\,z) = \lim_{z\uparrow\tilde{z}} E(m(x)\,|\,z) + \lim_{z\uparrow\tilde{z}} E(\varepsilon\,|\,z) + \lim_{z\uparrow\tilde{z}} E(D\,|\,z)\alpha = E(m(x)\,|\,\tilde{z}) + \lim_{z\uparrow\tilde{z}} E(D\,|\,z)\alpha + E(\varepsilon\,|\,\tilde{z})$.

Differencing (8) and (9) produces

$\lim_{z\downarrow\tilde{z}} E(y\,|\,z) - \lim_{z\uparrow\tilde{z}} E(y\,|\,z) = \alpha\left\{\lim_{z\downarrow\tilde{z}} E(D\,|\,z) - \lim_{z\uparrow\tilde{z}} E(D\,|\,z)\right\} = \alpha\delta$

or

$$(10) \quad \frac{\lim_{z\downarrow\tilde{z}} E(y\mid z) - \lim_{z\uparrow\tilde{z}} E(y\mid z)}{\delta} = \alpha \;.$$

One strategy to estimate the impact of the treatment (α) is to assume some flexible functional form for m(x) such as a high order polynomial and use linear (OLS) regression methods in the case of the sharp design. In the fuzzy design, two-stage instrumental variable estimation is used where an indicator variable I ($z \geq \tilde{z}$) is used as an instrument for D.

The GMS program's method of selecting participants fits into the fuzzy design because applicants whose scores on the non-cognitive test exceed the threshold value are still not guaranteed a GMS scholarship. As noted above, the applicant must also be eligible for a federal Pell grant. If we assume that the likelihood of Pell eligibility is similar among those applicants "near" the threshold, then a fuzzy design will yield consistent estimates of the treatment. Another issue has to do with the discrete nature of the non-cognitive test score which takes only integer values. As shown in Card and Lee (forthcoming), clustered standard errors are appropriate when z is discrete, with z used as the clustering variable.

As noted above, the GMS program has race-specific thresholds that may change with different cohorts of GMS scholars. When estimating effects when race and/or entering cohorts are combined, interaction variables of the non-cognitive test score (and its higher order polynomials) and race and/or cohort are also included as regressors. Our estimation strategy assumes that m(x) follows linear, quadratic, and cubic polynomial forms. As a further check of the robustness of our estimates to the specification of m(x), we apply non-parametric regression methods in order to estimate α (see Porter, 2003).

In particular, we employ local linear regression to estimate $\lim_{z \downarrow \tilde{z}} E(y \mid z)$, $\lim_{z \uparrow \tilde{z}} E(y \mid z)$

and $\lim_{z \downarrow \tilde{z}} E(D \mid z)$. Given that no applicant receives a scholarship if $z < \tilde{z}$

then $\lim_{z \uparrow \tilde{z}} E(D \mid z) = 0$. The details of the methods employed as well as the manner in

which we select the bandwidths for the non-parametric regressions are presented in the

Appendix.

IV. The Results

Two sets of results are discussed below. First, the IV quadratic regression results

are presented as they generally fit the Cohort II and III data better than a linear model

while models with cubic representations do not significantly improve fit. Second, the

non-parametric results are also reported although these estimates are based only on the

Cohort III data because, due to different a sampling scheme for Cohort II, the sample size

on the left-hand side of the cut point for the Cohort II sample is too small near the cut.

As new cohorts become available we will be able to increase the sample size available

which should make our non-parametric estimates more precise.[4]

A. *Estimating Retention to the Sophomore Year*

---

[4] Results discussed below are in the form of percentage differences between the treatment
mean (Y_t) and the counterfactual mean (Y_c) defined as $100 * [(Y_t - Y_c) / Y_c]$. Thus,
when discussing "differences," "increases," or "decreases" in outcomes below the
relevant comparison groups are students who received funding from the GMS program
relative to the counterfactual group, that is, the expected outcomes if GMS participants
did not receive the "treatment."

Given that most attrition from college happens during the freshman year or in the summer prior to the sophomore year (Tinto, 1987) there is interest in knowing the extent to which programs such as the GMS increase freshman to sophomore year retention.

Descriptively, the overall retention rate for African Americans is 98.1 percent, with GMS participants' retention at 99 percent and their non-participants colleagues being retained at 97.2 percent (p=.04). Asian American (Latino/a) students' overall retention is 98.6 (99.3) percent, with GMS awardees having retention levels of 99.2 (99.6) percent (respectively) compared to non-participants whose rates are 98.1 (98.9) percent, respectively (these differences are not statistically significant at conventional levels).

Given these extremely high rates of retention to the sophomore year, only the IV procedure with a linear term for the total non-cognitive score variable was estimated for the pooled sample and different ethnic subgroups. As shown in Table 2, there is evidence that the scholarship improved Asian American students' freshman-to-sophomore retention by about 21 percent.[5]

B. Estimating Retention through the Junior Year

We also investigate whether there is evidence that the GMS program produces retention differences through the junior year of college. Descriptive results indicate small differences in the overall and race/ethnic-specific retention levels through the junior year of college. Overall, African Americans and Asian Americans have nearly identical overall retention rates at 97.5 and 97.4 percent, respectively. Latino/a students'

average retention is slightly higher at 97.7 percent. Differences in retention for GMS participants compared to their non-participating colleagues are about two percentage points for Asians (98.8 vs. 96.5), less than a point for Latino/a students (97.9 vs. 97.3), and is 98.3 vs. 96.8 for African American students.

The pooled (overall sample) IV estimates indicate no significant differences in retention rates for the GMS students compared to their non-GMS counterparts. However we did find subgroup differences in retention through the junior year for Asian Americans, where our results provide evidence that the effect for these students is about 11 percent.

The non-parametric RD estimates for junior year retention are shown in Figures 6 and 7. Figure 6 shows the estimates of the numerator in (10) while Figure 11 shows the estimates of the denominator in (1). While these figures indicate negative effects for African American and Latino/a GMS scholars, based on bootstrap estimates using 1000 replications the point estimates are not statistically significant at conventional levels. We find, however, statistically significant differences in junior year retention for Asian GMS recipients compared to the counterfactual group, with the former having about 17 percent higher retention levels than the latter group.

C. Loan Accumulation in the Freshman Year

We also investigate whether there is evidence of a causal effect of the GMS program on the loan debt students accumulate in their freshman year. Descriptively, GMS recipients have lower loan debt than non-GMS recipients and this difference varies

---

[5] Formally, the improvement is for students near the cut point since we only identify the local treatment

by race/ethnicity. African Americans debt levels for GMS students is the highest at about $1,133; Asian students have loan levels about $1,000 and Latino/a students average loan amount in the freshman year is $878. Non-GMS students' average loan levels in the freshman year are $3,070 for African Americans, $3,166 for Asian American students, and $3,435 for Latino/a non-recipients. (All these differences are significant at p< .001).

The IV estimates provide evidence that overall loan debt in the freshman year is about 60 percent lower for GMS students and the race-specific results are as follows: The estimate is about -52 percent for African American students, indicating that GMS scholars have loan debt levels about 52 percent lower than the counterfactual group. The point estimates are about -81 percent for Latino/a GMS students and about -36 percent for Asian American students, though the latter two are measured imprecisely and not statistically significant.[6]

D. Loan Debt Through the Junior Year

Next we estimate whether the GMS is effective in reducing the total loan debt accumulated through the junior year of college. Descriptive statistics indicate that regardless of GMS affiliation, African American students have the highest debt load levels through their junior year of college (about $11,000 for non-GMS students vs. about $4,100 for GMS scholars). At about $2,800 for each group, the average debt level for Asian American and Latino/a GMS participants is substantially lower than that of their African American colleagues. Debt levels for non-GMS Asian and Latino/a

effect at the cut point.

[6] Although not reported, the non-parametric estimates indicate a 72 percent lower loan debt amount for African Americans, Asian American loan debt is 92 percent lower, and Latino/a students have freshman year loan debt amounts that average about 59 percent lower than the counterfactual group.

students is closer to the African American non-participant average at about $10,200 and $9,300, respectively. (All these differences are significant at p< .001).

The quadratic IV estimates indicate that for the full sample, accumulated loan debt is about 59 percent lower for GMS students compared to the counterfactual group. We also find the GMS program reduces accumulated loan debt for Latino/a students by about 77 percent. African American and Asian American students also have lower debt estimates, at 42 and 66 percent lower (respectively) than the counterfactual group. All the above results are significant at p < .05. The non-parametric estimates, shown in Figures 8 (and Figure 7) give similar results.

E. Hours Worked in the Freshman Year

When surveyed by NORC, GMS participants and their non-participating counterparts indicate how many hours they worked for pay during an average week for the academic year in question. Reductions in work may improve academic performance among GMS recipients if any reductions are converted to time studying and/or becoming engaged in the social and academic fabric of the institution.

Descriptively, there are again differences between GMS participants and non-participants and by race/ethnicity. Among GMS recipients, Latino/a students average the most hours worked at about 11.7 hours per week; African American students work about an hour less per week (10.6) and Asian students work the least at 8.7 hours. Non-GMS students work more hours, with African American working about 16 hours per week followed by Latino/a students who average 14.3 and Asian students working about 13.6 hours per week during the academic year.

The conditional IV estimates of hours worked in the freshman year for the full sample indicate GMS recipients work about 35 percent fewer hours than the counterfactual group. Asian American students work considerably fewer hours than overall, with hours worked about 76 percent lower, and the difference at the cut point for African American students is about -29 percent and the estimate for Latino/a students is about -12, however, the latter estimate and the African American result described above are not statistically significant at conventional levels.[7]

F. Hours Worked Through the Junior Year

We estimate the effects that the GMS had on hours worked in the junior year of college enrollment. Descriptive statistics indicate that non-GMS African American, Asian American, and Latino/a students worked more hours per week (19.2, 17.1, and 19.3 hours, respectively) than their GMS recipient counterparts (14.2, 12.1, and 13.2). The results are similar for the non-parametric estimates (see Figures 7 and 9).

V. Conclusions

As noted above, our initial results indicate that the GMS program improves a number of important student outcomes for low income, high ability, and minority students served by the program. As noted above, retention is higher for Asian GMS students in the junior year and large differences in loan debt and in some cases work hours during college for GMS recipients. We find that these results often vary by racial/ethnic group and whether we examine the freshman year or results through the

---

[7] In all cases the non-parametric estimates are not statistically significant.

junior year of college. Our estimates, however, formally apply only to individuals near the cut and this caveat should be kept in mind when interpreting our results.

Although the RD design avoids some of the distributional and other assumptions used in selection models, the RD approach is not "free of problems and difficulties" (Pedhazur and Schmelkin, 1991, p. 298). One of the potential problems is how much data is available on the left hand side of the cut point. Given small amount of data on the left of the threshold in the Cohort II sample, especially for certain race/ethnic groups, we were restricted in some cases to using the Cohort III data only when employing non-parametric methods. This limitation should be less of a problem when adding subsequent cohorts because NORC is now surveying larger random samples of non-recipients than they did for Cohort II. Adding additional cohorts will increase the sample size, and provide more support left of the cut point which should stabilize our non-parametric estimates.

Herein we investigated the program's effect on three important student outcomes. There are, however, many other important outcomes to be examined including academic performance while in college, graduation from college, continuation to graduate school, and whether the program improves student non-cognitive outcomes such as self-efficacy, leadership, and involvement in one's community. We are currently applying the methods detailed above to some of these student outcomes.

References

Card, D., Chetty, R., and Weber, A. (2006), "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," N.B.E.R. Working Paper #12639.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally.

Ehrenberg, R. G. and Sherman, D. R. (1987). Employment While in College, Academic Achievement, and Postcollege Outcomes. *Journal of Human Resources*, 22, 1-23.

Fan, J. (1992). Design Adaptive Nonparametric Regression. *Journal of the American Statistical Association*, 98: 998-1004.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. New York: Chapman and Hall.

Guryan, J. (2000). *Does Money Matter? Regression Discontinuity Estimates from Education Finance Reform in Massachusetts*. University of Chicago Working Paper.

Hahn, J., Todd, P. and Van der Klaauw, W. (2002). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1): 201-209.

Hardle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

Hardle, W. and Linton, O. (1994). Applied Nonparametric Methods. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*. New York: North Holland.

Heckman, J. J., Hotz, V. J., and Dabos, M. (1987) Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings? *Evaluation Review*. 11: 395-427.

Heckman, J. and Robb, R. (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer, (Ed.), *Drawing Inferences from Self-Selected Samples*. Mahwah, NJ: Erlbaum.

Jacob, B. A. and Lefgren. L. (2002). *The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago*. National Bureau of Economic Research Working Paper No. 8916.

Kane, T. J. (2003). *A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going*. National Bureau of Economic Research Working Paper No. 9703.

Manski, C. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review*, 80: 319-323.

Millet, C. M. (2003). How Undergraduate Loan Debt Affects Application and Enrollment in Graduate or First Professional School. *Journal of Higher Education*, 74: 386-427.

Moffit, R. (1991) Program Evaluation with Nonexperimental Data. *Evaluation Review*, 15:291-314.

Porter, J. (2003). Estimation in the Regression Discontinuity Model. Harvard University manuscript. Retrieved from http://www.ssc.wisc.edu/~jrporter/reg_discont_2003.pdf on February 18, 2006.

Sedlacek, W. E. (1998). Admissions in Higher Education: Measuring Cognitive and Non-Cognitive Variables. In Wilds, D. J. and Wilson, R. (Eds.) *Minorities in Higher Education 1997-98: Sixteenth Annual Status Report*. Washington, D.C.: American Council on Education.

Sedlacek, W. E. (2003). Alternative Measures in Admissions and Scholarship Selection. *Measurement and Evaluation in Counseling and Development,* 35: 263-272.

Sedlacek, W. E. (2004). *Beyond the Big Test: Noncognitive Assessment in Higher Education*. San Francisco: Jossey-Bass.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*, 51: 309-317.

Tinto, V. (1987). *Leaving College*. Chicago, IL: University of Chicago Press.

Trochim, W. M. K. (1984). *Research Design for Program Evaluation: The Regression-Discontinuity Approach.* Beverly Hills, CA: Sage Publications.

van der Klaauw, W. (2002). Estimating the Effects of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach. *International Economic Review*, 43(4): 1249-1287.

Willis, R. and Rosen, S. (1979). Education and Self-Selection. *Journal of Political Economy*, 87(5, part 2): 507-536.

Winship, C. and Mare, R. (1992). Models for Selection Bias. *Annual Review of Sociology*, 18: 327-350.

Appendix
Local Polynomial Regression Estimates and Optimal Bandwidth Determination

$$\alpha = \frac{\lim_{z \downarrow \tilde{z}} E(y \mid z \geq \tilde{z}) - \lim_{z \uparrow \tilde{z}} E(y \mid z < \tilde{z})}{\lim_{z \downarrow \tilde{z}} E(D \mid z \geq \tilde{z}) - \lim_{z \uparrow \tilde{z}} E(D \mid z < \tilde{z})} = \frac{\lim_{z \downarrow \tilde{z}} E(y \mid z \geq \tilde{z}) - \lim_{z \uparrow \tilde{z}} E(y \mid z < \tilde{z})}{\lim_{z \downarrow \tilde{z}} E(D \mid z \geq \tilde{z})}$$

where the left side of the equality follows in our situation because

$$\lim_{z \uparrow \tilde{z}} E(D \mid z < \tilde{z}) = 0 \text{ for all } z < \tilde{z}.$$

To derive a consistent estimator of α we need to consistently estimate $E(y \mid z \geq \tilde{z})$, $E(y \mid z < \tilde{z})$ and $E(D \mid z \geq z)$ in a neighborhood of $\tilde{z}$. To obtain consistent estimates of these three terms we apply local polynomial regression.

Consider the regression model

$$y = m(x) + \varepsilon$$

Local polynomial regression estimates of m(x) at a point x_0 by estimating a weighted polynomial regression where points near x_0 receive larger weights. Suppose that a local polynomial regression of order p is estimated. Let $\mathbf{X}$ be the matrix defined by

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix}$$

and let $\mathbf{y}$ be the vector

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}$$

Finally, define a diagonal weighting matrix $\mathbf{W}$ by

$$\mathbf{W} = \text{diag}\left\{ K_h(X_i - x_0) \right\}$$

where K_h is a kernel weighting function with bandwidth h and is defined by

$$K_h(\bullet) = K(\bullet / h) / h$$

for some kernel function. Throughout we use the Epanechnikov kernel function defined by $K(u) = \frac{3}{4}(1-u^2)$ for $-1 < u < 1$. The estimated local polynomial coefficients at x_0,

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

are then obtained from

$$\min_{\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) .$$

For theoretical reasons (See Fan and Gijbels, 1995 and Porter, 2003) it is preferable to estimate odd ordered polynomial models. In our estimates α we simply estimate a local linear regression (p = 1). For our estimate of α we then estimate three local linear regressions for $E(y \mid z \geq \tilde{z})$ and $E(D \mid z \geq \tilde{z})$ and use data from the right of the cut point only, and for $E(y \mid z < \tilde{z})$ we use data to the left of the cut point. Letting z_{-} be the closet point on our grid of z values to the left of $\tilde{z}$ (which in out case is $\tilde{z} - 0.1$) and z_{+} be the point closest on our grid of z values to the right of $\tilde{z}$ ($\tilde{z} + 0.1$), the estimated value of α equals

$$\hat{\alpha} = \frac{\hat{E}(y \mid z_{+}) - \hat{E}(y \mid z_{-})}{\hat{E}(D \mid z_{+})} .$$

To implement this technique it is necessary to choose a bandwidth. We choose the bandwidth that minimizes the asymptotic mean squared error. For a local linear regression model this optimal bandwidth equals (see Fan and Gijbels, 1995)

$$h_{opt}(x_o) = C(K)\left[\frac{\sigma^2(x_0)}{\{m''(x_0)\}^2 f(x_o)}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}$$

where

$$C(K) = \left[\frac{\int K^2(t)dt}{\{\int t^2 K(t)dt\}^2}\right]^{\frac{1}{5}} ,$$

$\sigma^2(x_0)$ is the variance of ε at x_0, $m''(x_0)$ is the second derivative of m at x_0, and f(x_0) is the density of x at x_0. For the Epanechnikov kernel function $C(K) = 1.719$. Several of these quantities are unknown and so we employ a two step method to obtain the optimal bandwidth.

In the first step we compute what is termed the "Rule of Thumb" (ROT) bandwidth which we denote h_{ROT}. To compute h_{ROT} a forth order polynomial is estimated globally (i.e., with all data weighted equally). From these estimates we compute

$$\tilde{m}(x) = \hat{\beta}_0 + \cdots + \hat{\beta}_4 x^4$$

which gives

$$\tilde{m}''(x) = 2\hat{\beta}_2 + 6\hat{\beta}_3 x + 12\hat{\beta}_4 x^2$$

and $\hat{\sigma}^2$ where

$$\tilde{\sigma}^2 = \sum_{i=1}^{N}(y_i - \tilde{m}(x_i))^2 / (N-5).$$

Then,

$$h_{ROT} = 1.719 \left[\frac{\tilde{\sigma}^2}{\sum_{i=1}^{n}\left\{\tilde{m}''(x_i)\right\}^2} \right]^{1/5}$$

In the second step we estimate a 3rd order local polynomial regression using bandwidth h_{ROT} to compute

$$h_{opt}(x_0) = 1.719 \left[\frac{\hat{\sigma}^2(x_0)}{\sum_{i=1}^{n}\left\{\hat{m}''(x_i)K_{h_{ROT}}(x_i - x_0)\right\}^2} \right]^{1/5}$$

where

$$\tilde{\sigma}^2 = \sum_{i=1}^{N}(y_i - \hat{m}(x_i))^2 / \mathrm{tr}\left\{\mathbf{W} - \mathbf{WX}(\mathbf{X'WX})^{-1}\mathbf{X'W}\right\}.$$

When computing the optimal bandwidth for $\hat{E}(y\,|\,z_+)$ and $\hat{E}(D\,|\,z_+)$ we use only data to the right of the cut point $\tilde{z}$ when $x_0 = z_+$. When computing the optimal bandwidth for $\hat{E}(y\,|\,z_-)$ we use the data to the left of the cut point and $x_0 = z_-$. We bootstrapped the 95% confidence intervals for $\hat{\alpha}$ using 1000 replications and recomputed the optimal bandwiths for each replication. We report the bias adjusted confidence intervals.

Table 1
Sample Means and Means Just Above and Below the "Cut Points"
for Demographic and High School Background Variables

Variable Name	Full Sample	All Applicants with Total Non-Cognitive Scores Equal to the…		p-value
		Cut Score or Cut Score + 1	Cut Score - 1 or Cut Score - 2	
ACT Composite Score	23.7	23.58	24.12	0.28
SAT Verbal+Math Score	1123.92	1110.76	1124.86	0.35
Attended Religious High School	0.06	0.06	0.04	0.25
Attended Private High School	0.07	0.08	0.04	0.12
Years of High School Math	3.87	3.88	3.86	0.37
Years of High School Science	3.66	3.63	3.69	0.22
Family Size	3.69	3.65	3.66	0.96
Born in U.S.	0.61	0.61	0.58	0.57
Family Owns Home	0.51	0.47	0.5	0.47
Father's education				0.5
Less Than High school	0.2	0.19	0.22	
High School	0.27	0.27	0.25	
Some College	0.21	0.2	0.23	
BA/BS Degree	0.14	0.15	0.09	
Post BA/BS Degree	0.1	0.1	0.12	
Mother's education				0.97
Less Than High School	0.19	0.21	0.18	
High School	0.25	0.26	0.27	
Some College	0.28	0.26	0.29	
BA/BS Degree	0.18	0.18	0.17	
Post BA/BS Degree	0.07	0.07	0.06	

Notes: Cohorts II and III combined. Cut scores for total non-cognitive score were 71, 72 and 68 for African Americans, Asian Americans and Latinos, respectively in Cohort II and 72, 75 and 69 for African Americans, Asian Americans and Latinos, respectively for Cohort III. All tests of differences were Fisher exact tests for equality based on categorical data except for family size, ACT and SAT scores which were simple t-tests for differences in means.

Table 2
Estimated Impact of GMS on Outcome Variables at End of Freshman Year of College

	Enrollment	Total Loans	Hours of Work
Combined	0.000	-1935.55	-3.89
	(0.001)	(625.73)	(1.96)
African Americans	-0.010	-1255.04	-3.33
	(0.018)	(461.57)	(2.05)
Asian Americans	0.169	-1475.71	-15.28
	(0.063)	(1324.43)	(4.71)
Hispanics	0.000	-3517.54	0.76
	(0.001)	(2043.65)	(2.85)

Source: Cohorts II and III of Gates Millennium Scholarship Follow-up Surveys. See text for details.

Notes: Standard errors that are clustered based on total cognitive score are reported in parentheses. Estimates for enrollment based on the instrumental variable probit model while estimates for total loans and hours of work are based on two-stage least squares. Controls for cohort, total non-cognitive score and its square and their interaction with cohort are included in the race specific estimates; the combined model also includes controls for race and interactions of total non-cognitive score and its square with race. Estimates for the probability of enrollment exclude the squared non-cognitive score variable and its various interactions.

Table 3
Estimated Impact of GMS on Outcome Variables at End of Junior Year in College

	Enrollment	Total Loans	Hours of Work
Combined	0.009	-6481.77	-4.42
	(0.016)	(1190.85)	(1.81)
African Americans	0.001	-5155.02	-6.41
	(0.024)	(1460.02)	(2.60)
Asian Americans	0.116	-6470.32	-8.60
	(0.081)	(3082.45)	(3.72)
Hispanics	-0.031	-8066.18	0.97
	(0.028)	(1782.39)	(2.82)

Source: Cohorts II and III of Gates Millennium Scholarship Follow-up Surveys. See text for details.

Notes: Standard errors are reported in parentheses and are clustered based on total cognitive score. Estimates for enrollment based on the instrumental variable probit model while estimates for total loans and hours of work are based on two-stage least squares. Controls for cohort, total non-cognitive score and its square and their interaction with cohort are included in the race specific estimates; the combined model also includes controls for race and interactions of total non-cognitive score and its square with race.

Figure 1
Distribution of Total Non-Cognitive Score by Race/Ethnicity: Cohort II
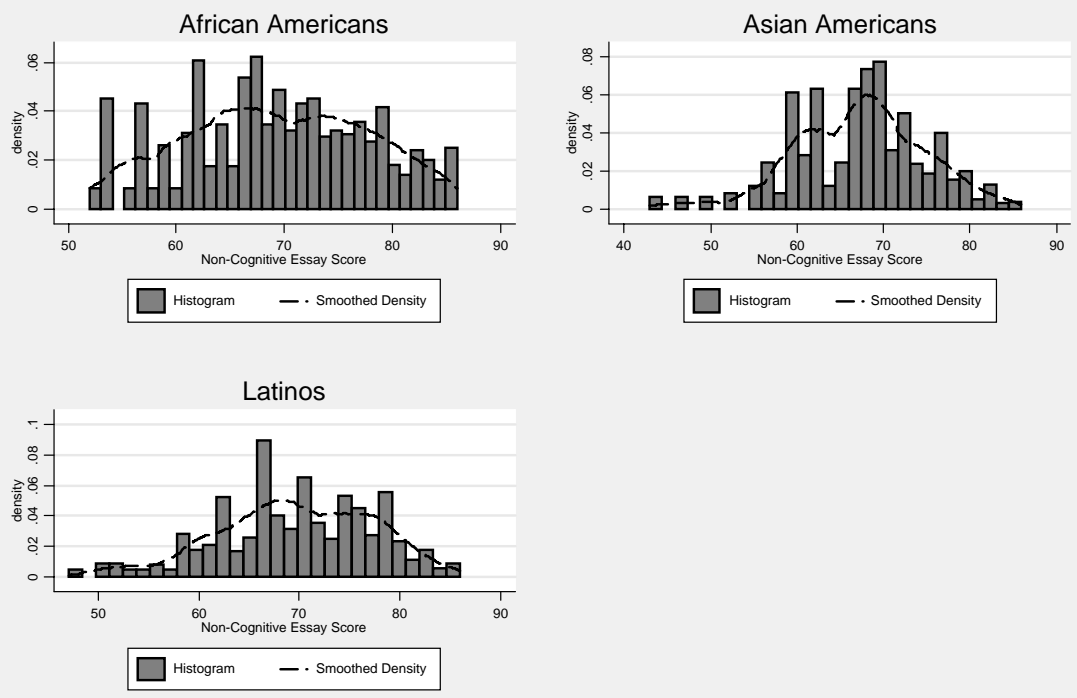
Source: Gates Millennium Scholar Surveys: Cohort II

Figure 2
Distribution of Total Non-Cognitive Score by Race/Ethnicity: Cohort III
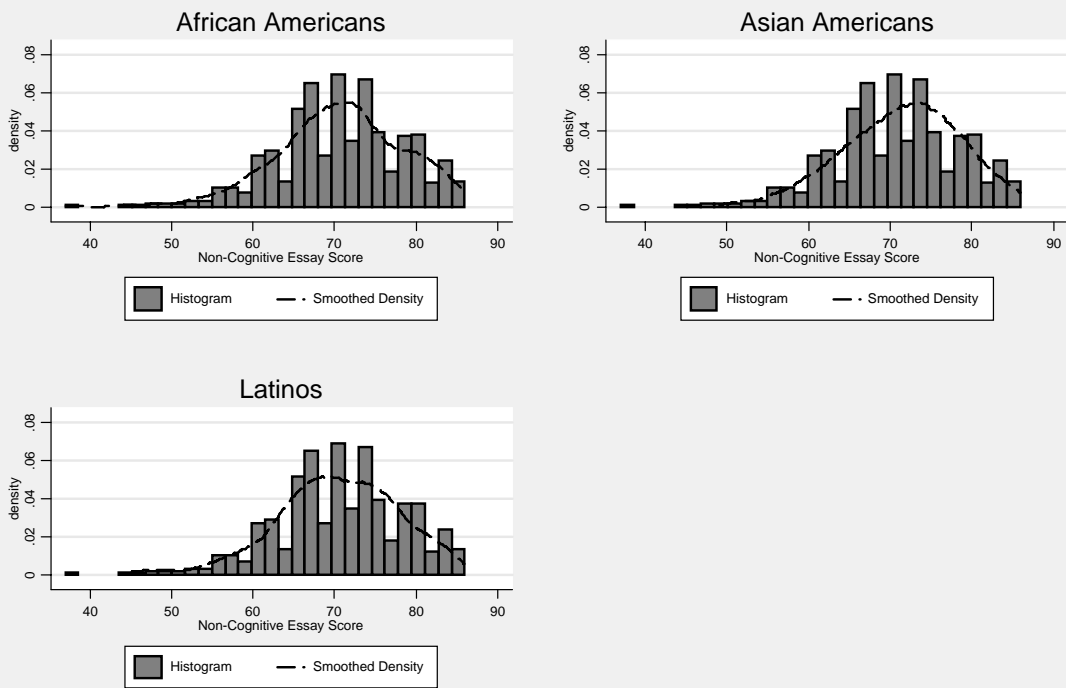
Source: Gates Millennium Scholar Surveys: Cohort III

Figure 3
Predicted Probability of Enrollment in Follow-up Survey by Total Non-Cognitive Score
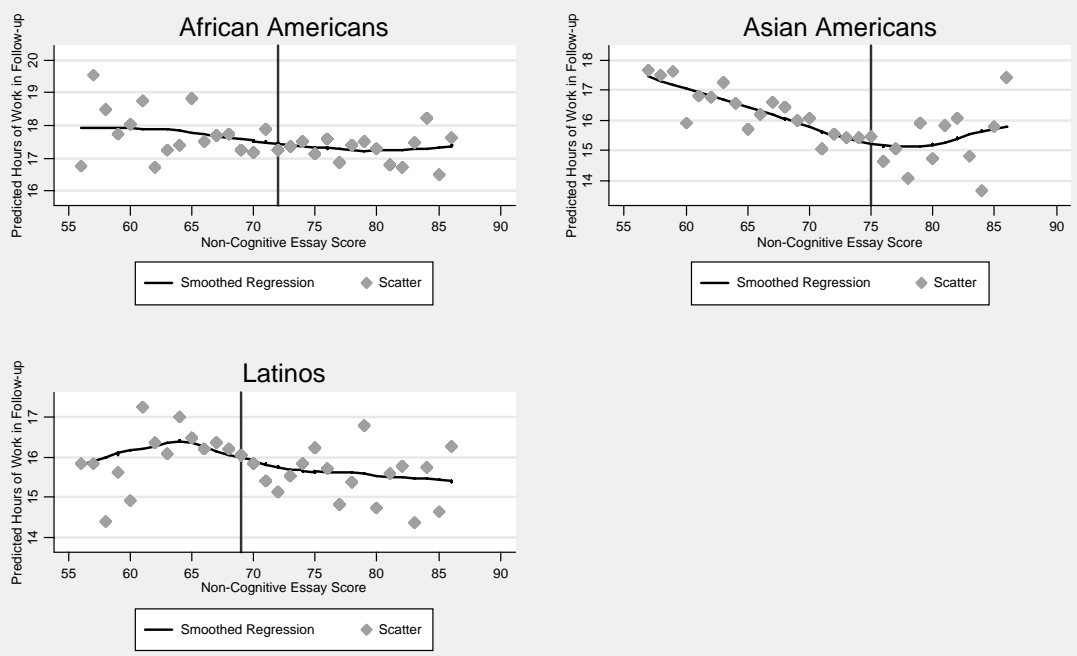Cohort III
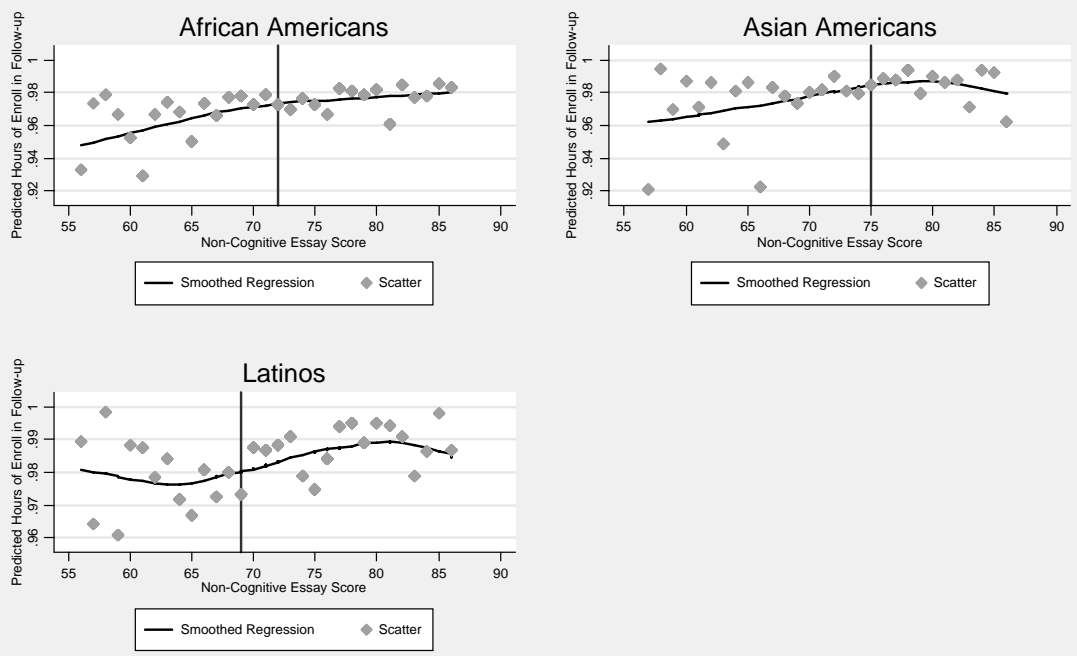
Source: Gates Millennium Scholar Surveys: Cohort III.
Notes: The vertical lines indicate the respective cut points for each ethnic group for the Gates Millennim Scholarship program
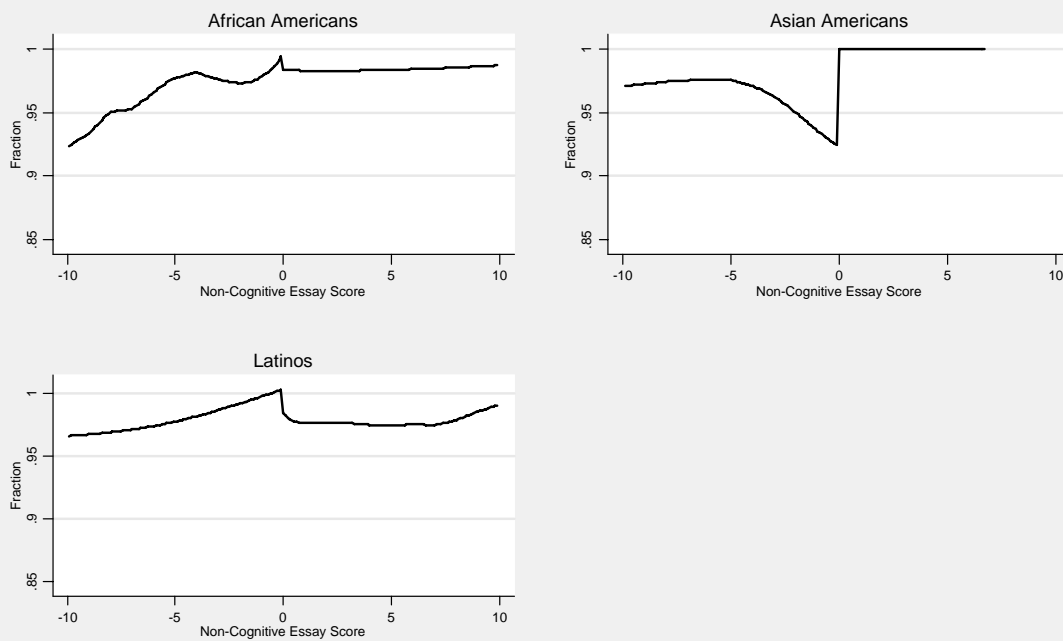
Figure 4
Predicted Total Amount of Loans in Follow-up Survey by Total Non-Cognitive Score
Cohort III
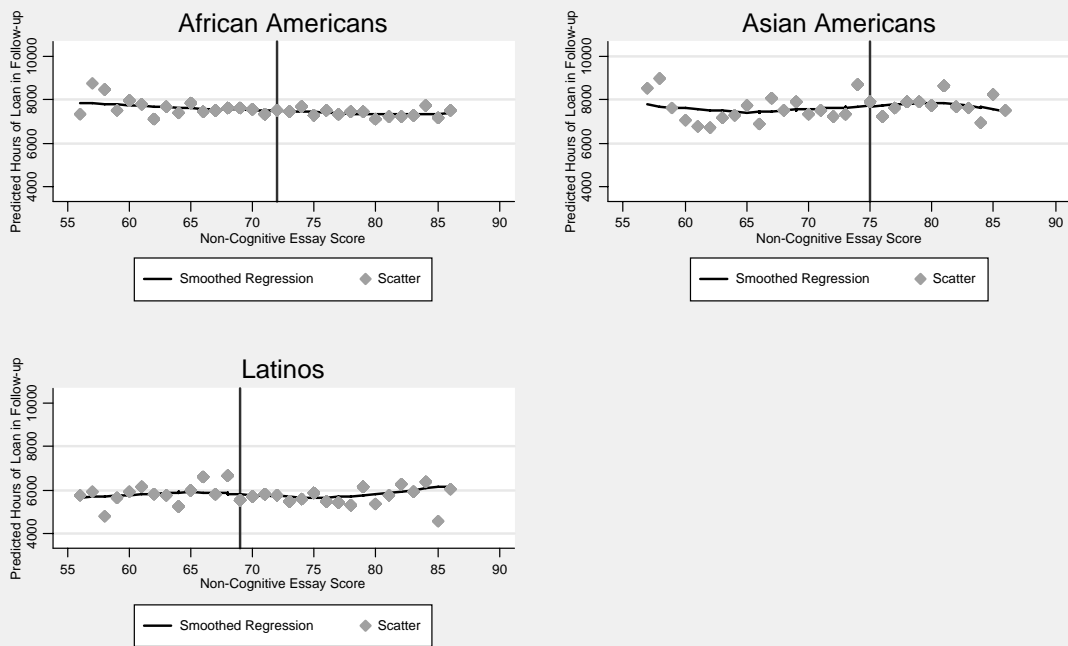
Source: Gates Millennium Scholar Surveys: Cohort III.
Notes: The vertical lines indicate the respective cut points for each ethnic group for the Gates Millennim Scholarship program
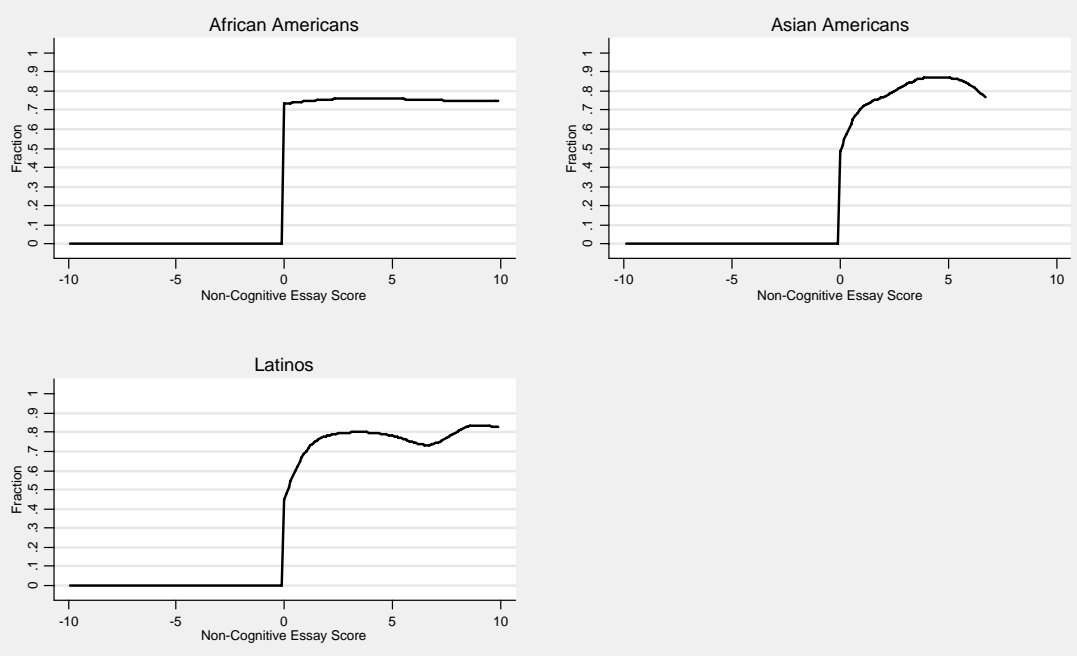
Figure 5
Predicted Hours of Work in Follow-up Survey by Total Non-Cognitive Score
Cohort III

African Americans

Asian Americans

Latinos

Source: Gates Millennium Scholar Surveys: Cohort III.
Notes: The vertical lines indicate the respective cut points for each ethnic group for the Gates Millennim Scholarship program
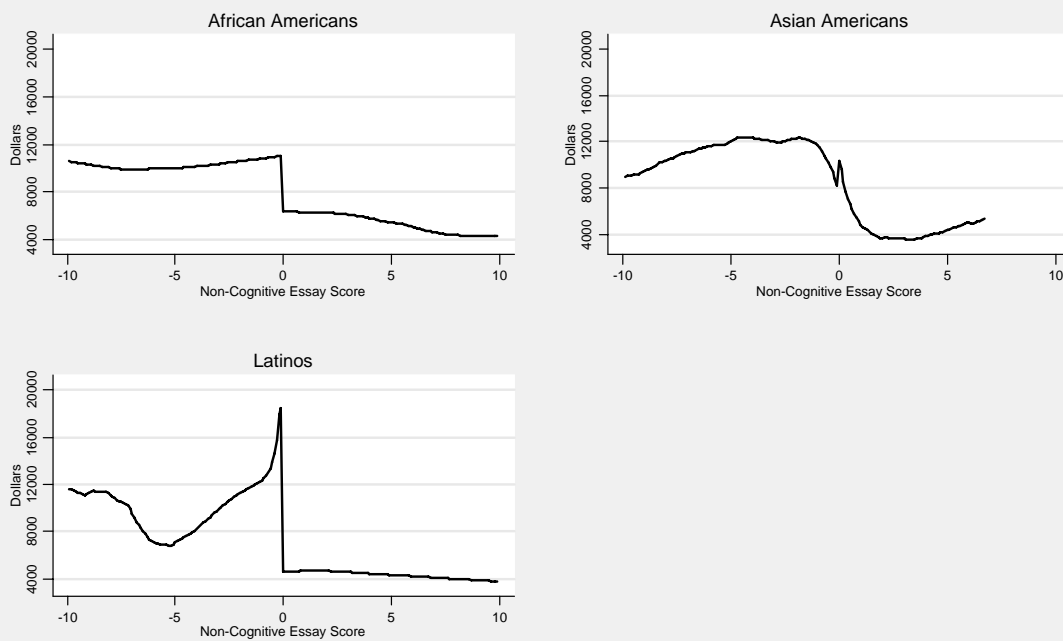
Figure 6
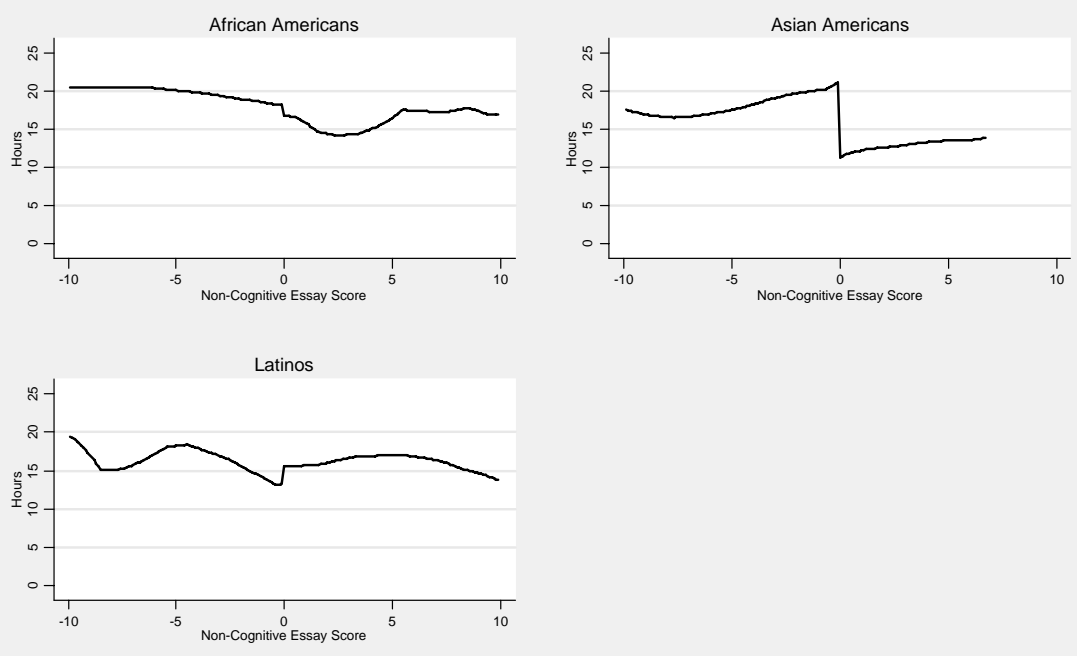## Fraction Enrolled in College: Fall 2004
Cohort III

Source: Gates Millennium Scholar Surveys: Cohort III
Notes: 1. Estimates based on local linear regression using optimal bandwidths.
2. Non-cognitive essay score measured as deviation from cut point.

Figure 7
Fraction of Individuals who are Gates Millennium Scholars
Cohort III

Source: Gates Millennium Scholar Surveys: Cohort III
Note: 1. Estimates based on local linear regression using optimal bandwidths.
     2. Non-cognitive essay score measured as deviation from cut point.

Figure 8
Accumulated Debt from Student Loans: Junior Year
Cohort III

Source: Gates Millennium Scholar Surveys: Cohort III
Notes: 1. Estimates based on local linear regression using optimal bandwidths.
2. Non-cognitive essay score measured as deviation from cut point.

Figure 9
Hours Worked per Week : Junior Year
Cohort III

African Americans

Asian Americans

Latinos