Evaluation of a Program to Help Minorities Succeed at College Math: U. C. Berkeley's Professional Development Program

Steven Chin[a]
Maoyong Fan[b]
Guido Imbens[c]
Jeffrey M. Perloff[d]

May 2006

[a] U.C. Berkeley Professional Development Program Director
[b] Graduate Student, Department of Agricultural & Resource Economics, U.C. Berkeley
[c] Professor, Department of Agricultural & Resource Economics and Department of Economics, U.C. Berkeley; and Professor of Economics, Harvard University
[d] Professor, Department of Agricultural & Resource Economics, U.C. Berkeley

Note: We thank the various members of the U.C. Berkeley Committee on Student Diversity and Academic Development and particularly Caroline Kane for advise and helpful comments

Contact:
Jeffrey M. Perloff (perloff@are.berkeley.edu)
Department of Agricultural & Resource Economics
207 Giannini Hall — MC 3310
University of California
Berkeley, CA 94720-3310

Table of Contents

Evaluation of a Program to Help Minorities Succeed at College Math: U. C. Berkeley's Professional Development Program

The University of California, Berkeley's Professional Development Program (PDP) was founded in 1974 to promote academic excellence and diversity for underrepresented students in mathematics, engineering, and the sciences throughout the educational pipeline from middle school through graduate school.[1] Does the program work? If so why? Is it because it provides extra training, uses innovative teaching techniques, creates a sense of community among participants and employs superior teachers, or benefits from sample selection?

To determine whether the program has an effect on the academic outcomes for students taking calculus at U.C. Berkeley, we use a variety of matching and other statistical techniques designed to compensate for a lack of a randomly assigned, controlled experimental design.

In the first section, we describe the various ways that students at Berkeley are taught calculus. Next, we lay out our four hypotheses as to why students in PDP might perform better than other students. In the third section, we describe our data set and present summary statistics. We use a matching technique to compare PDP students to students in other programs with similar (ideally identical) characteristics in the fourth section. In the fifth section, we compare our matching results to those from traditional analyses. Finally, we conclude that PDP is effective in raising performance in calculus and discuss each of our hypotheses about why it is effective.

---

[1] Over time, the population served by PDP has changed, with the removal of race or ethnicity as criterion for targeting services. Today for example, black students are a smaller part of the U.C. Berkeley student body, and a smaller portion of PDP students than in the 1970s and 1980s.

Various Approaches to Teaching Calculus at Berkeley

Math 1A and 1B are the gateway calculus courses for students in mathematics, physical sciences, engineering, and in some social sciences. In Math 1A and 1B, a professor gives three hours of lectures per week to up to 200 students. A student has the choice of taking calculus from two or three professors each semester. Each professor gives independent homework assignments, midterms, and finals. (When we compare students' grades, we always control for the particular class that students attend.)

Students can obtain additional instruction from their graduate student instructor (GSI) or from the professor during office hours. In addition, formal discussion sections led by GSIs are provided. In our sample, there are four different types of discussion sections. We start with the traditional discussion sections offered by the math department. Then we discuss the workshop-style sections offered by the Professional Development Program (PDP), the Student Learning Center (SLC), and (in the second half of our time period, 1997-2001) by the mathematics department.

Traditional Mathematics Department Discussion Sections

Traditionally, students attend a 50 minute discussion sections twice a week in which a mathematics department GSI provides mini-lectures, answers questions, and conducts quizzes and exams. Each section serves up to 25 students. Usually, the GSI lectures from the front of the room for most of the period.

Most students attend sections under the auspices of the mathematics department. However, some students attend alternative sections provided by PDP while other students obtain additional training from the SLC designed to supplement the mathematics department's sections.

PDP Intensive Discussion Sections

Between 1978 and 2003, 2,565 undergraduates participated in PDP calculus programs. Since 1988 (and during the entire period of our study, which starts in 1993), the PDP students attend the same lectures as other students, but they attend intensive workshop-style discussion sections. PDP works closely with from six to ten mathematics graduate student instructors, each semester, who use new teaching techniques designed to foster collaborative learning employing innovative curriculum materials in small study groups, limited to no more than 25 students.

Where mathematics department GSIs hold sections for 50 minutes twice a week, PDP GSIs conduct section meetings two or three times a week for up to 80 minutes each. PDP students receive an additional unit for attending these longer, intensive discussion sections and a pass or no pass grade.

Typically, the GSI spends a few minutes discussing the key issues to be studied that day. Then, the GSI hands out worksheets, students go to the blackboards and work in small groups to solve the problems. The GSI moves from group to group providing hints and advice as to how to solve the worksheet problems (which are often more difficult that homework assigned by the professor). Mini-lectures and quizzes are included as needed. Reviews before midterms and finals are held in the evening hours, with pizza and soda provided for the students.

In addition, an undergraduate tutor is assigned to each intensive section to help tutor students in their small groups and in additional office hours. Intensive discussion sections are held in PDP's facility on campus, which also houses study areas, computer workstations, a student lounge, a math curriculum library, and the advising staff. These additional amenities provide a familiar place for students to gather and socialize as friends. In these ways, PDP strives to help students form a community in which academic success is encouraged and rewarded.

SLC Adjuncts to Discussion Sections

The Student Learning Center (SLC) offers adjunct group-study courses that are linked to specific math lecture sections. These "adjuncts" were designed to help students improve study strategies, critical thinking, and exam-taking skills. Students regularly meet three hours a week to work on practice exams or additional problems; learn about time management; and work with peers to solve math problems. Upper division undergraduate or graduate students on staff lead the adjunct sections. Students receive one pass/not pass unit for attending the adjunct course. Students attend adjuncts in addition to their regular math department lecture and discussion section. Each adjunct serves up to 25 students. Unlike PDP, SLC does not provide an additional undergraduate tutor for each section, nor does it try to create a community among the students by providing meeting places and various inducements to meet out of class.

In addition, the SLC offers drop-in tutoring from upper-division undergraduates who could provide additional help for students. These additional resources are used on an as-needed basis: Most students who use the SLC tutor do so just before midterms and finals.

Math Workshop Discussion Sections in the Math Department

Given the apparent success of the PDP, the mathematics department switched to workshop–style discussion sections midway through our period of study. The math workshop discussion sections offered in the department of mathematics are variations on the PDP workshop model. Before the department of mathematics introduced these new sections, it offered discussion sections where classroom instruction was comprised of question and answer sessions, mini-lectures, and quizzes with the GSI in the front of the class for most of the twice weekly 50 minute discussion sections. In contrast, the mathematics workshop discussion sections meet twice a week for 80 minutes. The GSI assists up to 25 students with the completion of a worksheet of calculus problems that the students work on in small groups. The GSI circulates among the groups, giving advice as needed to the students. In addition, mini-lectures and quizzes are used as needed. Students in both models of discussion sections also attend math lectures given by the professor. In contrast to the PDP Intensive Discussion Sections, the workshop-style discussion sections are held in regular classrooms in the math department building with little or no opportunity for students to gather and socialize informally. Unlike PDP, no undergraduate tutors are available for the sections and there is no attempt to create a community among students.

Hypotheses

We examine four hypotheses for why students in PDP may perform better than those in the traditional mathematics sections. First, PDP provides longer section meetings and individual tutoring. Second, PDP uses an innovative workshop approach to teaching calculus, which is more effective than traditional lectures sections. Third, PDP creates a community for its students and it employs superior GSIs. Fourth, due to self-selection, better motivated students participate in PDP.

We examine the first hypothesis concerning the role of longer section meetings and individual tutoring by comparing mathematics department sections to PDP and SLC sections, which provide additional training. In addition, for each section, PDP assigns an undergraduate tutor who is available to meet students individually. SLC provides a drop-in tutor center. Thus, if this hypothesis is correct, we expect both PDP and SLC students to perform better than those in neither program.

The second hypothesis, that the workshop-style sections are superior, can be examined by mathematics department sections using the traditional approach to those using a workshop style. Both PDP and SLC use a workshop approach, so their students can be compared to those in the traditional sections as well.

We examine the third hypothesis, that the community created by PDP for its students increases performance, by comparing PDP to SLC students. Both PDP and SLC students participate in extra long sections, use a workshop approach, and provide extra tutoring. However unlike SLC, PDP creates a sense of community and makes an effort to hire superior GSIs.

We have two modeling approaches to dealing with the fourth, creaming hypothesis, as we explain in the next section in more detail. In one method, we try to match students in PDP to others so well that we do not have a creaming problem. In the other method, we test for and compensate for a possible sample selection bias.

Data and Summary Statistics

We have data on 28,645 Berkeley undergraduates from 1993 through 2001, of which 17,886 records were complete and usable. The data set includes students' demographic characteristics, which classes they took, their academic performance on math courses, and other measures of performance at Berkeley.

Table 1 shows how many students were in each racial and ethnic group and the average mathematics SAT score and high school grade point average (GPA) by group. We focus on a group of minority students that we label BHNA: blacks, Hispanics (Mexican/Mexican-American and other Spanish-American), and Native Americans (American Indians and Alaskan Natives). Over the decade of our study period, these minorities were 14.2% of all students, but 78.3% of PDP students, and 56.2% of SLC students.

Non-BHNA students had higher average mathematics SAT scores than BHNA students (700 versus 578) and higher high school GPAs (4.17 versus 3.78). On average, BHNA students in the PDP program had higher mathematics SAT scores (599) than those in SLC (500) or those who were in neither program “others” (583). Their high school grades followed the same pattern. Non-BHNA students in PDP and SLC had slightly lower SATs and high school GPAs than did others.

Table 2 shows how performance in Mathematics 1A (the first course in the two-course calculus sequence) varied by racial and ethnic group. Some students take the courses for a letter grade while others receive a “pass/no pass” grade. We use three measures of grades. First, for those students who receive a letter grade, we convert the letter grade into a cardinal index, where an F is 0, D- is 1, D is 2,…, and A+ is 12. Second, for all students, we record a zero-one variable that equals one if a student passes: receives a “pass” if taking the course pass/no pass or a C- or above if taking the course for a letter grade. Third, for students who receive a letter grade, we set a zero-one variable equal to one if the student received at least an A-.

Minorities (BHNA) receive lower grades on average than do non-minorities. However, by any of our three measures, minorities and non-minorities generally do better in PDP and SLC than otherwise. BHNA students in PDP averaged 6.5 on our 12 point scale, those in SLC had a 5.8, and those in neither program had a 4.8 average. Moreover, 17% of BHNA students in PDP and 18% in SLC received at least an A-, compared to 8% of the others.

Matching

We first consider a matching technique that allows us to simulate a controlled experiment after the fact. Intuitively, comparing the outcome for two individuals with the same predetermined characteristics where one is treated and the other is not, is analogous to comparing two individuals in a randomized experiment. We match the performance of students in PDP and SLC to similar students who are in neither program ("others"). We compare the performance of students in PDP to others, those in SLC to others, and PDP to SLC students.

Matching Methodology

We use a method where we compare subjects in a treatment group to those in a control group (Abadie and Imbens, 2002, and Abadie, Drukker, Herr, and Imbens, 2002). PDP or SLC is the "treatment" that only some students receive. Students in neither program are in the control group.

Let $\{Y_i(0), Y_i(1)\}$ denote the two potential outcomes (e.g., grades) of individual *i*, where $Y_i(1)$ is the outcome of individual *i* if exposed to treatment and $Y_i(0)$ is the outcome of individual *i* if not exposed to treatment. For each student in a given class, we observe only one outcome,

$$Y_i = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1, \end{cases}$$

where $W_i \in \{0, 1\}$ indicates whether or not individual *i* is in the treatment group. Because we observe only one result (either treated or untreated outcome), we need to find a way to estimate the other outcome without bias. The idea of the matching methodology is to estimate the counterfactual result by using average outcomes of similar individuals in the corresponding (treatment or non-treatment) group.

We match each individual to a number of similar individuals with respect to observed pretreatment characteristics. The optimal number of matching to each individual depends on specific problem. Then we can estimate the sample average treatment effect by averaging within-match differences in the outcome variable between the treated and the untreated individuals. The sample average effect is:

$$\tau = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i(1) - \hat{Y}_i(0)\right)$$

where N is sample size.

One of the crucial steps in this approach is to match a given individual to others with nearly identical characteristics who were treated differently. We use the observed characteristics of individuals to create our comparison groups. For example, if we wanted to evaluate the effect of a program on this semester's GPA, we might use last semester's GPA as an important matching covariate. Similarly, we might compare people with the same race, gender, or high school GPA and SAT scores.

An algorithm is used to match people in terms of the "distance" between their various characteristics. We choose a vector norm to decide who the closest neighbors to individual i are. Let $\|x\|_V = (x'Vx)^{1/2}$ be the vector norm with positive definite matrix V^2, we define $\|z - x\|_V$ as the distance between the vector x and z, where z represent the covariates for a potential match for individual i. Let $d_M(i)$ be the distance from individual i to the M^{th} nearest match with the opposite treatment. Consider the set of observed covariates for an individual i to be X_i, the set individual i will match with is

$$\Psi_M(i) = \left\{ l = 1,\ldots,N \,\middle|\, W_l = 1 - W_i, \|X_l - X_i\|_V \le d_M(i) \right\}$$

and $d_M(i)$ is defined as

---

2 We use the diagonal matrix, of which the diagonal elements are the inverses of the variances of X_i (the element of the set of covariates), as our weighting matrix V. The weighting matrix V accounts for the difference in the scale of the covariates.

$$\sum_{l:W_l=1-W_i} 1\left\{\left\|X_l - X_i\right\|_V < d_M\left(i\right)\right\} < M$$

and

$$\sum_{l:W_l=1-W_i} 1\left\{\left\|X_l - X_i\right\|_V \le d_M\left(i\right)\right\} \ge M$$

where $1\{\cdot\}$ is the indicator function, which is equal to 1 when the value in brackets is true and zero otherwise.

The simple matching estimator will be biased in finite samples when the matching is not exact. Although theoretically matching on multidimensional covariates can lead to substantial bias, the matching approach combined with bias adjustment often leads to estimates with little remaining bias.[3] Abadie and Imbens (2002) show that the estimators will have a term corresponding to the matching discrepancies (the difference in covariates between matching individuals and their matches) that will be of the order $O_p(N^{-1/k})$ with k continuous covariates.

The bias-corrected matching estimator adjusts the difference within the matches for the differences in their covariate values. The adjustment is based on an estimate of the two regression function: $\mu_\omega(x) = E\{Y(\omega)| X = x\}$ for $\omega = 0$ or 1. The regression functions are approximated by linear functions and estimated using least squares on the matched observations. If we are estimating sample average treatment effect, we could estimate the regression functions using only the data in the matched sample

$$\hat{\mu}_\omega\left(x\right) = \hat{\beta}_{\omega 0} + \hat{\beta}'_{\omega 1} x$$

for $\omega = 0$ or 1, where

[3] We use the nnmatch command in STATA, which allows each observation to be used as a match more than once.

$$\left(\hat{\beta}_{\omega 0}, \hat{\beta}_{\omega 1}\right) = \arg\min_{\{\beta_{\omega 0}, \beta_{\omega 1}\}} \sum_{i:W_i=\omega} K_M(i)\left(Y_i - \hat{\beta}_{\omega 0} + \hat{\beta}'_{\omega 1}x\right)^2$$

where $K_M(i) = \sum_{l=1}^{N} 1\left\{ i \in \Psi_M(l) \frac{1}{\#\Psi_M(l)} \right\}$, which is the number of times individual *i* is used as a match for all observations *l* of the opposite treatment group, each time weighted by the total number of matches for observations *l*.

We weight the observations in these regressions by $K_M(i)$, the number of times the unit is used as a match, because the weighted empirical distribution is closer the distribution of covariates in which we are ultimately interested. For this reason, we use the matched sample only in this step. Given the estimated regression functions, for the bias-corrected matching estimator we predict the missing potential outcomes as:

$$\tilde{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \dfrac{1}{\#\Psi_M(i)} \displaystyle\sum_{l \in \Psi_M(i)} \left(Y_l - \hat{\mu}_0(X_i) - \hat{\mu}_0(X_l)\right) & \text{if } W_i = 1 \end{cases}$$

and

$$\tilde{Y}_i(1) = \begin{cases} \dfrac{1}{\#\Psi_M(i)} \displaystyle\sum_{l \in \Psi_M(i)} \left(Y_l - \hat{\mu}_1(X_i) - \hat{\mu}_1(X_l)\right) & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1 \end{cases}$$

with the corresponding bias-corrected estimator for the average treatment effect

$$\hat{\tau}_M^{bcm} = \frac{1}{N} \sum_{i=1}^{N} \left[\tilde{Y}_i(1) - \tilde{Y}_i(0)\right].$$

To choose the covariates used in matching individuals, we first estimate a linear model with the relevant measure (e.g., course grade) as the dependent variable. The independent variables include PDP and SLC status (as relevant), gender, race, ethnicity, whether the student was in the coalition pool, whether the student was in Berkeley's Biology Scholar's Program

(BSP), verbal SAT score, mathematics SAT score, economic opportunity (EOP) status, high school grade point average, whether one's native language is English, whether one's father has a graduate degree, whether or one's mother has a graduate degree, dummy variables for the student's initial major, and math class dummies.[4]

We use the regression results to select three sets of variables: "exact match" variables (those for which we require a perfect match), "other match" variables (those for which we will accept a close, but imperfect match), and "bias adjustment" variables. The variables for which we require an exact match are the most important characteristics. We always require an exact match on race, ethnicity, and the math class dummy for the course that a given student is taking (so that all students' grades are comparable).[5] The other exact match candidates are those variables with t-statistics above 10 in the regression: typically, mathematics SAT score or high school GPA. The "other match" variables include all those for which we can reject the null hypothesis that the coefficient is zero at the 5% level based on a two-tailed t-test.

Because the sample matching estimator will be biased in finite sample when the matching is not exact, as we have already noted, we need to adjust for the bias. To do so, we use the bias adjustment variables, which are all variables with t-statistics larger than 4. Thus, the set of exact match variables and bias adjustment variables overlap.

[4] The coalition pool includes applicants to Berkeley who are minorities and those who come from low-income families or families in which neither parent attended college. The coalition students who are admitted at the beginning of the Fall semester receive a letter inviting them to join PDP. Berkeley undergraduates who are interested in majors related to biology can apply for BSP at the beginning of each semester. This program is intended to help those students who are in the coalition pool.

[5] In STATA's nnmatch command we impose the requirement of an exact match by multiplying the corresponding elements in the weight matrix by 1,000 relative to the wrights placed on the elements of ordinary covariates.

To ensure that this matching estimator is robust to our regression specification, we conduct robustness checks by trying other exact and other match variables. Our objective is to find the most similar person or persons in the other group based on their predetermined characteristics. The larger the number of variables that we use for matching, the less accurately we are likely to match on any given variable for which we do not require an exact match. Therefore, we consider two alternative approaches. In one, we used the criterion that the other matching variables group had t-statistics greater than 4 (rather than 1.96). The second alternative is the same as the first except that we put variables with very high t-statistics (e.g., high school GPA) in the other matching group instead of in the exact matching group.

For all three of these specifications, we match either one or four ($M = 1$ or 4) people from the other treatment group. Thus, because we have three specifications and match either one or four people, we have six estimation approaches.

For most analyses the six approaches produce similar results. We report only one of these six estimates. To choose among them, we examine how comparable are the matched groups. We use the following algorithm to select an approach.

We start by rejecting any approach in which we cannot exactly match the "exact match" variables. Next, we calculate a statistic that we use to select among the remaining approaches. For each "other match" variable, we calculate a ratio: the difference in means between the control and treatment group under any given approach divided by the difference in means of the two groups for the full sample. We then divide this ratio by the standard deviation within the entire sample for this variable to obtain our test statistic. If this test statistic is greater than 0.5 for any "other match" variable, we reject a given approach. Then, we compare the test statistics for all remaining approaches for the "other match" variable with the highest t-statistic in the original OLS regression and pick the approach which has the lowest test statistic.

Comparing Students in PDP or SLC to Others

We start by measuring the sample average treatment effects of PDP or SLC students relative to students in neither program (controls or "others") for both Math 1A and Math 1B. The top third Table 3 shows that students in PDP average higher grades than do similar students in neither program. The first two columns show that the PDP treatment effect on letter grades in both Math 1A and Math 1B range from 1.41 to 1.99. (A difference of one would raise a grade from, say, a B to a B+, whereas a difference of 2 would raise it from a B to an A-.) All the differences are statistically significantly different than zero at the 5% confidence level based on a two-tailed t-test.[6]

The last two columns show the sample average treatment effects of SLC on letter grades. The SLC treatment has a statistically significant beneficial effect on minority (BHNA) students, but not on non-minority students. The average effect on minority students is 1.49 for Math 1A and 2.4 for Math 1B.

Alternatively, we can examine the effect of these programs on whether students pass or fail. The sample average treatment effects are in the middle third of Table 3. The first two

[6] The matches are based on draws from the population with replacement. The PDP entire sample Math 1A treatment effect is calculated based on observations for 618 PDP and 7,547 students in neither program. The mean number of times that a PDP student is used as a match is 12.5 (minimum is 0 and the maximum is 490), whereas the mean for the control group is 0.08 (minimum is 0 and the maximum is 8). The comparable figures for the minority PDP Math 1A calculation is 504 PDP observations, 612 students in neither program, with the mean number of matches 1.2 for PDP and 0.8 for other (for both, the minimum is 0 and the maximum is 11). Finally, for the non-minority PDP Math 1A calculation, we used 112 PDP observations and 6,942 students in neither program. The mean number of matches is 63.5 (minimum 2, maximum 484) for PDP and 0.02 (minimum 0, maximum 2) for other.

columns show the sample average treatment effects of PDP. All six treatment effects are positive and statistically significant at the 5% level. Minorities in PDP are 15 to 18% more likely to pass than are minorities in neither program. The average PDP effect is more than twice as large for minorities as for non-minorities. The third and fourth columns show the sample average treatment effects of SLC. For minorities, the treatment effects are positive and statistically significant at 5% level and close to the PDP effects. For non-minorities, our SLC matching estimators are not statistically significant and the point estimates are virtually zero.

Finally, we examine whether these programs can help students get top grades (A- or above) in the bottom third of Table 3. For PDP, all but one of the six coefficients are statistically significantly different from zero (the exception is for non-minorities in Math 1B). Minorities in PDP are 9% more likely to receive a top grade in Math 1A than are others. Non-minorities are 23% more likely in Math 1A and 26% more likely in Math 1B to receive a top grade. SLC had no statistically significant effect for Math 1A but had very large effects for Math 1B: 37% higher for minorities and 18% for non-minorities.

Thus, both PDP and SLC students perform better than others. These results are consistent with the hypothesis that the extra training and tutoring both provide are effective in raising grades.

Comparing PDP and SLC Students Directly

We also compared PDP to SLC students using the same three performance measures. As Table 4 shows, none of the differences are statistically significantly different from zero for the numerical grades or the top grades (A-/A/A+), and only half of the differences for the passing measure are statistically significant. Apparently there is no statistically significant difference in the probability of passing for minorities, but non-minority PDP students have a statistically significant 11% greater probability of passing than SLC students. Thus, there is only limited evidence for the hypothesis that the superior performance of PDP students is due to their community or superior PDP GSIs. However, these comparisons are based on much smaller samples than in our other matching analyses.[7]

Workshop Style Sections

One hypothesis for the reason that PDP and SLC produce superior results is their use of workshop style sections. The mathematics department began to provide workshop style sections similar to PDP intensive discussion for Math 1A and Math 1B in the spring semester, 1997. Table 9 reports sample average treatment effects of mathematics department workshop style sections using observations after Spring 1997. The grades and the probability of receiving an A-/A/A+ are statistically significantly different from zero, while the probability of passing coefficients are not statistically significantly different from zero. Three of the four effects are negative. These results indicate that students do better in the traditional section than in the workshop sections when conducted by the mathematics department.

Other Matching Experiments

We now briefly summarize four other matching experiments. First, to determine if the introduction of the workshop style sections by the mathematics department half way through the observation period biased our basic results in Table 3, we divided the time period in half—before and after the mathematics department introduced the workshop-style sections—and repeated the analyses for each subperiod. We found no systematic differences between the periods and continued to find large positive differences in performance for PDP students compared to others.

[7] For example, for Math 1A matching, we have only 567 PDP students and 153 SLC students.

Second, we repeated the analysis in Table 3 for four demographic groups: blacks, Hispanics, Asians, and whites. The matching results in subgroup of blacks and Hispanics are consistent with those of our BHNA results in Table 3. And the matching estimators of subgroup Asian and whites are consistent with those with our non-BHNA analysis. In short, the results for the BHNA and non-BHNA groups hold for their subgroups.

Third, we investigated whether PDP and SLC had long-lasting treatment effects on participants' school performance in subsequent years. In particular, did participants have higher junior year GPAs or were they more likely to graduate in four years or less? According to our matching analysis, neither program had a statistically significant for either of these long-term criteria. For the minorities, the PDP treatment effect for graduating within four years was -0.64 (asymptotic standard error, 0.40) and the effect on the junior-year GPA was -0.031 (0.023). The results for other groups were similarly statistically insignificant.

Fourth, we examined whether PDP participants who initially majored in mathematics, physics or engineering were more likely to remain in those majors. Based on our matching analysis, we do not find statistically significant effects. For minorities, the treatment effect on the probability of remaining a math, physics, or engineering major was 0.015 (0.013). We also repeated our basic analysis using only engineering students and found similar results to those in Tables 3 and 4.

Other Statistical Analyses

We believe that our matching results are the most reliable way to analyze the effects of the PDP (and SLC) experiments. However, for completeness, we conducted two other, traditional analyses. First, we use ordinary least squares and an instrumental variables technique to estimate the effect of these programs on grades and test for creaming or endogeneity. Second, we use an ordered probit analysis so that we could treat grades as ordinal and not cardinal.

Ordinary Least Squares

For comparison purposes, we report in Table 6 ordinary least squares regressions of grades measured as a cardinal 0-12 index on various characteristics of students in Math 1A and in Math 1B. The explanatory variables are race and ethnicity, high school performance (grade point average, GPA, and SAT scores), major upon entering Berkeley, whether a student is in PDP or SLC, whether a student took a mathematics department workshop section, a 1997 year dummy (to capture to capture any change between the first and second halves of our sample unrelated to the mathematics department's introduction of workshop sections), whether the students parents graduated from college, and a dummy variable for each class (which are not reported to save space).

For Math 1A, the ordinary least squares estimate of the PDP effect is 1.75 (standard error, 0.15) and the SLC effect is 1.01 (0.24). Thus, for both programs, we can reject the null-hypothesis of no effect at the 0.05 level. The corresponding matching effects for the total sample in Table 3 are 1.75 (0.23) for PDP and 0.39 (0.30) for SLC.

For Math 1B ordinary least squares estimates are 1.30 (0.17) for PDP and 1.26 (0.32) for SLC. The corresponding matching estimates are 1.67 (0.33) for PDP and 0.04 (0.574) for SLC. Thus, the ordinary least squares and matching results are similar for PDP, but not for SLC.

In the matching technique, we try to compare such similar students that sample selection is not a problem. One might worry, however, that sample selection is a problem in the ordinary least squares regression. After all, students choose to be in PDP.

Thus, we re-estimated our equations using instrumental variable, where we treat being in the PDP program as endogenous and add three additional instruments: EOP status (whether the student is from a poor family), whether the student is in the Biology Scholars Program (BSP), and whether the student was in the coalition pool. Applicants in the BSP or/and coalition pool are women and minorities and those who come from low income families or/and families without a parent who attended college. The coalition students who are admitted at the beginning of the Fall semester receive a letter inviting them to join PDP (hence are more likely to be in PDP). The Berkeley undergraduates who are interested in majors related to biology could apply for BSP at the beginning of each semester.[8]

The instrumental variables estimates for PDP are larger: 2.49 (with an asymptotic standard error of 0.56) for Math 1A and 2.66 (0.92) for Math 2A. However, based on Hausman-Wu test statistics (the chi-square test statistics are 2.13 for Math 1A and 1.36 for Math 1B, each with 69 degrees of freedom), we cannot reject the null hypothesis of exogeneity for either course (i.e., the ordinary least squares estimates are consistent and efficient).

Ordinal Analysis

So far, our main measure of grades has been a cardinal index. We now consider using an alternative approach where we assume that the grades are only ordinal and estimate using an ordered probit model. Because not all 13 grade categories that we used before are observed in

---

[8] We estimated a probit equation for whether one is in PDP, where the explanatory variables are gender, SAT scores, high school GPA, ethnicity, race, whether the students' parents have graduate degrees, whether English was the student's native language, EOP status, whether one is in BSP, and whether one was in the coalition pool. For Math 1A, this equation correctly predicts PDP status with 93% accurately. It predicts that 6,918 students were not in PDP out of the 7,278 student who were not, and correctly predicts 207 students were in PDP out of the 352 who actually were in the program.

each class, we collapse the grades into four categories starting with the worst outcome, fail (F to D+, which is coded 0), then C- to C+ (coded 1), B- to B+ (coded 2), and A- to A+ (coded 3).

Table 7 shows the estimated ordered probit. We estimate a latent variable such that the grade switches from one category to another when the latent variable exceeds 0, μ(1), and μ(2). Students in PDP and SLC do measurably better than other students controlling for personal characteristics. The PDP coefficient is substantially larger than the SLC coefficient for Math 1A, but the two effects are virtually identical for Math 1B. Students in the mathematics department's workshop sections do statistically significantly worse in Math 1B, but not in Math 1A. We cannot reject the hypothesis that this dummy's coefficient is zero at the 5% level.

Table 8 shows how the distribution of grades would change for the average student as we change one variable at a time. For dummy variables, the change in the distribution is calculated as the difference in the estimated probabilities with the dummy variable equal to one and zero. For continuous variables, the change is the derivative.

As the table shows, a Math 1A student in PDP (SLC) has a 18.7% (11.9%) higher probability of getting an A and a 10% (6.4%) lower probability of failing. In Math 1B, the PDP (SLC) students have a 14.4% (15.0%) higher probability of an A and a 8.4% (8.7%) lower probability of failing. A mathematics department workshop section student has a 23.9% lower probability of getting an A in Math 1B.

Conclusions

We have used a variety of traditional and new techniques to analyze the effects of the University of California at Berkeley's Professional Development Program (PDP) and the Student Learning Center's (SLC) corresponding program. Both programs are clearly effective in improving the performance of minority and other students in basic calculus courses.

Unfortunately, according to our estimates, these programs are not effective in keeping students in mathematics, physics, and engineering; raising their grades in later years; or raising the probability of graduating within four years.

We formulated four hypotheses to explain why PDP was effective in raising grades. First, PDP and SLC provide longer section meetings and individual tutoring. Both are successful in raising grades relative to shorter mathematics department sections (where no individual tutoring is provided).

Second, PDP and SLC introduced an innovative workshop approach to teaching calculus, which is thought to be more effective than traditional lectures sections. However, when the mathematics department introduced such sections half way through our sample period, students in the workshop sections did no better and may have done worse than students in traditional sections. One possible explanation for this result is that PDP and SLC provide longer workshop sections and that this extra time is necessary.

Third, PDP creates a community for its students and it employs superior graduate student instructors (GSIs). The SLC program lacks these features. To the degree that PDP has superior results to SLC, this difference may be due to the community effect and the superior GSIs.

Fourth, due to self-selection, better motivated students participate in PDP. We used techniques to eliminate or control for sample selection. Our formal tests found no evidence of sample selection bias.

References

Abadie, A., and G. Imbens, 2002, *Large Sample Properties of Matching Estimators for Average Treatment Effects,* Department of Economics, University of California, Berkeley

Abadie, A., D. Drukker, J.L. Herr, and G. Imbens, 2002, *Implementing Matching Estimators for Average Treatment Effects in STATA*, Department of Economics, University of California, Berkeley

Greene, W. H., 1995, *Limdep Users' Manual*, version 7.0, Econometric software, Inc.

Table 1
Summary Statistics

		Percentage of			*Mean Math SAT*				*Mean High School GPA*			
	Number	*PDP*	*SLC*	*All*	*PDP*	*SLC*	*Other*	*All*	*PDP*	*SLC*	*Other*	*All*
BHNA	2,537	78.3	56.2	14.2	599	500	583	578	3.87	3.55	3.79	3.78
Black	731	24.5	22.9	4.1	571	506	554	552	3.70	3.39	3.60	3.60
Hispanic	1695	51.3	32.0	9.5	611	493	590	586	3.95	3.65	3.86	3.86
American Indian/ Alaskan Native	111	2.6	1.2	0.6	628	580	637	632	3.77	3.83	3.79	3.79
Non-BHNA	15,349	21.7	43.8	85.8	658	623	702	700	4.14	4.04	4.17	4.17
White	4293	5.9	8.8	24.0	678	638	692	691	4.14	4.04	4.14	4.14
Asian	8084	8.9	25.1	45.2	644	624	710	708	4.07	4.03	4.19	4.18
Other	2972	6.8	10.0	16.6	660	608	691	689	4.23	4.05	4.18	4.18
All	17,886	100.0	100.0	100.0	612	554	691	682	3.93	3.76	4.14	4.11

Notes: BHNA includes blacks, Hispanics, and Native Americans. Native Americans are American Indians and Alaskan Natives. Hispanic includes Mexican/Mexican-American and Other Spanish-American. Others are other Asians, other ethnic groups, and those who declined to state their ethnic background.

Table 2
Math 1A Outcomes

	Letter Grades				*Pass Percentage*				*A-/A/A+ Percentage*			
	PDP	*SLC*	*Other*	*All*	*PDP*	*SLC*	*Other*	*All*	*PDP*	*SLC*	*Other*	*All*
BHNA	6.5	5.8	4.8	5.6	84	80	67	75	17	18	8	12
Black	6.3	7.0	4.3	5.3	85	92	64	75	12	23	7	10
Hispanic	6.7	5.3	4.9	5.7	84	76	68	75	19	17	9	14
American Indian/ Alaskan Native	6.4	4.7	4.7	5.3	80	67	67	71	15	0	6	8
Non-BHNA	8.0	7.4	7.3	7.3	95	93	89	89	34	29	27	27
White	8.3	6.8	7.0	7.1	97	92	88	88	33	24	25	25
Asian	8.3	7.8	7.5	7.5	96	94	90	91	41	33	29	29
Other	7.2	6.9	7.1	7.1	90	90	88	88	24	23	25	25
All	6.8	7.0	7.0	7.0	86	89	87	87	20	26	25	25

Table 3
Sample Average Treatment Effect of PDP or SLC Compared to Untreated Students

	PDP		*SLC*	
	Math 1A	*Math 1B*	*Math 1A*	*Math 1B*
Grades				
All	**1.75** (0.23)	**1.67** (0.33)	0.389 (0.298)	0.04 (0.574)
BHNA	**1.81** (0.21)	**1.41** (0.23)	**1.49** (0.55)	**2.35** (0.65)
Non-BHNA	**1.58** (0.26)	**1.99** (0.41)	0.42 (0.27)	-0.18 (0.57)
Probability of Passing				
All	**0.088** (0.033)	**0.071** (0.03)	**0.079** (0.018)	-0.073 (0.053)
BHNA	**0.151** (0.029)	**0.177** (0.031)	**0.141** (0.071)	**0.185** (0.046)
Non-BHNA	**0.069** (0.016)	**0.066** (0.028)	0.04 (0.027)	-0.029 (0.035)
Probability of A-/A/A+				
All	**0.195** (0.047)	**0.253** (0.065)	0.043 (0.049)	**0.182** (0.074)
BHNA	**0.094** (0.02)	0.054 (0.031)	0.124 (0.085)	**0.367** (0.127)
Non-BHNA	**0.230** (0.072)	**0.256** (0.078)	0.041 (0.053)	**0.178** (0.078)

Note: A coefficient is bold if we can reject the null hypothesis of no difference at the 0.05 level using a two-tail t-test.

Table 4
Sample Average Treatment Effect Comparison of PDP and SLC

	Grades		*Probability of Passing*		*Probability of A-/A/A+*	
	Math 1A	*Math 1B*	*Math 1A*	*Math 1B*	*Math 1A*	*Math 1B*
All	0.426 (0.391)	0.095 (0.407)	-0.019 (0.058)	**0.103** (0.049)	-0.001 (0.06)	-0.09 (0.077)
BHNA	0.609 (0.657)	-0.691 (0.664)	-0.056 (0.071)	0.023 (0.054)	-0.067 (0.081)	-0.138 (0.098)
Non-BHNA	0.428 (0.377)	0.725 (0.495)	**0.108** (0.034)	**0.106** (0.044)	0.056 (0.055)	0.071 (0.072)

Table 5
Sample Average Treatment Effect of Mathematics Workshop Style Sections

	Math 1A	*Math 1B*
Grades	**-0.930** (0.230)	**-0.670** (0.257)
Probability of Passing	0.036 (0.030)	-0.011 (0.013)
Probability of A-/A/A+	**0.139** (0.349)	**-0.129** (0.031)

Notes:
Because the Math Department started to provide a choice of section Spring, 1997, only data from that semester forward are included. The sample includes only students in Mathematics Department sections, where students in the workshop sections are the treatment group.

Table 6
Ordinary Least Squares Regression of Grades (0-12) on various Characteristics

	Math 1A		*Math 1B*	
	Coefficient	*Standard Error*	*Coefficient*	*Standard Error*
Constant	**-10.96**	0.91	**-8.99**	0.68
PDP	**1.75**	0.15	**1.30**	0.17
SLC	**1.01**	0.24	**1.26**	0.32
Math Dept Workshop	**-1.23**	0.52	-0.05	1.17
Male	-0.08	0.07	**-0.31**	0.06
Verbal SAT/100	**-0.10**	0.04	**-0.15**	0.04
Mathematics SAT/100	**1.57**	0.06	**1.20**	0.06
High School GPA	**2.27**	0.11	**2.50**	0.11
Black	-0.24	0.22	0.33	0.26
White	0.14	0.10	0.15	0.10
Hispanic	**-0.53**	0.15	-0.25	0.17
Asian	0.09	0.10	0.08	0.09
Mathematics or Physics	0.29	0.18	**0.68**	0.13
Business or Economics	**0.47**	0.24	**0.41**	0.24
Chemistry	**-0.52**	0.13	-0.23	0.12
Biology	-0.09	0.08	-0.09	0.08
Engineering	**-0.23**	0.10	**-0.23**	0.08
English	**-0.35**	0.08	**-0.28**	0.07
Father College Grad	**0.21**	0.07	0.09	0.07
Mother College Grad	0.00	0.08	**0.28**	0.08
1997	0.37	0.88	-1.69	1.22
$\bar{R}^2$	0.24		0.18	
Number of Observations	7,630		9,695	

Note: Class dummy variables are not reported.
White robust standard errors are reported.

Table 7
Ordered Probit Model

	Math 1A		*Math 1B*	
	Coefficient	*ASE*	*Coefficient*	*ASE*
PDP	**0.639**	0.063	**0.471**	0.070
SLC	**0.407**	0.084	**0.489**	0.108
Math Dept Workshop	-0.479	0.259	**-0.782**	0.357
Male	-0.034	0.027	**-0.095**	0.025
Verbal SAT	**-0.000**	0.000	**-0.001**	0.000
Math SAT	**0.006**	0.000	**0.005**	0.000
HS GPA	**0.846**	0.040	**0.895**	0.037
Black	-0.047	0.086	0.158	0.096
White	0.072	0.041	0.031	0.037
Hispanic	**-0.180**	0.059	-0.060	0.060
Asian	0.051	0.039	0.019	0.033
Mathematics or Physics	**0.142**	0.072	**0.244**	0.055
Business or Economics	**0.198**	0.092	0.141	0.103
Chemistry	**-0.182**	0.048	-0.063	0.043
Biology	-0.020	0.033	-0.025	0.030
Engineer	-0.060	0.040	**-0.078**	0.030
Mother Native English Speaker	**-0.137**	0.031	**-0.103**	0.027
Father College Grad	**0.073**	0.031	0.038	0.026
Mother College Grad	0.018	0.034	**0.095**	0.029
1997	-0.008	0.600	0.307	2.222
Constant	**-5.333**	0.578	**-4.748**	2.202
$\mu(1)$	**1.071**	0.021	**0.970**	0.017
$\mu(2)$	**2.155**	0.025	**2.003**	0.021

Notes: The class dummy variables are not reported.

Table 8
Ordered Probit Marginal Effects on Grades

	Math 1A				*Math 1B*			
	F or D	*C*	*B*	*A*	*F or D*	*C*	*B*	*A*
PDP	-0.100	-0.144	0.057	0.187	-0.084	-0.096	0.035	0.144
SLC	-0.064	-0.092	0.036	0.119	-0.087	-0.100	0.037	0.150
Math Dept Workshop	0.075	0.108	-0.043	-0.140	0.130	0.159	-0.059	-0.239
Male	0.005	0.008	-0.003	-0.010	0.017	0.019	-0.007	-0.029
Verbal SAT	0.000	0.0001	0.000	-0.000	0.000	0.000	0.000	-0.000
Math SAT	-0.001	-0.001	0.001	0.002	-0.001	-0.001	0.000	0.001
HS GPA	-0.133	-0.190	0.075	0.248	-0.159	-0.182	0.067	0.274
Black	0.007	0.011	-0.004	-0.014	-0.028	-0.032	0.012	0.048
White	-0.011	-0.016	0.006	0.021	-0.006	-0.006	0.002	0.009
Hispanic	0.028	0.041	-0.016	-0.053	0.011	0.012	-0.005	-0.018
Asian	-0.008	-0.012	0.005	0.015	-0.003	-0.004	0.001	0.006
Math/Physics	-0.022	-0.032	0.013	0.042	-0.043	-0.050	0.018	0.075
Business/Econ	-0.031	-0.045	0.018	0.058	-0.025	-0.029	0.011	0.043
Chemistry	0.029	0.041	-0.016	-0.053	0.011	0.013	-0.005	-0.019
Biology	0.003	0.005	-0.002	-0.006	0.005	0.005	-0.002	-0.008
Engineer	0.009	0.013	-0.005	-0.017	0.014	0.016	-0.006	-0.024
English	0.022	0.031	-0.012	-0.040	0.018	0.021	-0.008	-0.031
Father College Grad	-0.011	-0.016	0.007	0.021	-0.007	-0.008	0.003	0.012
Mother College Grad	-0.003	-0.004	0.002	0.005	-0.017	-0.019	0.007	0.029
1997	0.001	0.002	-0.001	-0.002	-0.054	-0.063	0.023	0.094