Using a Matched Sampling Methodology to Evaluate Program Effects: An Illustration From the University of California Outreach Programs

CSE Technical Report 598

Denise D. Quigley,
Jorge Muñoz and Alison Jacknowitz
CRESST / University of California, Los Angeles

May 2003

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein is supported by a grant from the University of California Office of the President.

The findings and opinions of this report are the authors' and do not necessarily reflect the conclusions reached by those directly responsible for the implementation, administration and management of the UC's K-12 Outreach efforts.

USING A MATCHED SAMPLING METHODOLOGY TO EVALUATE PROGRAM EFFECTS: AN ILLUSTRATION FROM THE UNIVERSITY OF CALIFORNIA OUTREACH PROGRAMS

Denise D. Quigley,

Jorge Muñoz and Alison Jacknowitz

CRESST/University of California, Los Angeles

Introduction

Policymakers and educators are committed to increasing the competitive eligibility of high school students applying to the University of California and to increasing the representation of poor and underrepresented minority students on UC campuses. The University of California's primary strategy to accomplish these goals is through School University Partnerships and academic development programs. Policymakers in conjunction with the University of California have invested substantial resources toward these common goals and programs. As a result, policymakers and educators want to know whether and to what extent the investment in School University Partnerships with supporting academic development programs is making a difference in the choices and lives of students. The University of California Office of the President (UCOP) has funded a number of research and evaluation efforts to investigate these questions.

This report is the first in a series that will evaluate the systemwide effects of School University Partnerships with supporting academic development programs on individual student-level achievement and student-level eligibility. The purpose of this report is to establish the feasibility of using a matched sampling methodology to investigate program effects and the methodology's ability to establish the matched samples of elementary, middle, and high schools whose progress on key indicators can be tracked over time. The purpose of this report is not to evaluate the effects of the program. At this time, evaluating the effects of the program is premature. Moreover, the evaluation of program effects should be coupled with information about the implementation of the program and its individual

components. This report establishes the methodology and the matched samples. It has three main sections. The first (Background) summarizes the history of the UC programs. The next section (Methodology) reviews the literature on matched sampling methods and outlines the strategy used for establishing the comparison groups. The following section (Illustrative Analyses) illustrates how differences in demographic characteristics and achievement outcomes can be assessed using the established comparison groups for the UC elementary, middle and high Partner schools.

Background

In 1995, the University of California Board of Regents adopted a policy to eliminate consideration of race, ethnicity, and gender in UC admissions. A majority of California's voters reinforced the Regents' decision with the passage of Prop. 209. While the University of California has had a long-standing commitment to educating California's diverse population, the rethinking of the affirmative action policy has brought a range of substantial changes to the way the University considers its responsibility to educationally disadvantaged students.

For many years, the University has been actively involved in providing academic enrichment programs that bridge secondary and post-secondary education. The programs have developed on each of the 10 University of California campuses and within the Office of the President, which serves as the administrative center of the University system. However since 1995, the challenges have been redefined themselves in light of the policy shift.

In particular, over the past several years, the University has implemented newly designed outreach efforts directed to K-12 schools that raise specific challenges with regard to education reform. The central research question related to these new efforts is: To what extent can the University act as a partner toward the improvement of educational outcomes in low-performing schools? Further, how does a whole-school improvement strategy converge, and even overlap, with the academic enrichment programs that are targeted to individual students in the same or neighboring schools? These newly designed strategies are known as UC's K-12 "outreach" activities.

The UC's K-12 outreach activities are an integrated effort of activities with the School University Partnership Program (S/UP) at its core. This S/UP program was

designed to assist schools with the development of systemic educational change, focusing on teacher training, educational leadership, and curriculum development. It was established in late 1998 following the recommendations of the Outreach Task Force of the University of California. The task force envisioned a set of Partner schools for each of the UC campuses at which educationally disadvantaged students would benefit from comprehensive reform strategies that would be launched, over time, by the schools in partnership with UC Outreach staff. Schools that consistently performed in the lowest two quintiles of educational achievement statewide on the state's standardized testing program (STAR system) were selected as UC Partners.

The S/UP program has been designed with the flexibility to accommodate the needs of schools and the strengths of campus programs. The implementation of S/UP consists of varied education reform strategies that are agreed to by the Partner school and University staff. In all cases, the program combines a set of educational interventions around teacher professional development, curricular reform, and the development of educational leadership. Some partnerships include specific new courses, tutoring, and technology-based initiatives. In all cases, the Partner schools are developing strategies to increase the college-going rates of students. The strategies vary by grade-span but seek to increase the educational achievement patterns that ultimately lead to successful completion of the University of California A-G admissions requirements. At some sites, UC programs have existed for many years, and the formalization of the partnership has resulted in coordinated services that include those programs with a history at the school site.

In most cases, the Partner school is supported by a site liaison, hired by the University. The liaison helps coordinate the services of the University campus with the senior administration of the Partner school and teachers as a way of aligning the school's program goals with the strengths of the University's services. An example might be the coordination of math professional development programs offered by the University with a school objective of offering algebra instruction to all students. In all cases, the liaison assists in the development and utilization of data sharing agreements that allow the Partner school and University to assess progress during the course of the year.

Additionally, the S/UP program integrates any other UC academic development programs that might be present at the Partner School. These academic development programs are:

Early Academic Outreach Programs (EAOP)—Programs designed to support academic enrichment and informational access for students interested in higher education. Programs range from early grades through high school and are generally targeted to educationally disadvantaged students.

Mathematics, Engineering, Science Achievement (MESA)—A student-based enrichment program designed to supplement educational achievement in mathematics and science, working with students in high schools across the state.

Puente—An intensive, multi-year literacy program in high schools that promotes the involvement of students, families and the community in the development of educational achievement and college-going aspirations.

Advancement via Individual Determination (AVID)—A comprehensive program that combines many components of academic enrichment and informational access with systematic curriculum improvement and professional development. It provides intensive student support in study skills, college student mentor-tutors, test preparation, college information, family involvement, and motivational activities. Programs range from middle school grades through high school and are generally targeted to educationally disadvantaged students.

California Subject Matter Projects (CSMP)—An extensive series of nine curriculum-based programs throughout the state that work with subject-specific teacher specialists in the development of teaching practices and curriculum development.

Partner schools may have all of these UC academic programs or have different mixes of them based on the needs of the Partner school. EAOP and CSMP are the most common to all Partner schools. In all cases, the motivation of the School University Partnership Program (S/UP) is to advance the rate at which students graduate from high school with a comprehensive educational background that makes them eligible to attend a University of California school. Completion of the A-G required course pattern is the single best indicator of the accomplishment of this objective.

Methodology

The University of California's commitment to these goals is matched by a strong commitment to measuring the effectiveness of its efforts to increase UC eligibility and improve student achievement. In this effort, the University of California commissioned a report by Policy Analysis for California Education (PACE) to inform deliberations of the Outreach Task Force. The report reviewed current evaluations of outreach programs and recommended improvements for the programs as well as guidelines for methods to evaluate the programs. PACE recommended that the University of California use a comparison group strategy to measure value-added and program effects (Hayward, Brandes, Kirst, & Mazzeo, 1997). The University of California set out to determine the feasibility of implementing a comparison group methodology that could identify whether and to what extent the UC's K-12 outreach programs were having an effect, despite the complications inherent both in measuring program effects and in the complex structure of the K-12 outreach programs. UCOP was interested in whether or not the S/UP program, with its mix of supportive academic development programs designed to address the needs of an individual school, affects student eligibility, course-taking, and achievement. Simply, they wanted to identify a comparison group methodology that could estimate the effect of the presence of UC's K-12 outreach efforts.

Overview of Literature on Matched Sampling Methodologies

The most accepted and reliable method of detecting treatment effects is a randomized study; however, in educational settings this strategy is virtually impossible due to ethical and practical reasons. Since it is usually not possible to randomly assign students to a treatment group and a control group, observational studies are generally conducted in educational settings. An observational study is "an empirical investigation of treatments, policies, or exposures and the effects they cause when assignment of treatments is not controlled" (Rosenbaum, 1995). Essentially, observational studies construct equivalent groups to estimate effects, and the more sophisticated, optimal approaches construct probabilistically equivalent groups.

The common method of constructing probabilistically equivalent groups is matched sampling (Heckman, 1989; Lalonde, 1986; Manski & Garfinkel, 1992; Rosenbaum, 1989, 1995; Rosenbaum & Rubin, 1983). There are many other methods

of estimating program effects, such as difference-in-differences or hierarchical linear modeling; however, these other techniques require either a census population or a large representative sample population to estimate an effect. Because matched sampling is so refined in its selection of controls, effects can be estimated with a relatively small sample population. This is of extreme benefit when the overall population is small or the costs to collect data across many sites are high.

Basically, matched sampling is a method of selecting units from a large reservoir of potential controls to produce a control group of modest size that is similar to a treated group with respect to the distribution of observed covariates. Matched sampling allows for the selection of groups and for adjustments to achieve probabilistic equivalency across the groups. Multivariate matching methods use several variables to construct probabilistically equivalent groups on observed characteristics. The matching can be done either by matching a whole group or pairwise. Propensity score matching refers to a class of multivariate matching methods where multiple variables are summarized into one score, and groups are matched on the one score.

Propensity score matching is considered the analog of randomization in observational studies. Propensity score matching is far less complete than randomization because it only balances the distribution of observed covariates (Rubin & Thomas, 2000), whereas randomization balances both unobserved and observed covariates. However, propensity score matching is superior to other multivariate matching methods because of its ability to reconcile variation across many observed factors and still establish probabilistically equivalent groups (Dehejia & Wahba, 1998; Rosenbaum, 1995; Rosenbaum & Rubin, 1983, 1985).

The limitations of a matched sampling methodology, whether the methodology incorporates a propensity score or not, are the potentially significant amounts of unobserved bias and the potential likelihood of not finding a sufficient match for each treated subject. Despite the limitations of using a matched sampling methodology, using a propensity score in constructing control groups is widely accepted and considered the optimal method of establishing a comparison group for an observational study. For these reasons, UCOP has adopted this methodology.

Multivariate Matched Sampling Methodology That Incorporates the Propensity Score

To estimate the effect of the presence of UC's K-12 outreach efforts, we needed to construct probabilistically equivalent groups of control schools for the 72 UC Partner high schools, 54 UC Partner middle schools, and 92 UC Partner elementary schools. Employing a multivariate matched sampling methodology that incorporated the propensity score, we needed to identify for every UC Partner school a unique control school based on a range of characteristics for the year prior to the implementation of the S/UP program.[1] Once the matched treatment and control groups are established, the program effect is measured by a simple comparison between the outcomes of the two groups in a given year or in growth over time.

Identifying these control schools required several steps. The first step in the matching process is to identify the potential set of match schools across the state. Potential match schools are regular public schools that (1) are not currently (2000-01), or have not in the last three years (1997-'98, 1998-'99, 1999-2000) been, UC Partner schools, and (2) have not participated during at least the last four years in AVID, PUENTE, MESA, EAOP, or the California Subject Matter Projects. These programs are the complementary outreach programs sponsored by UC.[2]

Next, from the group of potential match schools across the state, we identified any school that had characteristics similar to a Partner school on a few key variables known as "match variables." The match variables are school quintile on the Stanford Achievement Test Series, 9th edition (SAT-9), from spring of 1998; total school enrollment based on the October 1998 California Basic Educational Data System (CBEDS) school information form; percent eligible for free or reduced meals from the 1999 SAT-9 student header sheets; percent English language learners (ELL) from the spring 1999 R30-LC; and the ethnic breakdown of the percent Hispanic, percent African American, percent Asian, and percent White students from the October 1998

[1] 1997/'98 is the baseline year. 1998/'99 is the first year of implementation.

[2] To establish the group of regular schools, we first excluded a total of 207 schools in the state that were either alternative (*N* = 20), community (*N* = 1), juvenile (*N* = 1), or charter schools (*N* = 91), or were not able to be categorized as an elementary, middle or high school (*N* = 131), or were missing Academic Performance Index, demographic, and test score data (*N* = 5). Then we excluded schools that worked with the UC campuses. Across the state, we found the potential set of 275 match high schools, 559 match middle schools, and 3,075 match elementary schools.

CBEDS data collection. These key variables were selected as match variables since they define the type of demographic and socioeconomic population that a school serves. The type of population a school serves and the size of the school are well-known correlates for how well a school performs. We included in the set of key match variables the quintile of the school on the SAT-9 from Spring of 1998 because the S/UP program targets schools that on average score in the bottom three quintiles on the SAT-9. Across these five key match variables, we considered a potential match school with a mean that is within one standard deviation of the Partner school's mean to be "similar." In theory, this results in each and every UC Partner school having a set of similar schools from which at least one school could be selected as the final control school; this would be considered a full match. In practice, some Partner schools had no similar matches. Other Partner schools yielded a set of similar matches that ranged on average from 1 to 10 schools.[3] Refer to the Appendix A for a step-by-step example of this step in the matching methodology.

Next, we estimated a propensity score (i.e., the probability of being a UC Partner school) for every school, including UC Partner schools and all potential match schools. To estimate the probability of "being a UC Partner," we ran a logistic regression using the dichotomous variable of "being a UC Partner" (=1) and "not being a UC Partner" (=0) including all schools in the potential match pool and all Partners. The estimated probabilities range from 0 to 1. Via the logistic regression,

[3] The process yielded the following results. Of the 92 **UC Partner elementary schools**, 3 UC Partner elementary schools did not have data and 1 UC Partner elementary school did not yield at least one similar school from the potential set of match schools. These 4 UC Partner elementary schools are not included in the final Elementary Partner Group (N = 88). Of the 57 **UC Partner middle schools**, 3 did not have data and 11 did not yield at least one similar school from the potential set of match schools. Therefore, for the middle schools we had two sets of schools—the 11 middle schools without matches and the 43 middle schools with matches. We then examined these two groups across the match variables, as well as school mobility (defined as the percentage of students who first attended the school in the current year as indicated on the Stanford 9 header sheet from 1999), and the percentages of full and emergency credentialed teachers from the October 1998 CBEDS data collection. Comparing these middle school groups, we found that the two groups differed only modestly on mean characteristics. The 43 UC Partners with a match had higher percentages of African American and Asian students, as well as smaller numbers of students entering into the schools than the 11 UC Partners without matches. The final Middle School Partner Group has 43 schools. Of the 72 **UC Partner high schools**, 4 UC Partner high schools did not have data and 35 did not yield at least one similar school from the potential set of match schools. These 35 schools are unique in having very low or non-existent percentages of White students and very high percentages of either Hispanic or African-American students, which is the reason that UC is working with them on academic development. The 4 schools without data and the 35 schools for which we could not find a match were not included in the final High School Partner Group (N = 33). The differences between the UC Partner high schools for which we found matches (N = 33) and did not find matches (N = 35) are examined and discussed in depth in the section Discussion of Results of the Matched Sampling.

each school is assigned a predicted value of "being a UC Partner" based on the variables in the model. The logistic regression prediction equation included the following independent variables from fall 1997 demographic CBEDS data: school quintile on the SAT-9 for 1998; total enrollment from October 1997; total number of certified teachers from October 1997; suburban and rural dummies with urban as a referent group, percent eligible for free and reduced meals from 1998 SAT-9 header sheets; percent English Language Learners from the 1998 R30-LC; and percent Hispanic, percent African American, and percent Asian, with percent White as a referent group from the October 1997 CBEDS data. This yielded a propensity score based on information from the 1997/'98 school year, which we refer to as propensity98.

Finally from the set of similar schools identified for an individual Partner school, we identified which "similar" school out of the set had the nearest or exact-value propensity score to the Partner school. The school with the nearest or exact-value propensity score is then chosen to be the final match control school for that particular UC Partner school. If a potential match school had the nearest or exact-value propensity score for two UC Partner schools, the choice was made by minimizing the overall propensity score distance between each of the Partner schools and the potential Matches.

Once we chose the unique match school with the nearest value propensity score for each UC Partner school, we established the group of match schools for the UC Partner schools as well as the unique pairwise matching of UC Partner schools to their controls.

Discussion of Results of the Matched Sampling

Once the comparison group was constructed using the multivariate matching process that incorporates the propensity score (as outlined above), we compared the two groups—the UC Partners and the Matches—in terms of their means and distributions on the variables of interest. Table B1 (see Tables B1-B29 in Appendix B) reports the test of equality of the means and distributions for the elementary school Partners and Matches. Table B2 reports the results for the middle school Partners and Matches. Table B3 reports the results for the high school Partners and Matches.

Elementary school results. Comparing the elementary Partner schools (N = 88) to the elementary Match schools (N = 88) in Table B1, we see that the group means of elementary schools do not differ statistically in terms of the probability of

being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent eligible for free/reduced meals, percent of English Language learners, total enrollment, percent Filipino, the number of students who entered into the school in 1999/2000 (this variable is called *mobility*), average class size for Grades K-3, average class size for Grades 4-6, total number of full-credentialed teachers, and the total number of emergency-credentialed teachers. We assumed unequal variance and tested at the 0.05 significance level. In addition, we tested whether the distributions of these variables were equal using the Kolmogorov-Smirnov (K-S) test. The distributions for the above variables were also statistically similar. Note that if the means differ, then there is evidence that the distributions differ. Any discrepancy between the K-S test and the *t*-test of means is because the K-S test is less sensitive in identifying differences. The K-S test is less sensitive because it tests for global differences, whereas the *t*-test of means focuses on particular properties.

In addition to the tests of equality of means and distributions, we ran a logistic regression using the UC Partner elementary schools and the Match schools for a total *N* of 176 to predict "being a UC Partner." If the UC Partner schools and Match schools are similar, then the logistic regression should not be able to predict being a Partner (i.e., distinguish between the Partner and Match schools). We can judge how well the logistic regression was able to predict being a UC Partner by examining the discordant and concordant pairs.[4] We found 44.1% discordant pairs, 43.4% concordant pairs, and 12.4% tied. In a perfectly random assignment of schools, one would expect a roughly equal percentage of discordant and concordant pairs. Also, one usually finds somewhat more concordant than discordant pairs even in a random assignment, because the true coefficients are not known and the estimated coefficients tend to favor higher scores for the treatment group (called *capitalizing on chance*).

[4] If there are n1 Partner schools and n2 other schools, n1 + n2 = n. Each school has two values: (1) 1 or 0 indicating Partner or other, and (2) the probability of being a Partner based on the logistic regression. If all possible pairs were formed, there would be n(n-1)/2 pairs. Of the pairs, many would be Partner-to-Partner and other-to-other; thus, a pair of both 0s or both 1s. Otherwise, each of the n1 Partner schools is paired with each of the n2 other schools, so there are n1*n2 pairs. In this pairing of Partner to other schools, if the Partner school has a higher predicted probability of being a Partner than the other school, then the pair is concordant (i.e., the two variable values both go in the same direction). On the other hand, if the Partner school in the pair has a lower predicted probability of being a Partner school than the other school, the pair is discordant. If the two schools have the same predicted probability of being a Partner, the pair is tied. If schools are randomly divided into two groups (Partner and other) and the true best regression equation is used, we expect an equal number of concordant and discordant pairs.

Moreover, this procedure of using a logistic regression predicting "being a UC Partner" also allows us to calculate a global measure of discriminability for any two groups by using the logistic regression equation on the two groups and calculating the proportions of concordant and discordant pairs. The coefficient gamma = (proportion concordant – proportion discordant) / (proportion concordant + proportion discordant) is a frequently used correlation-like measure of the discriminability of the groups, varying from +1 (perfect prediction) through 0 (random prediction) through –1 (perfect negative prediction). The group of 176 elementary UC Partners and Matches yielded a coefficient gamma of -0.008 (essentially zero), indicating that the regression was not able to predict the difference between the UC Partner schools and the Match schools.

Thus, the Match elementary schools were not distinguishable from the Partner elementary schools according to the composite of the variables used in the logistic regression. All indications suggest that the UC Partner and Match elementary groups are similar across the observed variables. Therefore, we selected indistinguishable elementary school matches.

Middle school results. Comparing the middle school Partner schools (N = 43) to the Match schools (N = 43) in Table B2 in the same way, we also see that the group means of the middle schools do not differ statistically in terms of the probability of being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent eligible for free/reduced meals, percent of English Language Learners, total enrollment, percent Filipino, the number of students who entered into the school in 1999/2000 (mobility), and the total number of emergency-credentialed teachers. They do differ, however, in the percent of full-credentialed teachers. The Kolmogorov-Smirnov test indicated that the distributions of these variables were also equal and therefore statistically similar. The logistic regression using the UC Partner middle schools and the Match middle schools with a total N of 86 that predicted "being a UC Partner" resulted in 34.4% discordant pairs, 65.1% concordant pairs, and 0.5% tied. The group of 86 UC middle school Partners and Matches yielded a coefficient gamma of 0.308, indicating that the regression was able to predict the difference between the UC Partner middle schools and the Match middle schools. Thus, the Match middle schools are slightly distinguishable from the Partner middle schools according to the composite of the variables used in the logistic regression. However, the means and distributions of the two groups are statistically similar across the observed variables, indicating that the two groups

match very well and can be used as comparison groups. Therefore, we found slightly distinguishable middle school matches.

High school results. Finally, comparing the high school Partner schools (*N*=33) to the Match schools (*N*=33) in Table B3, we also see that the group means of the high schools do not differ statistically in terms of the probability of being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent eligible for free / reduced meals, percent of English Language Learners, total enrollment, percent Filipino, the number of students who entered into the school in 1999 / 2000 (mobility), average class size in the core curriculum high school courses, total number of full-credentialed teachers, and the total number of emergency-credentialed teachers. The Kolmogorov-Smirnov test indicated that the distributions of these variables were equal and therefore statistically similar. The logistic regression using the UC Partner high schools and the Match high schools with a total *N* of 66 that predicted "being a UC Partner" resulted in 34.8% discordant pairs, 63.4% concordant pairs, and 1.8% tied. The group of 66 UC high school Partners and Matches yielded a coefficient gamma of 0.291, indicating that the regression was able to predict the difference between the UC Partner high schools and the Match high schools. Thus, the Match high schools are slightly distinguishable from the Partner high schools according to the composite of the variables used in the logistic regression. However, the means and distributions of the two groups are statistically similar across the observed variables, indicating that the two groups match very well and can be used as comparison groups. Therefore, we found slightly distinguishable high school matches.

Summary

Based on these results and methodology, we were able to find a unique match for every UC Partner elementary school in the state, which yielded statistically similar groups. We were able to find a unique match for only a subset of the UC Partner middle schools in the state (43 out of the 57 UC Partner middle schools). This subset of UC Partner middle schools, for which we found statistically similar matches, was found however to be representative of all the UC Partner middle schools. This partial but representative match resulted in statistically similar groups, except for the percent of fully credentialed teachers. For the UC Partner high schools, we found unique matches for about half (33 out of 68) of the UC Partner schools. For the subset for which we found matches, the match resulted in

statistically similar groups. However, the subset of UC Partner schools that had matches were not representative of the entire group of UC Partner schools.

The partial match for the UC Partner high schools was not representative of all of the UC Partner high schools. Tables B4 and B5 report the means and *t*-tests for the UC Partner middle schools and high schools with a match and without a match, respectively. Table B4 indicates that the 11 UC Partner middle schools for which we did not find matches were statistically similar to the 43 UC Partner middle schools for which we did find matches in terms of socio-economic status, English Language Learners, school size, full- and emergency-credentialed teachers, and the propensity score for being a Partner in 1998. There were a few minor differences—the UC Partner middle schools without a match were statistically different than those with a match in that they had more African American and Asian students as well as more incoming mobility. Despite these slight differences, the UC Partner middle schools with a match are representative of all of the UC Partner middle schools. Table B5 indicates that the UC Partner high schools for which we could not find a match are statistically different from the UC Partner high schools across almost all of the variables of interest for which we could find a match. This means that the subset for which we found matches are unique in the populations that they serve, their size, and performance. Table B5 indicates that the schools for which we could not find a match had a higher percentage of African American, Asian, and Hispanic students, and a lower percentage of White students than the UC Partner high schools for which we found matches. The schools without a match also had a higher percentage of students eligible for reduced/free lunch, a higher percent of English language learners, a larger total enrollment, fewer full-credentialed teachers, and more emergency-credentialed teachers than the UC Partner schools for which we found matches. In the end, the 35 UC Partner high schools without matches differ systematically from the 33 UC Partner high schools for which we did find matches, indicating that the partial match for the UC Partner high schools is not representative. Presumably, we could not find matches for UC Partner high schools with highly disadvantaged populations because UC is working with all of these schools either through AVID, EAOP, MESA, Puente, UC Partnerships, or the California Subject Matter Projects.

Overall, we were able to find a group of elementary schools in the state that are statistically similar to the UC Partner elementary schools and a group of middle schools that are statistically similar to the UC Partner middle schools. This enables

simple comparisons of the Partners to the Matches for both the elementary schools and the middle schools in terms of the means and distributions for outcome variables and achievement measures of interest and allows for comparisons of pair differences. Because of the slight difference found for the middle schools in the percent of fully credentialed teachers, we should compare these UC Partner schools and their matches adjusting for the covariance of this variable to remove the linear effect of the remaining difference. For the UC Partner high schools, comparisons can be made only across the 33 UC Partner high schools and their 33 match high schools, remembering, however, that these comparisons do not reflect the changes in the entire group of UC Partner high schools.

Illustrative Analyses of Effects

Based on our results showing that the UC Partner schools and their matched schools were similar across a desired set of observed characteristics in the baseline year, we can assess the change in these UC Partner schools' performance (i.e., the effect of the presence of the UC K-12 outreach efforts) by comparing either static outcomes in subsequent years (i.e., 1998/'99 and 1999/2000) or growth over time (i.e., from 1998 to 2000).

Since the purpose of this report is not to evaluate effects, but to show the feasibility of using a matched sampling methodology and to establish a protocol for evaluating changes in outcomes, we will outline a few illustrative analyses. We show in this section the simplicity of using comparison groups to determine statistically significant differences when the treatment and control matched groups are statistically similar. We discuss differences in achievement scores, since achievement gains are commonly understood outcome measures. However, note that gains in achievement are not a direct aim or outcome of the UC K-12 outreach programs. The best outcome measure of the UC's K-12 outreach efforts is the completion of the A-G required course patterns.

To assess overall achievement, we examined the school-level distribution of reading and mathematics STAR test results in 2000 as well as the growth in these scores from 1998 to 2000. It is important to note that SAT-9 test results at the school level are only general indicators of achievement at the school and are not intended to be sensitive to changes in student content knowledge or specific UC program elements. Therefore, this type of evaluation analysis is not sensitive to student-level changes, but is trying to identify larger, school-level changes or trends.

The 2000 STAR test results are provided in Tables B6 through 13. Table B6 and Table B7 report the Partner and Match elementary school STAR test results for the percent of students scoring above the 75th percentile, at or above the 50th percentile, and above the 25th percentile for reading and mathematics in 2000, respectively, as well as the *t*-test of their equality as groups. Table B8 and Table B9 report the reading and mathematics results for the Partner and Match middle schools, and Table B10 and Table B11 report the reading and mathematics results for the Partner and Match high schools. Additionally, Table B12 and Table B13 report the Partner and Match high school STAR test results of the percent of students scoring above the 75th percentile, at or above the 50th percentile, and above the 25th percentile for science and social science in 2000, respectively, as well as the *t*-test of their equality.

The 1998-2000 STAR test growth results are provided in Tables B14 through B21. Table B14 and Table B15 report the Partner and Match elementary school STAR test growth from 1998 to 2000 for the percent of students scoring above the 75th percentile, at or above the 50th percentile, and above the 25th percentile for reading and mathematics, respectively. Table B16 and Table B17 report the reading and mathematics results for the Partner and Match middle schools, and Table B18 and Table B19 report the reading and mathematics results for the Partner and Match high schools. Additionally, Table B20 and Table B21 report the Partner and Match high school STAR test growth from 1998 to 2000 for the percent of students scoring above the 75th percentile, at or above the 50th percentile, and above the 25th percentile for science and social science, respectively.

More importantly, Tables B14 through B21 include the independent sample *t*-test of their equality as groups. Because of the pairwise matched design, we also conducted paired-sample *t*-tests. Tables B22 through B29 report the differences in 1998-2000 Partner and Match school test score growth for the percent of students scoring above the 75th percentile, at or above the 50th percentile, and above the 25th percentile for reading and mathematics for the elementary, middle and high schools, as well as the paired sample *t*-tests. We found that the correlation between the pairs was generally above 0.50 for all three groups, elementary, middle, and high. This indicates that the paired sample *t*-tests are more powerful than the independent sample *t*-tests. Specifically, the correlations between the pairs for the elementary schools were generally above .50 (48 out of 88) with an average correlation of 0.91; for middle schools, the correlations were generally above .50 (18 out of 43) with an

average correlation of 0.93; and for the high schools, the correlations were generally below .50 (7 out of 33) with an average correlation of 0.31.

Discussion of Differences Between UC Partner Schools and Match Schools

Elementary school discussion. For the elementary schools, we found that the distributions of reading and mathematics 2000 STAR test results did not differ for the UC Partner and Match elementary schools for second, third, fourth and fifth grades (see Table B6 and Table B7). For sixth grade, it appears that the Match schools had a significantly higher percentage of students scoring about the 75th percentile, at or above the 50th percentile in reading, and at or above the 50th percentile in mathematics. These same patterns were found for spelling and listening across the grades. In addition, we looked at the growth of STAR test scores from 1998 to 2000 for the UC Partner elementary schools and their matches (see Table B14 and Table B15). In terms of growth or change in the distribution of students scoring at the 75th, 50th, and 25th percentiles in reading and math, we found that the distributions of the test scores across all the subject areas and all grades were statistically similar. There appear to be some slight differences in the mean scale score growth in reading, mathematics, and spelling across a few grades, but nothing noteworthy. Given the large number of tests, a small percentage would be expected to achieve significance at the 0.05 level by chance, and therefore, the test results should be considered significant only when they occur in a systematic pattern by grade level, or subject matter, etc.

In addition, based on the design of pair-wise matched sampling, each UC Partner school had one unique match school. This allows, as stated above, for a test of the differences in the two-paired outcomes by conducting paired *t*-tests. No differences were found in the reading, math, listening, and spelling mean scale scores or growth in these scores for the UC Partner schools and their matches, Grades 2 through 6, based on paired *t*-tests. Thus, the independent and paired sample *t*-tests yielded the same results.

Based on tests of group means and distributions as well as paired *t*-tests, we found that, in general, the reading, mathematics, listening, and spelling achievement patterns for Grades 2 through 6 for the UC Partner elementary schools and their matches were similar in 2000 and had similar growth from 1998 to 2000.

Middle school discussion. For the middle schools, we found that the distributions of reading and mathematics 2000 STAR test results did not differ for

the UC Partner and Match middle schools for sixth, seventh, and eighth grades (see Table B8 and Table B9). There were no differences found between the Partner and Match schools across grades in reading and math. For eighth grade, it appears that the Match schools had a significantly lower percentage of students scoring about the 75th percentile in language arts. In addition, we looked at the growth of STAR test scores from 1998 to 2000 for the UC Partner middle schools and their matches (see Table B16 and Table B17). In terms of growth or change in the distribution of students scoring at the 75th, 50th, and 25th percentiles in reading and math, we found that the distributions of the test scores across all the subject areas and all grades were statistically similar. Additionally, no differences were found in the reading, math, listening, and spelling mean scale scores or growth in these scores for the UC Partner schools and their matches, Grades 6 through 9, based on paired *t*-tests. In general, the reading, mathematics, listening, and spelling achievement patterns for Grades 6 through 9 for the UC Partner middle schools and their matches were similar in 2000 and had similar growth from 1998 to 2000. The paired sample *t*-tests yielded the same results.

High school discussion. For the high schools, we found that the distributions of reading, mathematics, science, and social science 2000 STAR test results did not statistically differ for the 33 UC Partner and 33 Match high schools for 9th, 10th, and 11th grades (see Tables B10 through B13). Notice that the Match schools have a significantly higher percentage of students scoring above the 25th percentile in reading in 11th grade, in science in the 10th grade, and in social science in the 11th grade. This indicates some slight differences in the bottom tail of the distribution in 2000, but is not an overall pattern across grades or subject areas. Also remember that the 33 UC Partner high schools are not representative of all of the UC Partner high schools. We could not find matches for UC Partner high schools with highly disadvantaged populations (see above for details). In addition, we looked at the growth of STAR test scores from 1998 to 2000 for the UC Partner elementary schools and their matches (see Tables B18 through B21). In terms of growth or change in the distribution of students scoring at the 75th, 50th, and 25th percentiles in reading, mathematics, science, and social science, we found that the distributions of the test scores across all the subject areas and all grades, 9 through 11, were statistically similar. No slight differences were found. Additionally, no differences were found in the reading, math, science, and social science mean scale scores or growth in these scores for the UC Partner schools and their matches, Grades 9 through 11, based on

paired *t*-tests. In general, the reading, mathematics, science and social science achievement patterns for grades 9 through 11 for the 33 UC Partner high schools and their 33 matches were very similar in 2000 and had similar growth from 1998 to 2000. Again, the paired sample *t*-tests yielded the same results.

Summary

Overall, we found that the 88 UC Partner elementary schools had test score means in 2000, test score distributions in 2000, and test score growth from 1998 to 2000 in reading, mathematics, spelling, and listening on the STAR achievement tests that were statistically similar to the test score means, distributions, and growth for a similar set of elementary schools in the state that did not have UC programs. We also found these similarities in achievement measures (static and growth) for the 43 UC Partner middle schools and their similar set of middle schools in the state that did not have UC programs. Moreover, we found that the UC Partner high schools that served the most disadvantaged populations (N = 35) did not have a comparison group. However, of the UC Partner high schools for which we did have a comparison group, the UC Partner high schools had test score means in 2000, test score distributions in 2000, and test score growth from 1998 to 2000 in reading, mathematics, science, and social science on the STAR achievement tests that were statistically similar to the test score means, distributions, and growth for a similar set of high schools in the state that did not have UC programs.

Based on these illustrative analyses, we found that since 1997/'98 we have not seen a shift or statistically large school-level change in achievement on the SAT-9 by the UC Partner schools that is any different than the changes and shifts by schools without UC programs across the state with similar school demographics, mobility, enrollment, and percent of full- and emergency-credentialed teachers. However, large or even small changes in these school-level academic outcomes were not expected considering that the UC programs only broadened their focus in late 1997/'98, changes in school-level averages and means take a long time to detect, and the SAT-9 academic outcomes are not very sensitive to a student's A-G course-level completion patterns. The UC's K-12 outreach programs require time to have a measurable school-level effect on achievement and more probably will have an earlier and more significant impact on A-G completion rates.

Conclusions

This study's value lies primarily in having identified an optimal approach to comparison for UC's evaluation purposes and in having identified, using a statewide sample of schools, a specific set of comparison schools at the elementary, middle, and high school levels that can be used for subsequent studies. Future evaluation work will build upon this methodology and sample to investigate the impact of the School University Partnership Program on A-G completion and UC eligibility.

References

Dehejia, R. H., & Wahba S. (1998). Propensity score matching methods for non-experimental causal studies (National Bureau of Economic Research Working Paper No. 6829). Cambridge, MA: National Bureau of Economic Research.

Hayward, G. C., Brandes, B. G., Kirst, M. W., & Mazzeo, D. (1997, January). *Higher education outreach programs: A synthesis of evaluations.* A Report commissioned by the Outreach Task Force, Board of Regents, University of California, written by Policy Analysis for California Education (PACE).

Heckman, J. J. (1989). Casual inference and nonrandom samples. *Journal of Educational Statistics, 14,* 159-168.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs. *American Economic Review, 76,* 604-620.

Manski, D. F., & Garfinkel, I. (1992). Introduction. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs* (pp. 1-22). Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (1995). *Observational studies.* New York: Springer-Verlag.

Rosenbaum, P. R. (1989). Optimal matching in observational studies. *Journal of the American Statistical Association, 84,* 1024-1032.

Rosenbaum, P. R., & Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70, 41-55.

Rosenbaum, P. R., & Rubin D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39,* 33-38.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of American Statistical Association, 95,* 573-585.

APPENDIX A: EXAMPLE OF MULTIVARIATE MATCHED SAMPLING THAT INCORPORATES THE PROPENSITY SCORE

This appendix guides the reader through an example of how the matching process works given a Partner school with the characteristics detailed in Table A1. In Table A1, the second column is the Partner's values for the match variables. The third column is the population standard deviation for a given variable, and in the fourth column is the range that potential match schools must fall within to be considered a match. The lower bound of the range for a given variable is calculated by subtracting the Partner population standard deviation from the individual Partner's value for that variable. The upper bound of the range is calculated by adding the Partner school's population standard deviation to the individual Partner's value for that variable. For example, this Partner school's enrollment is 1,456 students. By subtracting and adding 961.824 from and to 1,456, the range of 494.18 to 2,417.82 is obtained. This selection can be done for all variables simultaneously using a built-in feature in Microsoft Excel.[5]

Table A1

An Example of the Matching Process

Variable	Partner School	Population S.D.	Range[a]
Propensity98	0.887	0.170	Not Applicable
Enrollment	1,456	961.824	494.18 to 2,417.82
Pct_ell	0.199	0.150	0.049 to 0.349
Pct_F_R_meals	0.315	0.198	0.120 to 0.513
Pct_hispanic	0.090	0.263	0.000 to 0.353
Pct _africanam	0.431	0.194	0.236 to 0.625
Pct_asian	0.223	0.150	0.072 to 0.373
Pct_white	0.224	0.156	0.068 to 0.380
SAT Quint	3	0.980	Not Applicable

[a]If the lower bound of the range is a negative number, it is set to zero because it is impossible for these variables to have negative values.

[5] The advanced filter in Microsoft Excel was used in the matching process.

Table A2 displays the results of the first step of the matching process. One will notice that these two possible matches have values within the selected ranges for all of the match variables.

Table A2

An Example of the Matching Process—Results of the First Step

Variable	Range[a]	Potential Match 1	Potential Match 2
Propensity98	Not Applicable	0.761	0.823
Enrollment	494.18 to 2,417.82	1,323	1,141
Pct_ell	0.049 to 0.349	.145	.162
Pct_F_R_meals	0.120 to 0.513	.308	.371
Pct_hispanic	0.000 to 0.353	0.314	0.112
Pct _africanam	0.236 to 0.625	0.256	0.288
Pct_asian	0.072 to 0.373	0.093	0.171
Pct_white	0.068 to 0.380	0.261	0.239
SAT Quint	Not Applicable	5	4

[a]If the lower bound of the range is a negative number, it is set to zero since it is impossible for any of these variables to be negative.

The next step is to identify the schools that match on SAT quintile. From Table A1, the SAT quintile of the Partner school is 3. Therefore, for the two possible matches to be considered potential match schools they must have an SAT quintile of 3, 4, or 5. They both have an SAT quintile of 3, 4, or 5; therefore, these schools are considered potential match schools. Since the Partner school in this example has at least one potential match school, this part of the matching process is complete for this school.

Finally, the propensity score for the Partner school is compared to the propensity score of each of its potential match schools. The potential match school with the nearest (or exact) propensity score value is selected as the match for the Partner school. In our example, we compare the Partner's propensity score of 0.887 to the propensity scores values of 0.761 and 0.823. The propensity score of 0.823 for potential Match 2 is nearer to 0.887; therefore, Potential Match 2 is chosen as the final match for the Partner school. Once a match is selected for a Partner school, this school is excluded from being a match to any other of the Partner schools.

APPENDIX B: TABLES

Table B1

Test of the Equality of Means and Distributions for Elementary School Partners (N = 88) and Elementary School Matches (N = 88)

	Equality of distribution (Kolmogorov-Smirnov Test)		Equality of means (2-sided)	
Variable	Combined K-S	*p*-Value	*t*-Statistic	*p*-Value
Propensity98	0.0490	1.0000	0.1493	0.8815
Pct_aa	0.0795	0.9220	-0.3207	0.7488
Pct_as	0.0909	0.8200	-0.3070	0.7592
Pct_hi	0.0682	0.9800	0.2181	0.8276
Pct_wh	0.1136	0.5560	0.6361	0.5255
Pct_meal	0.1364	0.3240	-0.7343	0.4638
Pct_ell	0.0909	0.8200	0.0762	0.9394
Enroll	0.1247	0.4440	-0.9366	0.3503
Pct_fi	0.0909	0.8200	-1.3453	0.1803
Mobility	0.0682	0.9800	0.4097	0.6825
ACS_K3	0.1451	0.2580	1.9241	0.0560
ACS_46	0.0895	0.8410	0.7226	0.4709
Full Crd	0.0682	0.9800	-0.4418	0.6592
Emergency Crd	0.1477	0.2370	1.5957	0.1124

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B2

Test of the Equality of Means and Distributions for Middle School Partners (N = 43) and Middle School Matches (N = 43)

	Equality of distribution (Kolmogorov-Smirnov test)		Equality of means (2-sided)	
Variable	Combined K-S	*p*-Value	*t*-Statistic	*p*-Value
Propensity98	0.1229	0.9050	-0.6520	0.5162
Pct_aa	0.0930	0.9920	-0.2592	0.7961
Pct_as	0.1395	0.7970	-0.6814	0.4975
Pct_hi	0.0698	1.0000	0.2015	0.8408
Pct_wh	0.1860	0.4460	0.4741	0.6366
Pct_meal	0.1163	0.9930	0.6767	0.5005
Pct_ell	0.0930	0.9920	0.0851	0.9324
Enroll	0.1528	0.7040	0.4886	0.6264
Pct_fi	0.1628	0.6190	0.1982	0.8434
Mobility	0.1628	0.6190	1.5092	0.1350
Full Crd	0.2791	0.7070	**-2.0482**	**0.0437**
Emergency Crd	0.2093	0.3030	1.6294	0.1070

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B3

Test of the Equality of Means and Distributions for High School Partners (N = 33) and High School Matches (N = 33)

	Equality of distribution (Kolmogorov-Smirnov test)		Equality of means (2-sided)	
Variable	Combined K-S	*p*-Value	*t*-Statistic	*p*-Value
Propensity98	0.2764	0.2270	-0.1621	0.1107
Pct_aa	0.0909	0.9990	-0.3815	0.7041
Pct_as	0.1818	0.6460	-0.1854	0.8535
Pct_hi	0.1515	0.8430	-0.5163	0.6075
Pct_wh	0.1515	0.8430	1.0438	0.3005
Pct_meal	0.1818	0.6460	-0.1184	0.9061
Pct_ell	0.3030	0.0970	-1.8152	0.0742
Enroll	0.1799	0.6690	-0.0789	0.9374
Pct_fi	0.2424	0.2870	1.2533	0.2147
Mobility	0.1515	0.8430	-0.3981	0.6919
ACS_core	0.0833	1.0000	0.2743	0.7847
Full Crd	0.2727	0.1720	-1.9030	0.0606
Emergency Crd	0.2424	0.2870	1.4776	0.1444

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B4

Means and Test of the Equality of Means for Middle School Partners With a Match and Partners Without a Match

	Means		Equality of means (2-sided)	
Variable	Partners with Match (N = 43)	Partners without Match (N = 11)	t-Statistic	p-Value
Propensity98	0.20	0.28	1.78	0.081
Pct_aa	8.62	29.30	**4.72**	**0.001**
Pct_as	6.70	15.60	**2.55**	**0.014**
Pct_hi	61.26	40.90	-2.32	0.240
Pct_wh	17.58	9.10	-1.85	0.070
Pct_meal	59.44	73.00	1.99	0.052
Pct_ell	32.26	39.3	1.01	0.316
Enroll	929.05	1037.5	0.90	0.370
Pct_fi	2.93	1.70	-0.68	0.501
Mobility	15.37	24.40	**2.32**	**0.024**
Full Crd	83.72	85.30	0.47	0.637
Emergency Crd	15.42	14.80	-0.19	0.847

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B5

Means and Test of the Equality of Means for High School Partners With a Match and Partners Without a Match

	Means		Equality of Means (2-Sided)	
Variable	Partners with Match (N = 33)	Partners without Match (N = 35)	t-Statistic	p-Value
Propensity98	0.172	0.462	**6.3086**	0.0000
Pct_aa	9.515	21.371	**2.8033**	0.0066
Pct_as	9.666	6.942	-1.0789	0.2846
Pct_hi	48.878	61.428	**2.0837**	0.0411
Pct_wh	25.484	5.685	**-6.7638**	0.0000
Pct_meal	37.909	60.085	**4.2517**	0.0001
Pct_ell	20.727	33.714	**3.8455**	0.0003
Enroll	1473.34	2033.8	**3.2514**	0.0018
Pct_fi	2.575	2.000	-0.5977	0.5521
Mobility	15.212	16.657	0.4030	0.6882
ACS_core	27.272	27.735	0.7169	0.4760
Full Crd	85.848	78.771	**-3.2455**	0.0018
Emergency Crd	14.272	18.085	**2.1303**	0.0369

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B6

2000 Reading STAR Test Comparisons for Partner Elementary Schools vs. Matches

READING	Partner $N = 88$	Match $N = 88$	*t*-Statistic	H_0: Equal Variance	*p*-value
SECOND GRADE					
Percent Scoring:					
Above 75th percentile	14.85	13.25	-0.99	Yes	0.32
At or above 50th percentile	37.34	35.87	-0.53	Yes	0.59
Above 25th percentile	60.94	58.32	-0.85	Yes	0.39
THIRD GRADE					
Percent Scoring:					
Above 75th percentile	10.48	10.39	-0.06	No	0.93
At or above 50th percentile	30.93	29.18	-0.76	Yes	0.44
Above 25th percentile	57.60	55.36	-0.90	Yes	0.36
FOURTH GRADE					
Percent Scoring:					
Above 75th percentile	11.53	11.96	0.28	Yes	0.77
At or above 50th percentile	29.09	29.47	0.16	Yes	0.87
Above 25th percentile	55.73	53.62	-0.78	Yes	0.43
FIFTH GRADE					
Percent Scoring:					
Above 75th percentile	10.75	10.11	-0.44	No	0.65
At or above 50th percentile	28.62	25.46	-1.32	Yes	0.18
Above 25th percentile	52.73	48.55	-1.43	Yes	0.15
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	2.84	6.09	**2.72**	No	**0.007**
At or above 50th percentile	8.82	14.81	**2.28**	No	**0.024**
Above 25th percentile	19.10	28.40	1.95	No	0.052

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B7

2000 Mathematics STAR Test Comparisons for Partner Elementary Schools vs. Matches

MATHEMATICS	Partner $N = 88$	Match $N = 88$	*t*-Statistic	H_0: Equal Variance	*p*-value
SECOND GRADE					
Percent Scoring:					
Above 75th percentile	22.28	20.47	-0.93	Yes	0.35
At or above 50th percentile	45.07	42.50	-0.97	Yes	0.32
Above 25th percentile	69.29	65.80	-1.23	No	0.21
THIRD GRADE					
Percent Scoring:					
Above 75th percentile	20.87	19.59	-0.72	Yes	0.46
At or above 50th percentile	46.27	44.12	-0.89	Yes	0.37
Above 25th percentile	71.02	68.34	-1.16	No	0.24
FOURTH GRADE					
Percent Scoring:					
Above 75th percentile	16.96	16.70	-0.16	Yes	0.87
At or above 50th percentile	38.21	37.03	-0.49	Yes	0.61
Above 25th percentile	62.36	62.09	-0.11	Yes	0.91
FIFTH GRADE					
Percent Scoring:					
Above 75th percentile	14.84	13.48	-0.79	No	0.42
At or above 50th percentile	35.63	33.03	-1.04	Yes	0.29
Above 25th percentile	59.17	53.90	-1.73	No	0.08
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	5.93	10.16	1.94	No	0.05
At or above 50th percentile	12.44	21.58	**2.17**	No	**0.03**
Above 25th percentile	20.24	30.15	1.61	Yes	0.10

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B8

2000 Reading STAR Test Comparisons for Partner Middle Schools vs. Matches

READING	Partner $N = 43$	Match $N = 43$	*t*-Statistic	H_0: Equal Variance	*p*-value
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	9.09	6.22	-1.46	Yes	0.15
At or above 50th percentile	22.12	17.72	-1.11	Yes	0.27
Above 25th percentile	41.77	35.59	-0.90	Yes	0.37
SEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	12.26	11.00	-0.82	Yes	0.42
At or above 50th percentile	32.44	31.25	-0.47	Yes	0.64
Above 25th percentile	58.12	57.22	-0.34	Yes	0.74
EIGHTH GRADE					
Percent Scoring:					
Above 75th percentile	12.65	9.81	-1.60	Yes	0.11
At or above 50th percentile	36.53	32.69	-1.40	Yes	0.17
Above 25th percentile	64.49	62.06	-0.94	Yes	0.32

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B9

2000 Mathematics STAR Test Comparisons for Partner Middle Schools vs. Matches

MATHEMATICS	Partner $N = 43$	Match $N = 43$	*t*-Statistic	H_0: Equal Variance	*p*-value
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	14.72	10.31	-1.66	Yes	0.10
At or above 50th percentile	30.21	23.81	-1.28	Yes	0.21
Above 25th percentile	46.67	39.44	-0.98	Yes	0.33
SEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	14.37	13.12	-0.68	Yes	0.50
At or above 50th percentile	35.95	32.91	-1.19	Yes	0.24
Above 25th percentile	62.47	59.66	-1.20	Yes	0.24
EIGHTH GRADE					
Percent Scoring:					
Above 75th percentile	14.14	11.50	-1.50	Yes	0.14
At or above 50th percentile					
Above 25th percentile					

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B10

2000 Reading STAR Test Comparisons for Partner High Schools vs. Matches

READING	Partner $N = 33$	Match $N = 33$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	7.90	8.57	-0.47	Yes	0.63
At or above 50th percentile	25.51	28.09	-0.94	Yes	0.34
Above 25th percentile	52.93	56.27	-1.07	Yes	0.28
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	9.48	9.93	-0.31	Yes	0.75
At or above 50th percentile	24.87	28.51	-1.49	Yes	0.13
Above 25th percentile	47.27	52.45	-1.91	Yes	0.059
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	11.09	12.24	-0.71	Yes	0.47
At or above 50th percentile	27.36	31.60	-1.76	Yes	0.08
Above 25th percentile	54.63	60.69	**-2.36**	Yes	**0.02**

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B11

2000 Mathematics STAR Test Comparisons for Partner High Schools vs. Matches

MATHEMATICS	Partner $N = 33$	Match $N = 33$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	17.24	18.81	-0.69	Yes	0.48
At or above 50th percentile	43.81	44.54	-0.22	Yes	0.82
Above 25th percentile	71.90	71.60	0.09	Yes	0.92
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	12.9	16.03	-1.43	Yes	0.15
At or above 50th percentile	37.27	44.15	-1.65	Yes	0.10
Above 25th percentile	67.27	70.39	-1.50	Yes	0.13
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	18.30	21.09	-1.24	Yes	0.21
At or above 50th percentile	40.03	44.60	-1.51	Yes	0.13
Above 25th percentile	65.18	69.36	-1.68	Yes	0.09

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B12

2000 Science STAR Test Comparisons for Partner High Schools vs. Matches

SCIENCE	Partner $N = 33$	Match $N = 33$	t-Statistic	H_0: Equal Variance	p-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	8.90	9.90	-0.76	Yes	0.44
At or above 50th percentile	32.27	34.24	-0.74	Yes	0.45
Above 25th percentile	70.63	72.24	-0.56	Yes	0.57
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	12.57	14.90	-1.25	Yes	0.21
At or above 50th percentile	36.51	40.90	-1.59	Yes	0.11
Above 25th percentile	63.03	67.51	**-2.01**	Yes	**0.04**
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	15.30	17.30	-1.03	Yes	0.30
At or above 50th percentile	33.57	38.24	-1.77	Yes	0.08
Above 25th percentile	64.33	68.60	-1.80	Yes	0.07

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B13

2000 Social Science STAR Test Comparisons for Partner High Schools vs. Matches

SOCIAL SCIENCE	Partner $N = 33$	Match $N = 33$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	10.27	10.48	-0.13	Yes	0.89
At or above 50th percentile	38.06	40.03	-0.72	Yes	0.47
Above 25th percentile	70.66	71.63	-0.35	Yes	0.72
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	12.60	13.96	-0.82	Yes	0.41
At or above 50th percentile	29.57	32.54	-1.24	Yes	0.21
Above 25th percentile	54.75	58.39	-1.70	Yes	0.09
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	27.09	30.21	-1.36	Yes	0.17
At or above 50th percentile	52.78	57.06	-1.82	Yes	0.07
Above 25th percentile	73.54	77.24	**-2.19**	Yes	**0.03**

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B14

1998-2000 Reading STAR Test Growth Comparisons for Partner Elementary Schools vs. Matches

READING	Partner $N = 88$	Match $N = 88$	*t*-Statistic	H_0: Equal Variance	*p*-value
SECOND GRADE					
Percent Scoring:					
Above 75th percentile	4.45	4.25	-0.16	Yes	0.87
At or above 50th percentile	8.98	8.85	-0.06	Yes	0.95
Above 25th percentile	11.42	10.68	-0.30	Yes	0.76
THIRD GRADE					
Percent Scoring:					
Above 75th percentile	2.23	2.80	0.58	Yes	0.56
At or above 50th percentile	6.69	6.55	-0.10	Yes	0.92
Above 25th percentile	11.42	10.68	-0.30	Yes	0.76
FOURTH GRADE					
Percent Scoring:					
Above 75th percentile	1.92	1.43	-0.35	Yes	0.72
At or above 50th percentile	5.31	6.16	0.45	Yes	0.65
Above 25th percentile	11.42	10.68	-0.30	Yes	0.76
FIFTH GRADE					
Percent Scoring:					
Above 75th percentile	1.32	1.62	0.39	Yes	0.70
At or above 50th percentile	4.49	1.98	-2.04	Yes	0.43
Above 25th percentile	11.42	10.68	-0.30	Yes	0.76

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B15

1998-2000 Mathematics STAR Test Growth Comparisons for Partner Elementary Schools vs. Matches

MATHEMATICS	Partner $N = 88$	Match $N = 88$	*t*-Statistic	H_0: Equal Variance	*p*-value
SECOND GRADE					
Percent Scoring:					
Above 75th percentile	8.80	8.07	-0.40	Yes	0.69
At or above 50th percentile	13.44	11.25	-0.93	Yes	0.35
Above 25th percentile	12.89	11.13	-0.67	Yes	0.50
THIRD GRADE					
Percent Scoring:					
Above 75th percentile	10.30	9.45	-0.59	Yes	0.56
At or above 50th percentile	17.77	16.72	-0.50	Yes	0.62
Above 25th percentile	12.89	11.13	-0.67	Yes	0.50
FOURTH GRADE					
Percent Scoring:					
Above 75th percentile	7.38	6.54	-0.53	Yes	0.60
At or above 50th percentile	13.98	11.37	-1.20	Yes	0.23
Above 25th percentile	12.89	11.13	-0.67	Yes	0.50
FIFTH GRADE					
Percent Scoring:					
Above 75th percentile	5.13	5.13	-0.01	Yes	0.99
At or above 50th percentile	8.81	9.28	0.25	Yes	0.80
Above 25th percentile	12.89	11.13	-0.67	Yes	0.50

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B16

1998-2000 Reading STAR Test Growth Comparisons for Partner Middle Schools vs. Matches

READING	Partner $N = 43$	Match $N = 43$	*t*-Statistic	H_0: Equal Variance	*p*-value
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	3.12	1.19	-1.13	Yes	0.26
At or above 50th percentile	6.23	3.31	-0.95	Yes	0.34
Above 25th percentile	10.58	1.78	-1.76	Yes	0.08
SEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	2.77	1.81	-0.58	Yes	0.56
At or above 50th percentile	6.93	5.13	-0.59	Yes	0.56
Above 25th percentile	10.47	5.59	-0.93	Yes	0.36
EIGHTH GRADE					
Percent Scoring:					
Above 75th percentile	3.88	1.22	-1.43	Yes	0.16
At or above 50th percentile	8.16	3.53	-1.33	Yes	0.18
Above 25th percentile	11.04	3.09	-1.35	Yes	0.18

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B17

1998-2000 Mathematics STAR Test Growth Comparisons for Partner Middle Schools vs. Matches

MATHEMATICS	Partner $N = 43$	Match $N = 43$	*t*-Statistic	H_0: Equal Variance	*p*-value
SIXTH GRADE					
Percent Scoring:					
Above 75th percentile	6.47	4.22	-1.13	Yes	0.26
At or above 50th percentile	10.63	6.13	-1.27	Yes	0.21
Above 25th percentile	12.53	2.19	-1.84	Yes	0.07
SEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	5.19	3.56	-0.88	Yes	0.38
At or above 50th percentile	10.21	6.66	-1.14	Yes	0.26
Above 25th percentile	12.39	6.25	-1.12	Yes	0.27
EIGHTH GRADE					
Percent Scoring:					
Above 75th percentile	5.33	2.34	-1.73	Yes	0.09
At or above 50th percentile	9.56	6.28	-0.89	Yes	0.37
Above 25th percentile	12.72	4.00	-1.39	Yes	0.17

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B18

1998-2000 Reading STAR Test Growth Comparisons for Partner High Schools vs. Matches

READING	Partner $N = 33$	Match $N = 35$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	0.22	-0.03	-0.36	Yes	0.72
At or above 50th percentile	0.69	1.58	-0.78	Yes	0.44
Above 25th percentile	2.75	1.24	-0.55	Yes	0.59
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	1.25	-0.14	-2.28	Yes	0.03
At or above 50th percentile	1.31	0.83	-0.40	Yes	0.69
Above 25th percentile	1.06	1.00	-0.04	Yes	0.97
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	-0.63	-0.52	0.15	Yes	0.88
At or above 50th percentile	-1.09	0.93	1.57	Yes	0.12
Above 25th percentile	0.16	1.93	1.08	Yes	0.29

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B19

1998-2000 Mathematics STAR Test Growth Comparisons for Partner High Schools vs. Matches

MATHEMATICS	Partner $N = 33$	Match $N = 35$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	3.22	1.79	-1.01	Yes	0.31
At or above 50th percentile	6.00	1.83	-1.64	Yes	0.11
Above 25th percentile	6.41	2.31	-1.25	Yes	0.22
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	2.59	1.86	-0.75	Yes	0.46
At or above 50th percentile	5.13	3.62	-1.02	Yes	0.31
Above 25th percentile	5.19	3.24	-1.22	Yes	0.23
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	2.44	4.03	1.30	Yes	0.20
At or above 50th percentile	4.03	4.31	0.14	Yes	0.89
Above 25th percentile	3.59	4.10	0.30	Yes	0.77

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B20

1998-2000 Science STAR Test Growth Comparisons for Partner High Schools vs. Matches

SCIENCE	Partner $N = 33$	Match $N = 35$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	0.56	0.17	-0.44	Yes	0.66
At or above 50th percentile	3.00	0.21	-1.39	Yes	0.17
Above 25th percentile	3.69	0.59	-1.02	Yes	0.31
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	0.75	1.21	0.42	Yes	0.67
At or above 50th percentile	2.25	0.62	-1.12	Yes	0.27
Above 25th percentile	2.53	0.10	-1.50	Yes	0.14
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	0.25	1.48	1.11	Yes	0.27
At or above 50th percentile	0.78	2.72	1.03	Yes	0.31
Above 25th percentile	1.63	2.93	0.63	Yes	0.53

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B21

1998-2000 Social Science STAR Test Growth Comparisons for Partner High Schools vs. Matches

SOCIAL SCIENCE	Partner $N = 33$	Match $N = 35$	*t*-Statistic	H_0: Equal Variance	*p*-value
NINTH GRADE					
Percent Scoring:					
Above 75th percentile	0.53	-0.90	-1.49	Yes	0.14
At or above 50th percentile	2.53	0.86	-0.81	Yes	0.42
Above 25th percentile	4.31	1.34	-1.00	Yes	0.32
TENTH GRADE					
Percent Scoring:					
Above 75th percentile	0.31	0.41	0.12	Yes	0.90
At or above 50th percentile	0.28	-0.10	-0.32	Yes	0.75
Above 25th percentile	2.16	-0.31	-1.82	Yes	0.07
ELEVENTH GRADE					
Percent Scoring:					
Above 75th percentile	0.28	1.21	0.69	Yes	0.49
At or above 50th percentile	0.59	3.48	1.44	Yes	0.16
Above 25th percentile	2.75	3.00	0.15	Yes	0.88

Note: Bold indicates that the means are statistically different at the 0.05 level.

Table B22

1998-2000 Reading STAR Test Growth Paired Comparisons for Partner Elementary Schools vs. Matches

READING	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
SECOND GRADE			
Percent Scoring:			
Above 75th percentile	0.2414	0.1947	0.8461
At or above 50th percentile	-0.0230	-0.0124	0.9901
Above 25th percentile	0.5632	0.2332	0.8161
THIRD GRADE			
Percent Scoring:			
Above 75th percentile	-0.6092	-0.6686	0.5055
At or above 50th percentile	0.0115	0.0083	0.9934
Above 25th percentile	0.8851	0.5321	0.5960
FOURTH GRADE			
Percent Scoring:			
Above 75th percentile	0.5287	0.3903	0.6972
At or above 50th percentile	-0.8621	-0.4903	0.6251
Above 25th percentile	1.7011	0.7748	0.4406
FIFTH GRADE			
Percent Scoring:			
Above 75th percentile	-0.3218	-0.3989	0.6910
At or above 50th percentile	2.4943	1.9321	0.0566
Above 25th percentile	2.4368	1.4319	0.1558

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B23

1998-2000 Mathematics STAR Test Growth Paired Comparisons for Partner Elementary Schools vs. Matches

MATHEMATICS	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
SECOND GRADE			
Percent Scoring:			
Above 75th percentile	0.7471	0.4262	0.6710
At or above 50th percentile	2.0920	0.8286	0.4096
Above 25th percentile	1.7241	0.6180	0.5382
THIRD GRADE			
Percent Scoring:			
Above 75th percentile	0.8621	0.6692	0.5052
At or above 50th percentile	0.8966	0.4699	0.6393
Above 25th percentile	3.8161	1.6675	0.0990
FOURTH GRADE			
Percent Scoring:			
Above 75th percentile	0.8621	0.5663	0.5727
At or above 50th percentile	2.6782	1.2828	0.2030
Above 25th percentile	0.9425	0.4269	0.6705
FIFTH GRADE			
Percent Scoring:			
Above 75th percentile	-0.0230	-0.0191	0.9848
At or above 50th percentile	-0.5862	-0.3194	0.7502
Above 25th percentile	1.2989	0.6286	0.5313

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B24

1998-2000 Reading STAR Test Growth Paired Comparisons for Partner Middle Schools vs. Matches

READING	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
SIXTH GRADE			
Percent Scoring:			
Above 75th percentile	2.6875	1.3780	0.1803
At or above 50th percentile	4.2500	1.1726	0.2499
Above 25th percentile	10.5000	1.7288	0.0938
SEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	1.5625	0.7482	0.4600
At or above 50th percentile	2.9375	0.8122	0.4229
Above 25th percentile	6.8438	1.1304	0.2670
EIGHTH GRADE			
Percent Scoring:			
Above 75th percentile	3.7188	1.6553	0.1079
At or above 50th percentile	5.9688	1.3607	0.1834
Above 25th percentile	9.8125	1.3658	0.1818

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B25

1998-2000 Mathematics STAR Test Growth Paired Comparisons for Partner Middle Schools vs. Matches

MATHEMATICS	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
SIXTH GRADE			
Percent Scoring:			
Above 75th percentile	3.0313	1.4730	0.1508
At or above 50th percentile	5.8125	1.4858	0.1474
Above 25th percentile	11.3125	1.7079	0.0977
SEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	2.1250	0.9159	0.3668
At or above 50th percentile	5.1875	1.3436	0.1888
Above 25th percentile	8.1875	1.2501	0.2206
EIGHTH GRADE			
Percent Scoring:			
Above 75th percentile	3.5625	1.8923	0.0678
At or above 50th percentile	4.2188	1.0098	0.3204
Above 25th percentile	10.3438	1.4608	0.1541

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B26

1998-2000 Reading STAR Test Growth Paired Comparisons for Partner High Schools vs. Matches

READING	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
NINTH GRADE			
Percent Scoring:			
Above 75th percentile	0.1786	0.2274	0.8218
At or above 50th percentile	1.1071	0.6281	0.5352
Above 25th percentile	1.6071	0.5572	0.5820
TENTH GRADE			
Percent Scoring:			
Above 75th percentile	1.5357	2.3793	0.0247
At or above 50th percentile	0.5714	0.4766	0.6375
Above 25th percentile	0.3214	0.1947	0.8471
ELEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	-0.2500	-0.3423	0.7348
At or above 50th percentile	-2.4285	-1.6945	0.1017
Above 25th percentile	-2.5000	-1.3572	0.1859

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B27

1998-2000 Mathematics STAR Test Growth Paired Comparisons for Partner High Schools vs. Matches

MATHEMATICS	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
NINTH GRADE			
Percent Scoring:			
Above 75th percentile	1.5357	1.0411	0.3071
At or above 50th percentile	4.1071	1.5004	0.1451
Above 25th percentile	3.6786	1.0380	0.3085
TENTH GRADE			
Percent Scoring:			
Above 75th percentile	0.6071	0.6029	0.5516
At or above 50th percentile	1.2500	0.7491	0.4603
Above 25th percentile	1.1786	0.6979	0.4912
ELEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	-1.5000	-1.0000	0.3262
At or above 50th percentile	-0.2500	-0.0987	0.9221
Above 25th percentile	-1.0714	-0.5577	0.5816

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B28

1998-2000 Science STAR Test Growth Paired Comparisons for Partner High Schools vs. Matches

SCIENCE	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
NINTH GRADE			
Percent Scoring:			
Above 75th percentile	0.6071	0.5785	0.5677
At or above 50th percentile	3.0357	1.4246	0.1657
Above 25th percentile	2.7500	0.8565	0.3993
TENTH GRADE			
Percent Scoring:			
Above 75th percentile	-0.4286	-0.3648	0.7180
At or above 50th percentile	1.3571	0.7969	0.4324
Above 25th percentile	1.7857	0.9396	0.3558
ELEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	-1.7857	-1.4424	0.1607
At or above 50th percentile	-2.5357	-1.1542	0.2585
Above 25th percentile	-2.3571	-0.9241	0.3636

Note: Test conducted assuming equal variance. Bold indicates that the mean of the paired differences is statistically significantly different from zero at the 0.05 level.

Table B29

1998-2000 Social Science STAR Test Growth Paired Comparisons for Partner High Schools vs. Matches

SOCIAL SCIENCE	Mean Difference Partner - Match	*t*-Statistic	*p*-Value
NINTH GRADE			
Percent Scoring:			
Above 75th percentile	1.2857	1.1721	0.2514
At or above 50th percentile	1.1786	0.5241	0.6044
Above 25th percentile	2.5714	0.7708	0.4475
TENTH GRADE			
Percent Scoring:			
Above 75th percentile	-0.0357	-0.0448	0.9646
At or above 50th percentile	0.3571	0.2731	0.7868
Above 25th percentile	2.1786	1.5041	0.1442
ELEVENTH GRADE			
Percent Scoring:			
Above 75th percentile	-1.2143	-0.8077	0.4363
At or above 50th percentile	**-4.0000**	**-2.0852**	**0.0466**
Above 25th percentile	-1.2857	-0.7271	0.4734