



Published in final edited form as:

Counc Undergrad Res Q. 2009 ; 29(3): 39–45.

Quantifying Success: Using Control Groups to Measure Program Effectiveness

Linda DeAngelo and Tama Hasson

CIRP Assistant Director for Research Higher Education Research Institute University of California, Los Angeles 3005 Moore Hall Los Angeles, CA 90095-1521 lindade@ucla.edu

Director, Undergraduate Research Center–Sciences Academic Director, PEERS University of California, Los Angeles 2121 Life Science Building Los Angeles, California 90095-1606 tama@lifesci.ucla.edu

In the current era of accountability, words like “assessment” have become part of the necessary lexicon of program directors running undergraduate-research programs and other student-development programs designed to serve underrepresented students. The pressure to demonstrate student-learning gains gathered momentum with the publication of the Spellings Commission Report (U.S. Department of Education, 2006), which emphasized the need to assess higher-education programs based on performance. This charge has been further supported by national agencies that fund student programs. They now mandate assessment in their requests for applications and have developed requests that specifically focus on development of rigorous assessment mechanisms for enrichment programs in the STEM fields (science, technology, engineering, and mathematics).¹

Explicit in these calls is the need for assessment that goes beyond mere student satisfaction with programs. We need to design mechanisms to measure and quantify if the goals of a program are being met and to identify which program components are contributing effectively to student success. The extensive literature reviews done by Crowe & Brakke (2008) and Seymour, Hunter, Laursen & DeAntoni (2004) reveal that most studies of undergraduate-research programs are descriptive in nature. Other studies appear promising but either lack control groups and are therefore less able to quantify success (e.g., Hunter, Laursen, Seymour, 2006) or are only able to assess long-term outcomes and not the success of particular components of programs or short-term program goals (Ishiyama, 2001).

In this article we use assessment of the Program for Excellence in Education and Research in the Sciences (PEERS) to demonstrate how control groups can be used to effectively measure the success of undergraduate-research programs in STEM fields. PEERS is a two-year program for freshmen and sophomores at UCLA that enrolled its sixth cohort of students in the 2008–09 academic year. Approximately 80 students enter the program each year. All are physical and life science majors who come from underrepresented groups² and have had more than the typical challenges to overcome to reach UCLA.

¹See <http://grants.nih.gov/grants/guide/rfa-files/RFA-GM-09-011.html>, among others.

²PEERS students are underrepresented racial/ethnic minorities and/or socioeconomically underrepresented.

PEERS is a complex program with many components. We have found it necessary to take into account different elements to create an appropriate control group. This often involves trial and error. In other words, control groups that might serve to assess a particular program component are not necessarily appropriate controls for other components. We feel that this study of PEERS is particularly illuminating in terms of demonstrating the complexity and utility of using post-matching controls (Schlesselman, 1982) for program assessment.

The PEERS Program

PEERS has four major goals: 1) to facilitate the development of effective study skills that will assist PEERS students in achieving competitive grades in basic math and science courses; 2) to increase the likelihood that PEERS students will stay in a STEM major (improved retention); 3) to encourage and prepare PEERS students for entry-level undergraduate research; and 4) to foster PEERS students' interest in and commitment to preparing for careers in research, teaching, or health fields.

To achieve this set of goals, PEERS offers seminars that teach effective study skills, as well as workshops to supplement learning in core science classes. PEERS students are given one-on-one academic counseling, priority in course enrollment, and invited to partake in peer mentoring by upper-division under-represented science students involved in research. As part of their exposure to on-campus research, PEERS students attend two research talks by UCLA faculty members each quarter and meet undergraduate researchers from underrepresented groups at poster sessions and on student panels. Students attend a course on choosing majors and identifying the careers possible with a science major. They also are given a wide variety of opportunities to enter into research on the UCLA campus and, once involved, PEERS students are eligible for stipends to support their endeavors.

A major component and expense of running the PEERS program are the workshops. Termed EXCEL (Excellence through Collaboration for Efficient Learners), these workshops in math, chemistry, physics, and life sciences (biology) are one-unit, pass/no pass classes that supplement the main lecture in a course. The leader (or facilitator) is an experienced graduate student or a postdoctoral fellow with a proven record of exceptional teaching. The facilitator leads small groups through in-class problems, following the model first described by Uri Treisman (Fullilove, Treisman, 1990). The EXCEL workshops were designed to introduce students to the scientific method, to enhance understanding, and to build confidence. During the workshops the students bond together as a community of learners as they successfully meet the challenge of solving new problems.

Gafney & Varma-Nelson (2008) have collected the extensive body of literature on the implementation of collaborative workshops based on the Treisman Model. Most of the articles in their volume are program summaries, and those that do address student performance in the workshops most often use control groups in which the comparison group of students differs from students in the workshops with respect to race/ethnicity or incoming academic preparation (Born, Revelle, Pinto, 2002; Drane, Smith, Light, Pinto, Swarat, 2005; Gafney, 2001; Tien, Roth, Kampmeier, 2002). A select group of studies have taken these variables into account when considering workshop success, although what counts as

workshop success varies, with some studies stating that a B- or better is success and others stating that a passing grade of C-or better is sufficient (Chinn, Martin, Spencer, 2007; Murphy, Stafford, McCreary, 1998; Rath, Peterfreund, Xenos, Bayliss, Carnal, 2007).

Using Control Groups to Quantify Success

The goal of PEERS assessment is to determine whether the students who have participated in PEERS do better than similar students who are not in the program. This type of assessment meets the rigorous standards that funding agencies now require for continued program support. We describe two examples of how to create post-matching control groups to assess undergraduate-research programs in STEM fields. The examples are drawn from our ongoing assessment of the PEERS program. Although PEERS serves both socio-economically and racially or ethnically underrepresented students, the examples below are limited to underrepresented racial-minority students (URM)³. We use independent sample t-tests to test for group differences between the PEERS and control groups.



PEERS students asking questions after attending a faculty research talk (L to R): Michael Daniel, Dr. Mark Frye, Rebeca Fuentes Ramirez and Kayla Gurley.

Retention in STEM and Effectiveness of EXCEL Workshops

The end of the first year for a cohort of PEERS students provides an opportunity to assess the impact of PEERS on retention in the sciences and first-year grades in core science courses. Therefore, as a first criterion for constructing control groups, we limited the control group to declared science majors who would therefore be eligible for PEERS. In addition, because in the examples for this article we only are analyzing PEERS students from underrepresented racial-minority groups, the control pool was similarly limited to URM students. However, we found that simply restricting our control group to only this set of criteria was inadequate for assessing both of these questions.

Using UCLA's 2007 freshman class for this example, there were 284 underrepresented racial-minority students who entered as declared science majors; 40 of these students joined the PEERS program, leaving a potential control group of 244 students. Of these 244 students, 89.7 percent began their second year at UCLA as science majors. This compares to a 97.5-percent retention rate for students who began PEERS; only one student in that cohort entered the second year in a non-science major. Given the retention question and eligibility

³URM students are African American/Black, Mexican American/Latino or Native American.

for selection for PEERS, this was deemed an appropriate control group to examine retention in the first year.

Assessing the effect of the EXCEL workshops on core science courses is not as easy as assessing the effect of PEERS on retention in sciences in the first year. The first complication is with the PEERS group. Of the 40 students in the cohort, six students participated in the PEERS workshops for a portion of the year, and therefore likely received some—but not the full—benefits of participation. Thus we eliminated these six students from the PEERS group, and they could not be used in the control group because they had had some exposure to the PEERS workshops. This limited the PEERS group to the remaining 34 students. For this first example, we will use the original control group of 244, as we defined it above, as an example of how a control group that is appropriate to assess one area of a program is not appropriate to assess another area.

As shown in the first column of Table 1, there is a significant difference between the PEERS EXCEL workshop participants and the control group on core math and science grades. In math, the control group averaged a 2.49, between a C+ and a B– grade. This score is expected, as most introductory courses are curved such that the median grade is a C+/B–. PEERS workshop participants averaged a 3.16 or B/B+ grade. In chemistry, the control group received a C grade (1.98), whereas the PEERS workshop participants averaged a 2.48, between a C+/B– grade. But based on this evidence, should we be satisfied and ready to state that the EXCEL workshops are effective and that PEERS students are getting the desired bump in grades from their participation in these workshops?

To answer these questions we must return to our criteria for admission to PEERS. Applicants to PEERS are selected based on a variety of criteria, including grades in their high-school science classes; their MSAT score, a strong predictor of success in entry-level science classes (Burton, Ramist, 2001; Duncan, Dick, 2000); and their answers to a set of essays asking about their life challenges and career goals. In addition, each student admitted to UCLA is given a Life Challenge (LC) Score. Among other factors, the LC scores range from zero (no challenges) to 12 and take into account the quality of the high school attended, access to Advanced Placement courses, the socioeconomic level of the family, the parents' education level, and whether English is spoken in the home. The median LC score at UCLA is a 2. The median LC score for the 34 PEERS students is 5.0 (3.4 SD). For the control group of 244, the median LC score was a 6.0 (3.7 SD), a difference that is not statistically significant, suggesting that this group could indeed serve as an appropriate control group.

On the other hand, the average MSAT for PEERS students was 632 (73 SD) and the average for the control group was 556, statistically different ($p < .001$). This result is not entirely unexpected however. Beginning with the 2006 PEERS class, a 540 cutoff for the MSAT was instituted for entry into PEERS. The EXCEL workshops are honors workshops and not remedial in nature and, from our experience, a student with a score lower than 540 need assistance with academics that is not provided in the workshops.⁴ Therefore, it is necessary

⁴As an example, in 2007 44% of students with MSAT < 540 failed (received a D or F) in at least one science class their first year. Whereas, this figure is 25% for students with MSAT > 540.

to eliminate from the control group those students with MSAT scores less than 540. This reduced the control group to 135 students, and brought the average MSAT up to 619 (60 SD) and the LC to 4.3 (3.6 SD). Both of these scores were not statistically different from the PEERS group.

We then repeated our analysis of the impact of the EXCEL workshops on math and chemistry grades in the first year. As shown in the second column in Table 1, restricting the control group to students with MSAT > 540 increased the average grades of students in the control group for both math and chemistry. In chemistry the control group now averaged a grade of 2.20 (C/C+), a grade that was not statistically different than the grades for the PEERS students in the workshops (2.48, C+/B-). In math the PEERS students in the workshops maintained their statistically significant advantage over the controls—3.16 (B/B+) for PEERS and 2.69 (B-) for the controls.

Here we again asked ourselves, is this the appropriate control group for evaluating the success of the EXCEL workshops? PEERS students are required to take two sciences courses each term, a total of six per year. While not a variable considered in any of the other published studies we could find on evaluating collaborative-learning workshops, we considered enrollment in science courses a potentially significant factor.

It turns out that among students in the control group, the average number of science courses taken per year was 3.43, whereas among the PEERS students the average was 5.85, a statistically significant difference ($p < .001$). We argue that it might be possible that the control group achieved higher grades in their core science courses than they might otherwise have because they were experiencing a less-demanding science course load in their first year. Therefore we decided to restrict our control group to the students with a statistically similar course load, creating a control group of 66. This control group continues to be statistically no different from the PEERS group on MSAT, but had a significantly lower LC score ($p < .05$), suggesting that this control group could actually outperform the PEERS group.



PEERS students attending UCLA Science Poster Day 2008. (L to R): James Aboagye, Yolanda Tinajero, Ivan Rodriguez-Pinto (poster presenter), Ifedayo Akinyemi, Miguel Alvarez, and Krystal Jimenez.

Our hypothesis that a less-demanding course load might have artificially inflated the average grades of our control group in their core math and science grades did not hold; in fact, the

opposite was true. As shown in the third column of Table 1, in math the average grade among the controls increased 4 percent to 2.80, and in chemistry the average grade among the control inched up 1.8 percent to 2.24. With this control group, the PEERS workshop participants continue to hold a statistically significant advantage in math grades and no advantage in chemistry.

In the final analysis, there was little to gain in terms of explanatory value by restricting the control group to only students with similar course-taking patterns and, in fact, this control group proved too restrictive given that its Life Challenge score was statistically different than the PEERS group. That said, without having delved into further comparative analysis between the controls and the PEERS group, we might have biased our results in favor of the control group. In terms of the PEERS program, these results indicate that we are seeing the predicted grade bump for students enrolled in math; however, the chemistry workshops are not functioning as well. Given these results, a major focus of the current and coming years will be to improve our EXCEL workshops, the chemistry workshops, in particular, to ensure they are meeting students' needs.

Independent Undergraduate Research Engagement⁵

Assessing the effect of participation in PEERS on research engagements, a long-term outcome of the PEERS program, is even more complex than assessing a short-term outcome like retention in the sciences and grades in core science courses in the first year. In order to capture research engagement during the entire undergraduate experience, we will use the 2003 PEERS cohort as an example. We start with a control group of 154 students who entered UCLA in 2003 as declared science majors and did not participate in PEERS and a group of 31 students who completed the PEERS program.

As shown in the first column of Table 2, among the PEERS graduates, 19 (61.3 percent) participated in a significant research experience, and the median length of their research engagement was eight terms (two full years at UCLA). Among the control group, 29 (18.8 percent) participated in a significant research experience, the median length of which was two terms, a statistically significant difference. Among the PEERS graduates, all did their research in the sciences. In contrast, of the 29 in the control group who tried research, only 17 (58.6 percent) did research in the sciences.

Further examination of the initial control group revealed that of the 154 students, 15 (9.7 percent) were dismissed from UCLA for academic reasons, and another 15 (9.7 percent) never took a science course and left the science major, although one of these non-science students did research outside of the sciences. In addition, 51 (33.1 percent) took fewer than five quarters of science in the first two years and left the science major before the end of their sophomore year. Four of these non-science students did research outside of the sciences. Since you must be a science major and take at least five quarters of science to graduate from PEERS, all of these students had to be eliminated from the control group,

⁵Research engagements included enrollment in lower division honors independent research courses (Course 99), enrollment in upper division independent research course (Course 196, 198, 199), participation in summer research programs on the UCLA campus, participation in funded research programs administered through the URC, participation in departmental and campus wide poster session.

reducing it to 77 students with backgrounds similar to the PEERS students. All had taken at least five quarters of science and were science majors at the end of their sophomore year. Before running our analysis, we compared this control group to the PEERS group on MSAT and LC and found no significant differences. This indicated to us that this was the appropriate control group and that no further analysis of difference between the groups would be necessary.

Repeating our analysis, our results for the control group now show that 24 students participated in research (31.2 percent), but the median length of research remained two terms, significantly different from that of the PEERS group (Table 2, second

FRESHMAN YEAR IN PEERS
Fall Freshman Seminar: Study Skills, Time Management, Adjustment to University Life & Intro to Research
Collaborative Learning EXCEL workshops to accompany core science classes (3-6 per year)
Enroll in and complete 5-6 science classes in the first year
PEERS research talks (6 per year)
PEERS social events (2 per year)
Meetings with PEERS and Departmental Counselors (4 meetings per year minimum)
Lab Tours (2 per year)
Undergraduate Science Research Poster Sessions (2 per year)
Invitation to begin undergraduate research in the summer following Freshman year
SOPHOMORE YEAR IN PEERS
Fall Sophomore Seminar: Choosing a Major, Choosing a Career, Entry into Independent Research
Collaborative Learning EXCEL workshops to accompany core science classes (3-6 per year)
Enroll in and complete 5-6 science courses in the first year
PEERS research talks (6 per year)
PEERS social events (2 per year)
Meetings with PEERS and Departmental Counselors (4 meetings per year minimum)
Counseling on Research entry
Invitation to apply to funded research programs on campus
For more information about the UCLA PEERS Program please visit the program web site: http://www.uclaeducation.ucla.edu/urc-care/progpeers.htm

column). Notably, of the 24 students in the control group who participated in research, four did research outside of the sciences. From these results it is clear that participating in PEERS and finishing the program has a significant impact not only on the chance that a student will participate in research, but also on the length and substance of that research engagement.⁶

Discussion and Conclusion

The use of post-matched control groups has allowed us to pinpoint areas within the complex PEERS program where there is programmatic success or a specific need for improvement. This impacts our decision making as we strive to meet the aims of PEERS and meet students' needs. It also provides the necessary data on short-, medium- and long-term outcomes to fulfill the accountability mandates set by our funding agencies.

When considering using control groups to assess student-research programs, it might seem wise at first to use a direct matching approach, whereby individuals who exactly match the incoming characteristics of each student in a program are selected to be part of a control group. However, even at a very large university like UCLA, this was impossible due to the lack of a sufficiently large sample size of STEM students from underrepresented racial-minority groups. Based on our experience, we would argue that using post-matched controls is a realistic and valid approach, even if additional time is required to settle on an appropriately limited control group for each area assessed. It is only through this trial-and-

⁶In addition, it is notable that only 3 (7.5%) of the 2003 PEERS students graduated or are set to graduate with a major other than science, whereas among the control 21 (27.3%) graduated or are set to graduate with a major other than science. Additionally, 11 (14.3%) of the control group were dismissed from UCLA for academic reasons or withdrew prior to graduation. None of the PEERS graduates were dismissed or withdrew prior to graduate

error approach that a researcher can confidently state the impact of a program and its components on student success.

The identification of control groups, even for post-matching controls, requires a large incoming class of underrepresented racial-minority students, as well as extensive access to both admissions records and enrollment records. For assessment of research involvement for the control group, further connections with departmental counselors (who enroll students in independent-research classes) and the scholarships office that supports undergraduate research are necessary. Most directors of STEM programs are scientists and, as a result, do not have access to these critical records. Such records are private, and access is subject to institutional approval by human-subjects boards. Therefore, collaboration between program directors and institutional evaluation experts is a vital link for the success of STEM program evaluation.

Biographies

Linda DeAngelo is Assistant Director for Research for the Cooperative Institutional Research Program at the Higher Education Research Institute, UCLA. She previously was a postdoctoral scholar at the Center for Educational Assessment at UCLA, where she conducted research and assessment on the PEERS program and other innovative science programs on campus.

Tama Hasson is Director of the Undergraduate Research Center for Science, Engineering & Math and directs the PEERS program, among others, on the UCLA campus. She previously was an assistant professor of cell and developmental biology at the University of California, San Diego.

References

- Born WK, Revelle W, Pinto LH. Improving biology performance with workshop groups. *Journal of Science Education and Technology*. 2002; 11(4):347–365.
- Burton, NW.; Ramist, L. *Predicting Success in College: SAT Studies of Classes Graduating Since 1980*. College Entrance Examination Board; New York: 2001.
- Chinn D, Martin K, Spencer C. Treisman workshops and student performance in CS. *ACM SIGCSE Bulletin*. 2007; 39(1):203–207.
- Crowe M, Brakke D. Assessing the impact of undergraduate research experiences on students: An overview of current literature. *CUR Quarterly*. 2008; 28(4):43–50.
- U.S. Department of Education. *A Test of Leadership: Charting the Future of U.S. Higher Education*. U.S. Department of Education; Washington DC: 2006.
- Drane D, Smith HD, Light G, Pinto L, Swarat S. The gateway science workshop program: Enhancing student performance and retention in the sciences through peer-facilitated discussion. *Journal of Science Education and Technology*. 2005; 14(3):337–352.
- Duncan H, Dick T. Collaborative workshops and student academic performance in introductory college mathematics courses: A study of a Treisman model math excel program. *School Science and Mathematics*. 2000; 100(7):365–373.
- Fullilove RE, Treisman PU. Mathematics achievement among african american undergraduates at the University of California, Berkeley: An evaluation of the mathematics workshop program. *Journal of Negro Education*. 1990; 59(3):463–478.
- Gafney L. Evaluating student performance. *Progressions the PLTL Project Newsletter*. 2001; 1(2):3–4.

- Gafney, L.; Varma-Nelson, P. Peer-Led Team Learning: Evaluation, Dissemination, and Institutionalization of a College Level Initiative. Vol. Vol 16. Springer; 2008.
- Hunter A-B, Laursen SL, Seymour E. Becoming a scientist: The role of undergraduate research in students' cognitive, personal and professional development. *Science Education*. 2006; 91(1):36–74.
- Ishiyama J. Undergraduate research and the success of first generation, low income college students. *CUR Quarterly*. 2001; 22:36–41.
- Murphy TJ, Stafford KL, McCreary P. Subsequent course and degree paths of students in a Treisman-style workshop calculus program. *Journal of Women and Minorities in Science and Engineering*. 1998; 4(4):381–396.
- Rath KA, Peterfreund AR, Xenos SP, Bayliss F, Carnal N. Supplemental instruction in introductory biology I: Enhancing the performance and retention of underrepresented minority students. *CBE Life Sci Educ*. 2007; 6(3):203–216. [PubMed: 17785403]
- Schlesselman, JJ. *Case-Control Studies Design, Conduct, Analysis*. Oxford University Press; Oxford: 1982.
- Seymour E, Hunter A-B, Laursen SL, DeAntoni T. Establishing the benefits of research experiences for undergraduates: First findings from a three-year study. *Science Education*. 2004; 88:493–534.
- Tien LT, Roth V, Kampmeier JA. Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *Journal for Research in Science Teaching*. 2002; 37:606–632.

Table 1

Comparison of First-Year Math and Chemistry Grades at Various Levels of Control 2007 PEERS URM Workshop Attendees and Other Non-PEERS STEM URM Students

	Math****		Chem****		Math***		Chem		Math*		Chem	
	GPA	SD	GPA	SD	GPA	SD	GPA	SD	GPA	SD	GPA	SD
Control	2.49	.98	1.98	.98	2.69	.92	2.20	.94	2.80	.89	2.24	.95
PEERS	3.16	.75	2.48	.90	3.16	.75	2.48	.90	3.16	.75	2.48	.90

Note:

**** p<.001,

*** p<.01,

* p<.05

Table 2

Comparison of Independent Undergraduate Research Engagement at Various Levels of Control 2003 PEERS URM Graduates and Other Non-PEERS STEM URM Students

	Control Level 1		Control Level 2	
	Control N=154	PEERS N=31	Control N=77	PEERS N=31
Average Number of Qtrs Research	.2***	4.6	.38***	4.6
Median Qtrs All Students	0	2	0	2
Median Qtrs Research Participants	2***	8	2***	8
Number of Research Participants	29	19	24	19
Percentage of Research Participants	18.8	61.3	31.2	61.3

Note:

p<001,

**
p<01