CBE—Life Sciences Education
Vol. 9, 119–132, Summer 2010

Article

Promoting Undergraduate Interest, Preparedness, and Professional Pursuit in the Sciences: An Outcomes Evaluation of the SURE Program at Emory University

Benjamin Junge,* Catherine Quiñones,† Jakub Kakietek,‡ Daniel Teodorescu,§ and Pat Marsteller†

*Department of Anthropology, State University of New York/New Paltz, New Paltz, NY 12561; †Center for Science Education, ‡Department of Political Science, and §Provost's Office, Emory University, Atlanta, GA 30322

Submitted August 14, 2009; Revised January 6, 2010; Accepted January 17, 2010
Monitoring Editor: Debra Tomanek

We report on an outcomes assessment of the Summer Undergraduate Research Experience (SURE) Program at Emory University in Atlanta, GA. Using follow-up survey data and academic transcripts, we gauge SURE's impact on levels of interest in, preparedness for, and actual pursuit of graduate study and professional careers in the sciences for the program's first 15 summer cohorts (1990–2004). Our follow-up survey indicated significant increases in all research preparedness skills considered, notably in ability to give a poster research presentation, to discuss research at a graduate school interview, and to apply research ethics principles. About a third of SURE graduates went on to complete a graduate degree >90% considered SURE as important or very important in their academic development. Respondents reported postprogram increases in the level of interest in academic and research careers, and reported high levels of employment in science careers and job satisfaction. Regression analyses of Emory SURE participant transcripts revealed that participants take significantly more science courses as seniors and earn higher grades in those courses than nonparticipants. This trend held after correcting for indicators of prior interest (first-year course work, GPA, and math SAT scores), gender, and minority status. We also report on an external survey completed by SURE participants.

INTRODUCTION

Since the late 1990s, undergraduate science education policy, programming, and funding in the United States have shifted increasingly toward the integration of teaching and research. Earlier proponents of undergraduate research (Okorodudu, 1970; Fields, 1971; Pace, 1979; Carsrud, 1980; Peppas, 1981; Palladino, 1982; Prentice-Dunn and Roberts, 1985; Malanga, 1988; Kremer and Bringle, 1990; Mauger, 1990; Delagarza *et al*., 1991; Schowen, 1998) saw their recommendations validated in the Boyer Commission on Educating Undergraduates in the Research University (1998) report call for a modern undergraduate education blueprint featuring inquiry-based, hands-on learning. Fueled by National Science Foundation's (NSF) Integration of Research and Education grants and with support from foundations such as the Howard Hughes Medical Institute, appreciation for the multiple pedagogic, institutional, and professional benefits conferred by undergraduate research has grown (Table 1) and so has the number of undergraduate research (UR) programs in major research universities and small liberal arts colleges (Mervis, 2001a,b).

Following calls in the late 1990s for evaluation components in UR programs (Manduca, 1997; and Spilich, 1997; Schowen, 1998) and concerns over the quality of previous evaluations (Adhikari and Nolan, 2002; Bauer and Bennett, 2003), a growing number of authors have sought to quantify the benefits conferred by UR. Studies range from quantitative before-after surveys to ethnographic accounts, and include both individual case studies and large-sample studies that pool data from multiple UR settings (Table 1). The

DOI: 10.1187/cbe.09–08–0057
Address correspondence to: Benjamin Junge (jungeb@newpaltz.edu).

Table 1. Student gains, benefits of UR, and recent evaluation models

Student gains and benefits of UR
Thinking, working and self-identifying as a scientist; personal and professional skills relevant to a scientific career; clarification, confirmation, and refinement of educational/career path; enhanced preparation for advanced (graduate) training and science career; development of technical skills and ability to work independently (Seymour *et al*., 2004)*
Understanding concepts relevant to research field; using primary literature; identifying research question; formulating research hypothesis and designing an experimental or theoretic test thereof; understanding the importance of controls; observing, collecting and statistically analyzing data; interpreting data relative to original hypothesis and reformulating research question as appropriate; interpreting results in context of research field; communicating research orally and in writing; thinking independently (Kardash, 2000)*
Understanding the research process and how scientists work on real problems; readiness for more demanding research; learning lab techniques; obstacle tolerance; working independently; interpreting results; analyzing data and integrating theory and practice; understanding how knowledge is constructed and that assertions require supporting evidence; becoming part of the learning community; understanding science and how scientists think; understanding the primary literature; learning ethical conduct; clarifying career path; developing oral presentation and science writing skills (Lopatto, 2004a,b, 2007)
Self-confidence gains (Gregerman, 1999; Adhikari and Nolan, 2002; Ward *et al*., 2002; Lopatto, 2004a,b, 2007; Frantz *et al*., 2006; Seymour *et al*., 2004)
Development of an identity as a scientist (Gafney, 2001; Burnley *et al*., 2002; Schowen, 2002; Merkel, 2003; Lopatto, 2004b, 2007; Millspaugh and Millenbah, 2004; Seymour *et al*., 2004*; Knox *et al*., 2006)
Beneficial for women (Campbell and Skoog, 2004)
Beneficial for underrepresented minorities (Alexander *et al*., 1998; Gregerman, 1999; Gafney, 2001; Matsui *et al*., 2003; Barlow and Villarejo, 2004; Crowe, 2006; Summers and Hrabowski, 2006)
Beneficial to first-generation, low-income college students (Ishiyama, 2001)
Capacity to improve the faculty-student relationship (Mervis, 2001b; Knox *et al*., 2006; Winthrop University Undergraduate Research Advisory Committee, 2006)
Improve undergraduate retention rates (Jonides, 1995; Nagda *et al*., 1998; Weaver *et al*., 2008)
Improve the reputation of the host institution (Page *et al*., 2004; Coleman, 2005; Elgren and Hensel, 2006)
Survey of UR assessment
Single-site assessments (Kremer and Bringle, 1990; Foertsch *et al*., 1997; Alexander *et al*., 1998; Kardash, 2000; Adhikari and Nolan, 2002; Burnley *et al*., 2002; Hathaway *et al*., 2002; Schowen, 2002; Zydney *et al*., 2002; Bauer and Bennett, 2003; Shachter, 2003; Barlow and Villarejo, 2004; Dirks and Cunningham, 2006; Frantz *et al*., 2006; Page *et al*., 2004; Gum *et al*., 2007)
Multiple-site assessments (Mabrouk and Peters, 2000; Merkel, 2001, 2003; Lopatto, 2004a,b, 2007; Seymour *et al*., 2004; Russell *et al*., 2006, 2007)
Assessment literature review (with summaries of papers grouped by programs evaluation vs. student experience evaluation focus) (Crowe and Brakke, 2008)

An asterisk denotes studies that included responses from student researchers and their research mentors.

growing corpus of evaluative studies of UR to date, however, focuses minimally on the intensive, mentor-based summer program model (cf. Kardash, 2000; Seymour *et al*., 2004). To be sure, summer UR programs abound and many of them have been described or, in some manner, assessed (Kremer and Bringle, 1990; Foertsch *et al*., 1997; Alexander *et al*., 1998; Kardash, 2000; Adhikari and Nolan, 2002; Burnley *et al*., 2002; Schowen, 2002; Zydney *et al*., 2002; Shachter, 2003; Lopatto, 2004b; Page *et al*., 2004; Frantz *et al*., 2006; Seymour *et al*., 2004; Gum *et al*., 2007; Butler *et al*., 2008). Few studies, however, use methodologically rigorous designs and include both program participants and nonparticipants in assessing outcomes linked to participation.

This article reports on an evaluation of the Summer Undergraduate Research Experience (SURE) Program at Emory University in Atlanta, GA. The evaluation began in the spring of 2005 to assess short- and long-term outcomes associated with participation for the program's first 15 summer cohorts. By means of two methodologically distinct studies—one study grounded in self-reported follow-up survey data (1990–2004 cohorts) and the other study in academic performance data from course work transcripts from a subsample from the 1990–2008 cohorts—we sought to gauge levels of interest in, preparedness for, and actual pursuit of graduate study and professional careers in the sciences. To address gains by more recent participants, we present data gathered by David Lopatto's SURE II-III surveys (www.grinnell.edu/academic/psychology/faculty/dl/sure&cure) that summarize self-perceived gains by 2007–2009 Emory SURE program respondents. We also briefly discuss measures of participant satisfaction per our internal end-of-program survey.

THE SURE PROGRAM AT EMORY: PARTICIPANTS AND COMPONENTS

SURE is a 10-wk residential program that provides full-time, mentored research training in mathematics, life and natural sciences, psychology, and anthropology to undergraduates from Emory and other U.S. undergraduate institutions. SURE is organized and administrated out of the Emory College Center for Science Education (ECCSE), which focuses on the promotion of undergraduate access, interest, and participation in the sciences. Funded primarily by the Howard Hughes Medical Institute (HHMI), and with variable support from sources such as the NSF, Emory University, and the National Institutes of Health, the SURE program has been in operation since 1990. The ECCSE focuses on recruitment and training of U.S. citizens in science fields

(NSF, 2005) and encourages "young people's interest in science, prepare(s) them for science-related careers, and increase(s) science literacy among all students, including non-science majors" (HHMI; www.hhmi.org/grants/institutions). SURE began by serving students working with biomedical sciences faculty. In 1995, we began supporting students in psychology and anthropology; an ethics component was added that year as well. For the period addressed in this study, 218 colleges were represented, and >320 Emory faculty served as mentors. Internal, formative assessments have been conducted each year to review program content and participant satisfaction.

Although applicants can seek appointments with any Emory science faculty, we maintain an online directory of potential research mentors. The majority of our mentors have hosted multiple students through the years, sometimes in tandem. Our program's success hinges on our mentor's belief that SURE activities and infrastructure facilitate and enhance their ability to provide meaningful, well-rounded research experiences.

Program Characteristics and Components

SURE takes place from late May until early August, during which time participants reside in an on-campus dormitory and devote 40 h/wk to program activities. After orientation and training in laboratory safety and research ethics, participants spend the bulk of their time at their mentor's research facility (typically a laboratory). In this setting, students devise, refine, and implement an individual research project, with supervision and training in methods and data analysis from the faculty mentor. Fellows share their research proposal (which includes background on their project, a clearly stated hypothesis, methodology, a 10-wk timeline, discussion of foreseeable differences and data analysis plan, and references; see Supplemental Material 1) on the second week of the program via our intranet. Sharing research plans early in the program paces student progress and allows students to learn about one another's work. Students receive additional preparation and guidance at weekly "Science Careers" meetings, which address career options, mentorship and professional skills, graduate school funding options, and how to prepare and present a scientific poster. To encourage audience participation and keep contents current, whenever feasible, we use a panel format (e.g., separate graduate student panel and faculty recruiter panels present a layered, multiple-perspective view of the graduate school application process; nonacademic scientists discuss their training, career paths, and current employment trends in their areas of expertise).

The program's ethics training component involves approximately 15 h of contact time, and begins with a day-long role-playing activity that introduces topics such as authorship, collaboration, data collection and ownership, and research integrity; these topics are discussed more fully in weekly small group meetings. These meetings are student-led, with a directive that the presentation should include background on the topic and include audience participation (e.g., via the use of a case study). We encourage students to query their research team as they research their topic (e.g., ask each laboratory member questions such as "how is authorship determined in this laboratory?"); mentors report these inquiries often trigger discussions that might otherwise not happen and which benefit all lab members. Fellows also have the option of attending weekly "Lunch with a Scientist" meetings (informal networking opportunities with faculty), participating in library skills workshops, attending a skills session for the Graduate Record Examinations, and meeting with MD/PhD program students. The final week concludes with a Research Poster Symposium that is judged by advanced graduate students and open to the Emory community. Both participants and mentors have access to judging criteria early in the program. Participants receive the judges' feedback by mail and the best posters receive cash prizes that are announced at our closing banquet and on the program website.

Eligibility and Selection Criteria

Applicants must be enrolled as full-time students in either a 4-yr college (working toward a bachelors degree) or a community/preparatory college (working toward an associates degree) and cannot have graduated before program participation. Applicants from Emory must identify a research mentor before applying and submit a mentor-approved proposal as part of the application as well as a letter of recommendation from the supporting mentor (a requirement that dramatically improved the quality of submitted proposals when we implemented it in the late 1990s). We typically review >300 external applications, and 60–70 applications from Emory students. Approximately half of applicants accepted for SURE participation are from Emory, and the remaining half typically represent 20–25 schools. Following the priorities of the U.S. government and HHMI to increase racial and gender diversity in the sciences, ECCSE also endeavors to accept a sizable proportion of women, underrepresented minorities, and small liberal arts college students into its annual participant pool.

Profile of SURE Participants, 1990–2004

During its first 15 yr, 822 individuals participated in SURE, with cohort size increasing from 10 students in 1990, to between 20 and 30 students for 1991–1994, and to 50–70 students from 1995 to 2004 (overall mean cohort size, 54.8 students). Most participants are rising seniors (70.4%) or rising juniors (23.7%). Over the years, approximately half (50.4%) of SURE participants have been from Emory and half from other schools. Of non-Emory students, more than two-thirds (68.9%) came from schools outside of Georgia, and more than one-quarter (28.5%) were from small liberal arts colleges. Overall, slightly more than one-quarter of participants (26.9%) were minority students and 15.9% came from historically minority colleges. More women participated in SURE than men (59.0 vs. 41.0%) and mean age overall was 21.3 yr old (range, 17.7–51.0, SD 2.6).

Emory students were more likely to be men (49.1 vs. 32.4%; $p < .001$) and less likely to be of a minority race/ethnicity (13.0 vs. 41.7%; $p < .001$). Comparison by cohort cluster revealed that the proportion of minority SURE participants increased significantly during the 1990–1994/1995–1999/2000–2004 clusters (20.0, 26.1, and 30.7%, respectively; $p = 0.050$). More than 90% of SURE participants pursue science-intensive majors at their home institutions (with more than three-quarters in the life and natural sciences, and with biology the most common major at 40.7%). One-fifth of participants (19.8%) pursue double majors, and the degree most typically completed by SURE participants is a bachelor of science (79.2%). The overall self-reported undergraduate grade point average (GPA) of SURE graduates is 3.63.

Within a subsample of 2002–2004 SURE participants who completed an intake questionnaire ($n = 95$), most (96.8%) had laboratory-related course work backgrounds. However, less than half (47.4%) had any prior experience in a mentor-supervised UR setting, and two-fifths (40.0%) have no UR experience whatsoever. Finally, two-thirds of the 2002–2004 cohort (67.4%) of SURE participants have no background in research ethics training.

METHODS

This evaluation draws upon four sources of data: 1) a study of follow-up data for 1999–2004 participants gauging interest, preparedness, and pursuit (as related to science careers outcomes); 2) transcript analysis of Emory SURE fellows relative to peers who did not participate in SURE (1999–2008) participants; 3) postprogram survey data (SURE 2007–2009) provided by the SURE II-III team (D. Lopatto, Grinnell College, Grinnell, IA); and 4) participant satisfaction data from exit surveys completed by SURE participants (2002–2009).

Study 1. Online Follow-Up Survey

This study gauges associations between SURE participation and three sets of outcome variables, conceptualized as "interest," "preparedness," and "pursuit." Interest and preparedness variables relate to short-term outcomes, namely, the extent to which SURE stimulates interest in the sciences and provides skills and experience germane to graduate study or a career in the sciences, respectively. Pursuit variables concern SURE's impact on postundergraduate academic training and professional career trajectories.

Self-report data come from an online follow-up survey (hereafter referred to as OFS) conducted during spring 2005. (See Supplemental Material 2 for survey questions.) To recruit respondents, all SURE graduates for whom a viable email or street address was available were contacted and invited to complete the online questionnaire (following an informed consent protocol approved by Emory University's Institutional Review Board). Of 743 individuals contacted, 250 completed the survey (with a response rate of 33.6%). Survey completion rates were marginally higher among racial/ethnic minorities (32.6 vs. 25.0%; $p = 0.040$) and noticeably higher among more recent SURE cohorts (with more than half of respondents coming from the 2000–2004 cohorts). For other variables considered (i.e., gender and undergraduate home school), the OFS sample was comparable with the overall SURE participant base.

The OFS addressed each of the project's three key themes: interest, preparedness, and pursuit, each described briefly below.

Interest. The OFS addresses interest as a subjective variable, asking respondents to recall their level of interest ("not interested at all," "slightly interested," "interested," and "very interested") in a variety of science-related graduate and professional careers at entry and completion of SURE (baseline/exit). Table 2 lists the 12 career options individually rated by each respondent. (See Supplemental Material 2, question 17 for exact wording.) We compared baseline and exit proportions of expressed interest for each of the options mentioned above, using a population proportion test to gauge statistical significance (the population proportion test is a normal approximation of a binomial distribution). Two types of interest are conceptualized: First, overall interest is calculated by comparing baseline and exit proportions of respondents who selected "interested" or "very interested" for the given option. Second, high-level interest is calculated by limiting the comparison to "very interested." To test for significant change in (recollected) interest level at baseline and exit, we used a Mantel–Haenszel chi-square test, which examines the significance of linear relationship between two ordinal variables (baseline and exit proportions of response categories). This test is nonparametric, that is, it does not assume normal distribution of the data, only of standard deviations. Using a one-tailed test, the criterion for statistical significance is a p value of <0.050.

Preparedness. Preparedness is assessed using a similar approach. Specifically, respondents were asked to recall how prepared they felt ("no ability," "weak ability," "basic ability," and "strong ability") at the beginning and conclusion of SURE with respect to a range of research design, methodology, and presentation skills, as well as familiarity with research ethics. Skills considered fall into three categories: research design skills, data presentation and publication skills, and graduate study skills. Baseline and exit preparedness comparisons were made using the same statistical measures and significance criteria as with interest (assessing proportions of overall preparedness and high-level preparedness for each skill).

Pursuit. Finally, the OFS assessed the pursuit of graduate training and professional careers in the sciences after completing college. In the absence of a non-SURE comparison group, outcomes identified among SURE graduates cannot be definitively attributed to program exposure. These outcomes are, however, contextualized with subjective assessment of the level of importance attributed to their SURE experiences for subsequent academic/professional development and career paths.

Study 2. Course Work Transcript Analysis

In contrast to the self-reports in study 1, which are subjective in nature (asking participants how interested and prepared they felt), study 2 provides a more objective measurement of interest demonstrated through course selection and academic performance. Specifically, we use academic records (transcripts) for Emory students to gauge associations between SURE participation and three outcome variables related to interest in science: 1) number of science courses taken, 2) number of advanced-level science courses taken (i.e., 300- or 400-level courses, or graduate-level courses), and 3) academic performance (GPA) in science course work.

("Science course work" refers to any courses in the natural and life sciences, as well as empirically oriented courses in psychology and anthropology.) Because the majority of SURE students participate in the program during the summer between junior and senior years, we would expect to see program impact during the following academic year. Therefore, these three variables are calculated for the senior year only. Our principal research hypothesis is that SURE participants will have higher levels of interest as measured in each of the three response variables—that is, they will have taken more science classes, have taken more advanced-level science classes, and have done better in science classes during their senior year—than students who did not participate in SURE. We emphasize that due to data availability, this analysis applies only to Emory students who participated in SURE.

Data Sources. Transcript data (courses taken, grades), as well as gender and ethnic minority status,[1] were provided by the Registrar's Office at Emory University. These data were available for a total of 21,364 students, including 475 SURE participants, who entered the Emory College in fall 1987 or later and who graduated in or before spring 2008. We limited our analysis to students who graduated within 4 years, because most students participate in the SURE program during the summer after the third year. Data on high school GPA and SAT math and verbal scores were obtained from the Emory Admissions Office. Admissions data were only available from the year 2000 onward, for a total of 5073 students, including 114 SURE participants.

For students who did not take any science classes during their senior or freshman year, science GPA for that year was coded as zero, rather than as missing. We believe the resulting adjusted science GPA (aGPA) captures course work choice (a real-time science interest indicator) and allows us to retain these students in our sample.

Analysis. To compare differences between SURE participants and nonparticipants with respect to number of science courses taken, number of advanced-level science courses taken, and science course work GPA, we used two different statistical measures: bivariate two-sample t tests and ordinary least square (OLS) multivariate regression. Each of these tests produces an odds ratio and a measure of association (p value). The OLS regression model accounts for possible confounding factors.[2] Comparing our measures of interest in science (science GPA, number of science classes, and advanced-level science classes taken during senior year), we would expect that students who chose to enroll in SURE would score higher than students who did not enroll in the program, simply because those students who did enroll were more interested in science to begin with. (That is why they decided to participate in the program in the first place.) Thus, any association we detect between participation in SURE and higher scores on our measures of interest in science might simply be due to students' prior interest in science, rather than to their participation in SURE (selection bias). One way to avoid selection bias is in the program design stage by assigning subjects randomly to the intervention group. This clearly was not an option in the case of SURE; students chose to participate in the program, or, in other words, self-selected.

Another way to deal with the selection bias is to include control variables in a multivariate regression model in the analysis stage. In particular, we need to control for (or, in other ways, "take out of the equation") students' prior interest in science, which would make them enroll in the program and have better senior-year science GPA and take more science classes their senior year. Therefore, we included the following three variables in the multiple regression model variables that we believe capture well the extent to which a student was interested in science before enrolling in SURE: number of science classes taken freshman year, freshman-year science GPA, and math SAT. (We restrict the first two indicators to the freshman year, because beyond this point, some students will have participated in SURE.) Additional controls include the student's gender, ethnic minority status, and high school GPA.

Thus, simply put, a multiple regression model compares students who had high existing interest in science (measured as their SAT math score and freshman science GPA) who participated in SURE and students with similar prior interest (similar math SAT score and freshman science GPA) who did not participate in the program. By including control variables into a multiple regression model, we are able to isolate the association between SURE and interest in science and ensure that this association is not driven by students' prior interest in science (thus limiting the possible effects of selection bias).[3]

Because data on high school GPA and SAT scores were available only for a subsample of the students, we estimated two sets of regression models: The first set of models included all 21,364 records from 1987 through 2008 and did not include Admissions' Office data. The other set of models did include admission data but only a subsample of records ($n = 5065$) for which the data were available was included.

Study 3. SURE II-III Survey (External)

The SURE survey is available to all summer research programs, and its scope is outlined in Lopatto (2007). The survey focuses on learning gains identified in Lopatto's

[1] Racial/ethnic minority was defined as Black, Hispanic, or American Indian/Alaskan Native.

[2] The number of science classes and advanced-level science classes are discrete counts; therefore, a count model would be more appropriate to assess the association between SURE participation and other independent variables with those two measures of interest in science. We have reestimated all the statistical models and found no substantive differences in the results. Because, with dependent variables with larger ranges the inefficiency of an OLS model compared with a count model is not substantial (Long, 1997) and OLS models are easier to interpret, we decided to report the results of OLS models instead of count models.

[3] In considering the possible association between SURE participation and interest in science during the senior year, we have considered the possibility that this association would be difficult to show given that most science majors presumably take a heavy senior-year science course load whether or not they have participated in SURE. We postulate, however, that even given this likelihood, there is nonetheless variation in the level of science courses (and advanced-level science courses) taken and that some of this variation can be explained by comparatively higher interest in science attributable to prior SURE participation. By this logic, any effect we can demonstrate would be conservative. Following this rationale, we have not included major as a control variable in our regression model.

prior research, which included queries to summer program directors on areas of training explicitly or implicitly pursued by their programs. Program directors receive a report that compares their participants' answers to those of students attending programs hosted at institutions with similar profiles (broadly defined, colleges vs. universities). Emory SURE students are invited to complete the survey on the last week of the program, and receive at least one additional email reminder during the fall semester. We have participated in this survey since 2003. The response rate for our program participants has ranged from 40 to 67%, with response rate decreasing as the program has grown larger (for 2003–2009, cohort size ranged from 51 to 84 participants).

Study 4. Annual Evaluation (Internal)

Participants are invited to complete an internal evaluation tied to student identification. This survey asks students to rate their satisfaction with various administrative program aspects (organization, adequate communication) and services (accommodations, dining facilities, social activities, parking, transportation options); perceived usefulness of various program activities and presentations; satisfaction with the assigned mentor and laboratory; asks respondents to give the program an overall grade; and asks respondents for ways in which the program may be improved. We also collect information about whether students have prior research experience and prior formal training in research ethics.

RESULTS

Study 1. Online Follow-Up Survey

Interest (Table 2). For overall interest, the highest baseline levels were reported for "scientific research" (72.4%), "an academic career" (51.6%), and "doing a Ph.D. in a scientific field" (48.4%), with only negligible interest in careers in dentistry (0.8%) and allied health (3.2%). For overall interest, baseline-to-exit changes are statistically significant (at $p < .050$) for half of the 12 options considered, namely, "doing a Ph.D. in a scientific field" ($p = .001$), "a job in science education/college" ($p < .001$), "a job in science writing/journalism" ($p < .001$), "an academic career" ($p < .001$), "a public health career" ($p < .001$), and "an allied health career" ($p = .001$). Among those options with significant change, the highest degree of change were for "an academic career" (14.0% increase), "a public health career" (11.2% increase), and "a job in science education/college" (10.0% increase). For high-level interest, baseline-to-exit changes were highly significant for all options (at $p < .001$) except "going to medical school," "a job in science education/K–12," "a career in dentistry," and "an allied health career." Those with the highest degree of significant change were "scientific research (generally)" (16.8% increase), "an academic career" (14.8% increase), and "doing a Ph.D. in a scientific field" (14.4% increase).

Preparedness (Table 3). At baseline, more than half of respondents indicated "basic" or "strong" overall ability in the following areas: "conceptualizing research objectives/hypotheses" (60.4%), "collecting data" (56.8%), and "understanding published research" (53.6%). The weakest skills at baseline were "writing grant applications" (13.2%), "using a statistical software package" (23.6%), and "identifying graduate funding" (26.4%). For overall ability (that is, baseline and exit proportions of respondents selecting "basic ability" or "strong ability"), substantive and highly significant increases were reported for all skills addressed (no p values >0.001). The three most notable increases were for the ability to prepare a poster presentation (45.2% increase), to discuss research at a graduate school interview (34.4% increase), and to select an appropriate data analysis strategy (34.0% increase). The smallest increase (at 16.4%) was for the ability to write a grant application. For high-level ability (that is, baseline and exit proportions of respondents selecting "strong ability"), increases were reported for all skills with statistical significance at $p < .001$. The largest increases were observed for the ability to prepare a poster presentation (45.2% increase), to apply principles of research ethics (38.4% increase), and to collect data (36.4%). The lowest increases (although all significant) were for the ability to write a grant application (4.8% increase), to use a statistical

Table 2. Baseline/exit comparisons of overall and high-level interest

Graduate study/career option	Overall interest				High-level interest			
	Baseline (%)	Exit (%)	Change	p value	Baseline (%)	Exit (%)	Change	p value
Scientific research (generally)	72.4	76.4	+4.0	0.088	40.0	56.8	+16.8	<0.001
Doing a Ph.D. in a scientific field	48.4	58.0	+9.6	0.001	26.4	40.8	+14.4	<0.001
M.D./Ph.D.	20.4	24.4	+4.0	0.070	5.6	10.8	+5.2	0.001
Going to medical school	39.6	37.6	−2.0	0.281	28.4	29.6	+1.2	0.360
A job in science education/K–12	7.2	10.4	+3.2	0.039	2.0	2.4	+0.4	0.384
A job in science education/college	29.6	39.6	+10.0	<0.001	8.0	16.8	+8.8	<0.001
A job in science writing/journalism	8.4	18.0	+9.6	<0.001	1.6	4.8	+3.2	0.001
An academic career	51.6	65.6	+14.0	<0.001	18.0	32.8	+14.8	<0.001
A career in industry	22.8	24.4	+1.6	0.296	3.2	7.2	+4.0	0.001
A public health career	21.6	32.8	+11.2	<0.001	5.2	13.6	+8.4	<0.001
A career in dentistry	0.8	0.8	+0.0	0.677	0.4	0.4	+0.0	0.736
An allied health career	3.2	7.6	+4.4	0.001	1.2	2.4	+1.2	.083

Table 3. Baseline/exit comparisons of overall and high-level ability

Skill	Overall ability				High-level ability			
	Baseline (%)	Exit (%)	Change	*p* value	Baseline (%)	Exit (%)	Change	*p* value
Research design								
Conceptualize objectives and hypotheses	60.4	83.2	+22.8	<0.001	10.0	43.6	+33.6	<0.001
Determine appropriate research methods	54.8	82.0	+27.2	<0.001	8.8	38.8	+30.0	<0.001
Select appropriate study design	44.8	67.6	+22.8	<0.001	6.8	25.2	+18.4	<0.001
Select data analysis strategy	38.0	72.0	+34.0	0.001	3.6	18.8	+15.2	<0.001
Apply principles of research ethics	50.0	80.4	+30.4	<0.001	9.6	48.0	+38.4	<0.001
Collect data	56.8	83.2	+26.4	<0.001	12.4	48.8	+36.4	<0.001
Use statistical software package	23.6	50.4	+26.8	<0.001	2.8	14.0	+11.2	<0.001
Data presentation and publication								
Prepare poster presentation	35.2	80.4	+45.2	<0.001	6.4	51.6	+45.2	<0.001
Understand published research	53.6	82.0	+28.4	<0.001	15.6	40.4	+24.8	<0.001
Graduate study								
Prepare graduate school application	49.2	70.0	+20.8	<0.001	8.8	28.4	+19.6	<0.001
Write grant application	13.2	29.6	+16.4	<0.001	0.4	5.2	+4.8	<0.001
Discuss research at grad school interview	40.0	74.4	+34.4	<0.001	6.0	39.6	+33.6	<0.001
Identify graduate funding	26.4	49.6	+23.2	<0.001	2.4	16.0	+13.6	<0.001
Identify mentors	49.6	74.4	+24.8	<0.001	11.6	40.8	+29.2	<0.001

software package (11.2% increase), and to identify graduate funding (13.6%).

Postundergraduate degree pursuit and completion (Table 4). We first report on the proportion of SURE alumni who, subsequent to completion of their undergraduate degree, have pursued graduate-level degree training. Nearly one-third (31.6%) completed a graduate degree, 86.5% of which were in a scientific field. Master's' degrees were completed by 15.4% of SURE graduates and 17.6% completed a doctoral degree. The most common graduate degrees completed were the M.D., M.A., and Ph.D., accounting for 30.1, 23.7, and 16.1% of all graduate degrees completed, respectively; graduate degrees were most commonly in a health-related field (41.9%) or in a life or natural science (26.9%). Of the 141 SURE graduates who had not yet completed a graduate degree at the time of the survey, 21 (14.9%) reported being currently enrolled in a graduate program. Accordingly, we estimate that as many as 40% (79 + 21/250) of SURE alumni pursue graduate training of some form. Of the 120 individuals who had neither completed a graduate degree nor were enrolled in a graduate program, eight (6.7%) said they intended to pursue another degree in the near future.

Employment (Table 5). At the time of the survey, approximately one-half of the entire OFS sample was employed on a full- or part-time basis (44.4 and 6.0%, respectively). Among employed respondents, the two most common work categories were research and development (44.6%) and professional services (35.6%); nearly a half of respondents (48.9%) were working in an academic setting (i.e., a college or university) and more than one-quarter (28.9%) in industry. Most (83.7%) were working in a science field; slightly less than one-half (44.6%) were in a research field. Three-quarters (77.7%) of employed respondents reported being either "satisfied" or "very satisfied" with their current positions. Among SURE graduates who completed their undergraduate degree in 2004 or earlier (an exclusion criterion imposed to allow at least 1 yr for the given outcome to occur), the most common pursuit during the first year after graduation was to attend graduate school (61.1%). When

Table 4. Degrees completed

Characteristic	Degree	*n*	%
Pursuit of graduate degrees	Graduate degree (any)	79	31.6
	Graduate degree in science	70	28.0
	Master's degree	39	15.6
	Doctoral degree	44	17.6
Specific graduate degrees completed[a]	M.A.	22	23.7
	M.S.	13	14.0
	M.B.A.	0	0.0
	M.F.A.	1	1.1
	M.P.H.	6	6.5
	R.N.	1	1.1
	Ph.D.	15	16.1
	Psy.D.	1	1.1
	M.D.	28	30.1
	D.M.D.	1	1.1
	D.P.T.	1	1.1
	J.D.	4	4.3
Graduate degree fields	Life & Natural Science	25	26.9
	Math & Computer Science	2	2.2
	Social Science	12	12.9
	Arts & Humanities	2	2.2
	Health	39	41.9
	Professional/Trade/Applied	12	12.9

[a] The *n* here corresponds to the 93 degrees obtained by the 79 individuals who completed at least one graduate degree.

Table 5. Employment status at the time of survey completion

Characteristic	Status	n	%
Employment status (current)	Currently employed, FT	111	44.4
	Currently employed, PT	15	6.0
	Unemployed/seeking	7	2.8
	Homemaker	0	0.0
	Not employed/not seeking	21	8.4
	Student/undergrad	15	6.0
	Student/grad	78	31.2
	Other	3	1.2
Subsample: currently employed (n = 126)			
Employment type	R&D	45	44.6
	Teaching	8	7.9
	Administration	5	5.0
	Professional services	36	35.6
	Other	7	6.9
Employment place	4-year college/university	44	48.9
	Elementary/secondary school	2	2.2
	U.S. federal government	10	11.1
	U.S. state government	1	1.1
	Nonprofit organization	2	2.2
	Industry/business	26	28.9
	Self-employed/owner	5	5.6
Current position in a science field?	Yes	103	83.7
	No	20	16.3
Current position is research-related?	Yes	45	44.6
	No	56	55.4
Satisfaction level w/current position	Very dissatisfied	19	15.1
	Somewhat dissatisfied	9	7.1
	Satisfied	41	32.5
	Very satisfied	57	45.2
Subsample: completed undergraduate degree before 2005 (n = 203)			
Pursuits during first year after completion of undergraduate degree	FT employment	57	28.1
	Graduate school	124	61.1
	Time off/travel	16	7.9
	Other	5	2.5

FT, full time; PT, part time; R&D, research and development.

inquired on level of importance attributed to SURE for subsequent academic development and career paths, nearly three-fifths (56.8%) considered SURE to be "very important" in this respect. More than nine of 10 respondents felt SURE to be at least "important" (data not shown).

Study 2. Course Work Transcript Analysis

Profile of Study Sample. Demographic and baseline characteristics of the study sample (n = 21,364) as well as data on science course work and performance are presented in Table 6. The overall sample (Emory graduates from the classes 1990–2004) includes 475 SURE participants and consists of slightly higher proportions of women than men (54.5 vs. 44.5%). In terms of ethnicity, the sample was primarily (89.0%) nonminority status. The average overall number of senior-year science courses was 0.78, and 0.49 for advanced level courses. GPA for science course work was 1.90 and 1.19 for freshman and senior years, respectively. (This figure, as noted above, is artificially driven down due to our having coded the GPA of students who took no science courses as 0.) For the subsample for whom admissions data were available (n = 5065), the average high school GPA was 3.52; average math and verbal SAT scores were 645 and 623, respectively. Overall undergraduate GPA was 3.21.

Associations between SURE Participation and Interest in Science (Table 7). Bivariate analysis shows that participating in SURE was associated with higher interest in science: Students who participated in the program took on average 1.16 more science classes and 1.11 more high-level science classes during their senior year than students who did not participate. Their senior-year adjusted science GPA was also higher by 1.65 points. Multivariate analysis, which controlled for students' prior interest in science, confirmed those results. The model which controlled for freshman adjusted science GPA, and number of science classes taken during freshman year, and the graduation GPA showed that SURE participants took on average 0.97 more science classes and 0.92 more high-level science classes than their peers who did not participate in the program, and their senior-year adjusted science GPA was higher by 1.38 points. The model which also included SAT math and verbal scores as well as high school GPA showed that SURE participants took on average 0.82 more science classes and 0.83 high-level

Table 6. Demographic characteristics and science course work/performance ($n = 21{,}364$)

Characteristic	Status	*n*	%
SURE participant?	Yes	475	2.2
	No	20,889	97.8
Gender	Female	11,648	54.5
	Male	9716	45.5
Ethnic minority	Yes	2356	11.0
	No	19,008	89.0
Subsample: admissions office data available ($n = 5065$)			
SURE participant?	Yes	114	2.2
	No	4951	97.8
High School GPA	Mean, min./max	3.52	0/4
SAT score, math	Mean, min./max	645	0/800
SAT score, verbal	Mean, min./max	623	0/800
Complete sample ($n = 21{,}364$)			
No. of science classes, senior year	Any level	0.78	
	Advanced	0.49	
GPA, all courses	Senior year	3.90	
GPA, science courses	Freshman year	1.90	
	Senior year	1.19	

science classes than their peers who did not participate in the program and their senior-year adjusted science GPA was higher by 1.15 points. All results were statistically significant at the 99% confidence level.

Results of our analysis of control variables, principally prior interest in science, are presented in Table 8. This analysis shows that senior-year interest in science was associated with the proxy indicators for prior interest in science: number of freshman-year science classes, freshman-year science GPA, and math SAT. (Neither gender nor racial/ethnic minority status were significant in our regression models.) Analysis of the subsample for which admissions data were available (bottom portion of Table 8) revealed that higher SAT verbal scores were associated with a lower number of science classes and advanced-level science classes taken but not with lower senior-year adjusted science GPA.

Studies 3 and 4. SURE II-III Survey (External) and Annual Evaluation (Internal)

Figure 1 reports aggregated responses for our 2007–2009 participants as provided by the Lopatto SURE survey team (Grinnell College). We have noted gains in recent years relative to previous years.

Response to our internal survey has ranged from 50 to 76.7% for years 2002–2009. For this period, >87% of survey respondents indicated the mentorship they received was superb and exceeded their expectations (52.5%) or met their expectations (35.4%). Eighty-four percent state the program met or exceeded their expectations, and 85% give the program an overall grade of "B+" or above. More than 84% of respondents indicate the program's clear communication of timelines is excellent or very useful and regard program organization similarly. Our most popular activity has been our "Lunches with Scientists," with >89% of respondents rating these activities as excellent or quite useful. The primary criticism of this activity is that demand is higher than supply, and some students are not able to attend as many lunches as they would like. Although less attended, the M.D./Ph.D. program information sessions were also highly rated (85% judged it excellent or very useful). Information presented in panel format was typically rated as being excellent or very useful by ~60% of respondents, and so was our presentation on funding graduate careers. Our Mentorship and Professional skills session was judged excellent or very useful by 43% of respondents and as somewhat useful by 42% of respondents.

More than 60% of respondents judged our ethics training to be very good or excellent, with 13% suggesting there is room for improvement. Approximately 67% of our 2002–2004 participants did not have ethics training before participating in SURE; for 2005–2009 participants, this percentage rises to 74%, probably reflecting the increase in younger first- and second-year participants.

DISCUSSION

SURE's Reach

During its first 15 yr, the SURE Program at Emory provided UR experience to >800 students. Among the roughly half of SURE participants not from Emory, more than two-thirds

Table 7. Associations between SURE participation and measures of interest in science[a]

Type of analysis	SURE participant performance (relative to nonparticipants)		
	No. senior-year science classes, β (SE)	No. senior-year advanced level science classes, β (SE)	Senior-year science aGPA, β (SE)
Bivariate association ($n = 21{,}364$)[b]	1.16 (0.02)	1.11 (0.04)	1.65 (0.07)
Multivariate (OLS) regression ($n = 21{,}364$)[c]	0.97 (0.05)	0.92 (0.04)	1.38 (0.07)
Multivariate (OLS) regression, admissions data subsample ($n = 5{,}063$)[d]	0.82 (0.10)	0.83 (0.08)	1.15 (0.13)

[a] All observed differences were significant at $p < 0.01$. Reported values are odds ratios, which indicate effect size.

[b] Measure of strength of correlation between variables.

[c] Predictive value; controlled for freshman aGPA, number of science courses in first year, and graduation GPA.

[d] Predictive value; controlled for SAT math and verbal scores and high school GPA.

Table 8. Control variables for prior interest in science for SURE participants vs. nonparticipants

	Participated in SURE?					
	Yes mean		No mean		Mean difference	Significance
Full sample (n = 21,364)						
Senior-year science classes[a]	1.58		0.47		1.16	<0.01
Senior-year advanced-level science classes[a]	1.92		0.76		1.11	<0.01
Senior-year science aGPA[a]	2.80		1.12		1.11	<0.01
Freshman-year science classes[a]	1.91		1.09		0.82	<0.01
Freshman-year science aGPA[a]	2.79		1.88		0.91	<0.01
Admissions data subsample (n = 5065)						
High school GPA[a]	3.67		3.51		0.16	0.020
SAT Math[a]	683		645		38	0.010
SAT Verbal[a]	663		622		41	0.005
	n	%	n	%	% Difference	Significance
Other control variables						
Gender[b]						
Female	245	51.6	11,403	53.4	1.8	0.193
Male	230	48.4	9486	46.6	1.8	
Ethnic minority?[b]						
Yes	58	12.2	2,298	10.8	1.4	0.690
No	417	87.8	18,591	89.2	1.4	

[a] Two-sample t test (continuous variables; compares means of participants vs. nonparticipants).

[b] Mantel–Haenszel chi-square test (categorical variables; linear relationship baseline and exit proportions of response categories).

were from schools outside of Georgia and 28.5% were from small liberal arts colleges, underscoring that SURE's reach is national and not limited to students from large research universities. SURE has also been successful in enrolling growing numbers of minority students, with increases over the years from approximately one-fifth to one-third of each summer cohort. Although the majority of SURE participants have been science majors, a steady presence of nonmajors (~1/10) suggests that SURE has also contributed to HHMI's goal to increase science literacy among the general undergraduate population. The average self-reported undergraduate GPA of SURE graduates (3.63, corresponding roughly to "B+") also suggests that SURE is successfully involving not only the highest academic achievers but also students with room for improvement in grade performance. Importantly, we note that the majority of participants come to SURE with no prior experience in a mentor-supervised research setting, a gap which SURE fills.

Study 1: Online Follow-Up Survey

Methodological Reflections on Alumni Surveys. Existing evaluation studies have addressed the benefits and shortcomings of alumni surveys (Hakim, 1998; Bauer and Bennett, 2003). On the one hand, such surveys may be constrained by recall limitations (i.e., memory issues) or a greater interest in responding among those participants who had a positive experience. On the other hand, alumni may have less temptation to speak well of their UR program post facto, because they are more distant from ongoing funding concerns on the part of program organizers. Alumni may also possess a broader critical perspective or insight into the positive or negative effects of their UR experience that could only come from the passage of time since completion of that experience. Finally, our confidence in the alumni survey approach is increased in light of previous studies that have shown correlation between a person's assessments of their achievements gauged during their undergraduate years and as alumni, and between their own assessments and those of faculty mentors (Pace, 1979; Seymour *et al.*, 2004).

SURE's Influence on Interest in the Sciences. Even in the absence of a comparison group, the range and degree of reported increases in both general and high-level interest suggests that SURE is effectively stimulating interest in graduate study and professional careers in the sciences. Especially noteworthy are increases in general and high-level interest in an academic career and high-level interest in scientific research generally. (We attribute the lack of significant change in general interest in scientific research to the fact that baseline interest level, at 72.4%, was already quite high, leaving less room for improvement.) We are encouraged at the lack of significant change in general interest levels for medical school, a career in industry, and dentistry, as these career trajectories fall outside the purview of SURE's programmatic foci as well as the aims of HHMI and NSF. It is also worth underscoring SURE's effect on general interest in college-level science education and, to a lesser degree, K–12 science education, suggesting the program's contribution to training the next generation of students in the sciences.

Calculations of change in high-level interest (meaning, differences between baseline and exit proportions of students who were "very interested" in the given option) are useful to gauge SURE's capacity to help students identify those professional trajectories that hold particularly intense

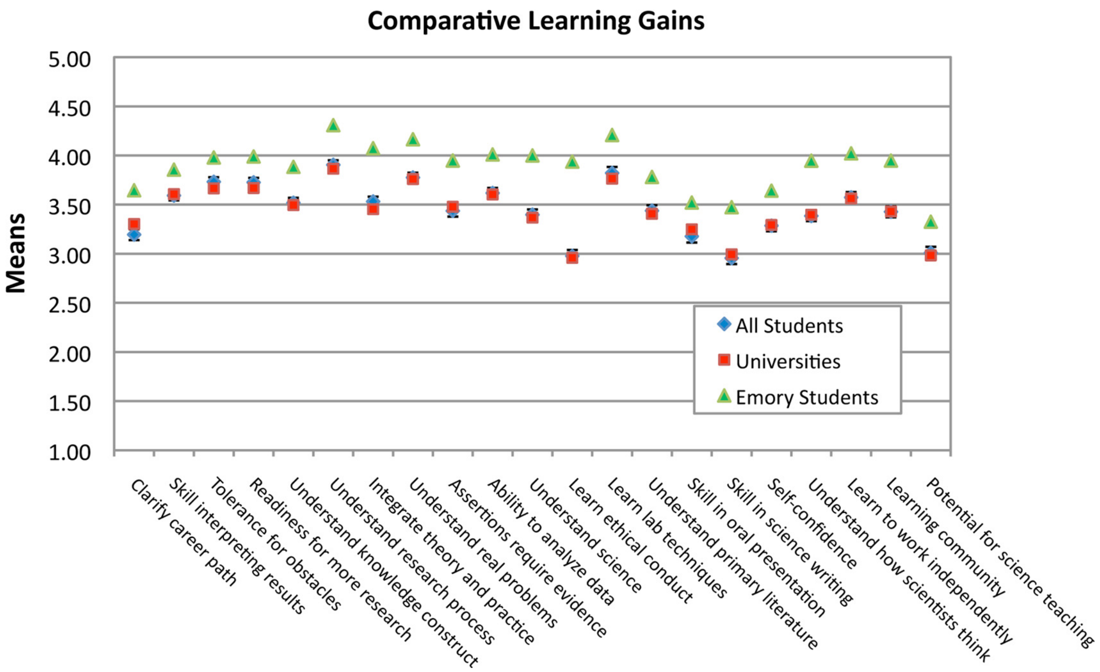

Figure 1. Comparative means on the 21 learning gain items in the SURE II-III survey (Lopatto, Grinnell College). The mean learning gains from Emory SURE data ($n \leq 97$) are depicted as green triangles. Blue diamonds represent 1665 responses to the SURE survey from 2007 to 2009; responses from programs conducted at universities are depicted by red squares ($n \leq 628$). The vertical lines in the "All Student" means depict ±2 SEs. Emory SURE participant response rates are as follows: 2007 (55% response rate; 60 program participants), 2008 (48% response rate, 68 participants), and 2009 (40% response rate, 84 participants).

appeal. Beyond scientific research and an academic career, we note the relatively high degree of significant change in high-level interest in carrying out a Ph.D. in a scientific field (a 150% increase from 26.4 to 40.8%). It is extremely difficult to ascertain whether a student will pursue a medical degree or a Ph.D. based on an application. A savvy student can target his or her essay to mask a pre-med focus. Applicant interviews are not feasible in most cases, and even when possible, a student (or letter writer) may feel it is not unethical to secure a research fellowship that will impress a medical school admissions committee. It can also be argued that basic research skills are an asset to any physician and as such, our program doesn't fail science at large. Recent publications (McGee and Keller, 2007; Villarejo *et al.*, 2008) have begun to tease out possible pre-M.D. versus pre-Ph.D. indicators, and we will consider adding some kind of rubric addressing these metrics to the recommendation letter that is currently part of the application.

SURE's Influence on Preparedness in the Sciences. Significant increases in all response options, for perceived overall and high-level ability, suggest that SURE is effectively improving participants' abilities in key scientific skill areas and hence their preparedness for graduate study and professional careers in the sciences. By the same token, the degree of increase varied substantially, suggesting greater effectiveness in some training areas over others. We note, for example, that the skills for which respondents reported the lowest general ability at baseline—"writing a grant application" (13.2%), using a statistical software application (23.6%), and "identifying graduate funding" (26.4%)—were also those for which the smallest positive change was reported from baseline to exit (29.6, 26.8, and 23.2%, respectively). This finding suggests important areas for programmatic improvement. It is also important to remember that some skills sets (grant application skills, for example) require years to fully develop and, therefore, that SURE may impel its students in a positive pedagogic direction even in the absence of a sense that their "ability" has improved.

In terms of overall ability, the strongest increase was reported for the preparation of a poster presentation (a 228% increase from 35.2 to 80.4%). We are also encouraged by appreciable increases for the selection of an appropriate data analysis strategy (+34.0), discussion of research at a graduate school interview (+34.4), and applying research ethics principles (+30.4%). The latter increase (research ethics) strikes us as particularly significant as this area is all too often left out of undergraduate science education.

SURE's Influence on the Pursuit of Graduate Study and Careers in the Sciences. Notable in this set of results is the finding that nearly three of 10 (28.0%) of SURE graduates go on to complete a graduate degree in a scientific field. This rate substantially exceeds rates reported in a recent alumni survey of former science majors at Emory (unpublished data). Whereas Ph.D.s and M.D.s were completed by 5.4 and 11.8% of alumni surveyed, respectively, SURE graduates completed these respective degrees at the rates of 16.1 and 30.0%. Moreover, the proportion of our sample that completed M.P.H. degrees was nearly double that of the science alumni survey (6.5 vs. 3.0%). That the M.D. was the most frequently pursued degree raises questions about what sorts of careers SURE graduates ultimately aim to pursue (i.e., research vs. clinical), and our finding that research and development was the most commonly reported employment category (at 44.6%) assuages concern that SURE might be a research-oriented detour leading to a clinical medical career.

Among respondents who were employed at the time of the survey, we found academic settings to be most common (at 44.6%), with industry (28.9%) and federal government settings (11.1%) also prevalent. Results also indicate high levels of employment in science fields (83.7%) and job satisfaction (with more than three-quarters reporting "satisfied" or "very satisfied"). Finally, the majority (>90%) consider SURE to have been important or very important to their academic development and career paths.

Study 2: Course Work Transcripts

The major finding from the regression analyses (limited to Emory students) is the statistically significant association between SURE participation and three outcome measures of interest in the sciences, even after controlling for prior interest in science, gender, and minority status. Specifically, we found that during their senior year former SURE participants take more science classes (both introductory and advanced level) and receive higher grades in these courses than their nonparticipating peers.

It is worth underscoring the ability of a regression design to address a common shortcoming of comparative outcome studies between program participants and nonparticipants, namely the inability to address selection bias. The particular bias we were concerned with was the likelihood that SURE participants are an undergraduate subset with a high pre-existing interest in science, which we would therefore expect to take more science courses and to do better in those courses irrespective of SURE participation. The data indeed showed that students who enrolled in SURE had higher freshman-year adjusted science GPA and SAT math scores and took more science classes their freshman year than students who did not participate. We addressed this issue by including indicators of students' interest in science before their enrollment in SURE (freshman adjusted science GPA, number of science courses taken during freshman year, and SAT math scores) in the regression models. Our analysis showed that SURE participants had higher interest in science after completing the program compared with their peers with similar pre-existing interest in science, but who did not participate in SURE. Thus, the breakthrough in this study was our ability to control for these variables and still detect significant associations with the three given outcomes. This methodological innovation lends powerful new evidence that participation in SURE is stimulating interest and enhancing academic performance in senior-year science course work.

Our online survey and transcript analyses also point toward important areas for future research. Absent in our own research, and seldom addressed in UR scholarship generally, is systematic attention to the role of faculty mentors (Hathaway *et al.*, 2002; Zydney *et al.*, 2002; Gafney, 2001; Frantz *et al.*, 2006). All too often, mentors are presumed to be a homogeneous group, each of whom devotes equal time and energy to their student researcher. Moreover, some faculty develop collaborative projects with their student researchers, whereas others leave the student to undertake their project more or less independently. As these variations have obvious implications for the students' qualitative experience of UR programs, they should be integrated into formal evaluation studies. A related area concerns the distinction between UR programs at public versus private institutions (see Merkel, 2003) and at large research universities versus small liberal arts colleges. Also called for in future evaluation studies is greater attention to those attributes that distinguish the summer intensive model from those UR programs taking place during the school year. What, in other words, might be the value added by summer research as a particular form of UR experience? Finally, there is a paucity of evaluation research on ethics training components of UR programs (but see Sweeting 1999; Shachter, 2003).

Studies 3 and 4. SURE II-III Survey (External) and Annual Evaluation (Internal)

Combined with our internal evaluation, SURE II-III (Lopatto) survey data offer a powerful, annual opportunity for reflection on our program offerings and how we may adapt activities to ensure explicit discussion of all topics we consider relevant to a UR experience for future scientists. This explicit articulation of objectives is important for all students (awareness of developing expertise inspires confidence) but possibly more so for students who study at institutions that are not as research-intensive as ours and who may not have a family history of participation in science careers. We also believe that this increased awareness helps students better reflect on their experience, and in turn allows them to offer more robust program feedback. It is also important to review activity feedback with an eye to student development and experience: the benefit of an activity (e.g., discussion on the importance of mentorship or how authorship is determined) may not be immediately apparent to a student in the short-term but may become more so upon experiencing new learning and work environments.

CONCLUSIONS

Despite limitations in the available data, this study strongly indicates that participation in the SURE program enhances research skills and interest in academic research careers. More importantly, the data indicate that participation in our summer research program significantly increases the number of science courses taken and GPA, postparticipation, in particular among women and minorities. The results suggest that planning SURE-like experiences for first- and second-year students might improve success in science and interest in academic research careers. We plan to conduct a study of first- and second-year SURE participants to test this hypothesis since in recent years we have added such students to our program. First- and second-year students may require different kinds of preparation, perhaps a research apprenticeship during the academic year or a research course. Future studies should also include attention to the quality of mentoring and to preparing current and future faculty to mentor undergraduates in research. We also hope to study different outcomes for students who participate in research for credit versus those who complete an honors thesis.

ACKNOWLEDGMENTS

We are grateful to the Emory Office of the Registrar and Office of Admissions for providing data used in study 2. We acknowledge the generous help of Tan Tran (Emory's Office of Institutional Planning and Research) with data management. We also thank David Lopatto and Leslie Jaworski (Grinnell College) for allowing us to participate in their survey and sharing their data each year. This research was supported through funding from the Howard Hughes Medical Institute, the National Science Foundation, Emory University's Division of Biological and Behavioral Sciences, and SURE mentors.

REFERENCES

Adhikari, A., and Nolan, D. (2002). "But what good came of it at last?": how to assess the value of undergraduate research. Notices AMS *49*, 1252–1257.

Alexander, B. B., Foertsch, J. A., Daffinrud, S., and Tapia, R. (1998). The Spend a Summer with a Scientist (SaS) Program at Rice University: A Study of Program Outcomes and Essential Elements, 1991–1997. Madison, WI: Wisconsin Center for Education Research.

Barlow, A.E.L., and Villarejo, M. (2004). Making a difference for minorities: Evaluation of an educational enrichment program. J. Res. Sci. Teach. *41*, 861–881.

Bauer, K. W., and Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. J. High. Educ. *74*, 210–230.

Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing Undergraduate Education: A Blueprint for America's Research Universities. Carnegie Foundation for the Advancement of Teaching. Boyer Commission on Educating Undergraduates in the Research University, State University of New York, Stony Brook, NY. www.sunysb.edu/boyerreport (accessed 19 February 2010).

Burnley, P. C., Evans, W., and Jarrett, O. S. (2002). A comparison of approaches and instruments for evaluating a geological sciences research experiences program. J. Geosci. Educ. *50*, 15–24.

Butler, P. J., Dong, C., Snyder, A. J., Jones, A. D., and Sheets, E. D. (2008). Bioengineering and bioinformatics summer institutes: meeting modern challenges in undergraduate summer research. CBE Life Sci. Educ. *7*, 45–53.

Campbell, A., and Skoog, G. (2004). Preparing undergraduate women for science careers. J. Coll. Sci. Teach. *33*, 24–26.

Carsrud, A. L. (1980). Undergraduate Research in Psychology: Its Past and Its Future. Texas. Paper presented at American Psychological Association Annual Meeting, Montreal, Canada, 1980.

Coleman, J. (2005). Undergraduate research participation as an essential component of a research university. Counc. Undergrad. Res. Q. *25*, 154–155.

Crowe, M. (2006). Creative scholarship through undergraduate research. Peer Rev. *8*, 16–18.

Crowe, M., and Brakke, D. (2008). Assessing the impact of undergraduate research experiences on students: an overview of current literature. Counc. Undergrad. Res. *28*, 43–50.

Delagarza, J. M., Anderson, S. K., and Lee, J.A.N. (1991). Undergraduate research experience through summer internships. Eng. Educ. *81*, 384–385.

Dirks, C., and Cunningham, M. (2006). Enhancing diversity in science: is teaching science process skills the answer? CBE Life Sci. Educ. *5*, 218–226.

Elgren, T., and Hensel, N. (2006). Undergraduate research experiences: synergies between scholarship and teaching. Peer Rev. *8*, 4–7.

Fields, E. F. (1971). Analysis and Evaluation of Undergraduate Research Training Projects. Final Report. Washington, DC: National Center for Educational Research and Development.

Foertsch, J. A., Alexander, B. B., and Penberthy, D. L. (1997). Evaluation of the UW/Madison's summer undergraduate research programs: final report. Madison, WI: University of Wisconsin–Madison.

Frantz, K. J., DeHaan, R. L., Demetrikopoulos, M. K., and Carruth, L. L. (2006). Routes to research for novice undergraduate neuroscientists. CBE Life Sci. Educ. *5*, 175–187.

Gafney, L. (2001). The impact of research on undergraduates' understanding of science. Counc. Undergrad. Res. Q. *21*, 172–176.

Gregerman, S. R. (1999). Improving academic success of diverse students through undergraduate research. Counc. Undergrad. Res. Q. *20*, 54–59.

Gum, A., Mueller, K., Flink, D., Siraj, S., Batsche, C., Boothryod, R., and Stiles, P. (2007). Evaluation of a summer research institute in behavioral health for undergraduate students. J. Behav. Health Serv. Res. *34*, *2*, 206–218.

Hakim, T. (1998). Soft assessment of undergraduate research: Reactions and student perspectives. Counc. Undergrad. Res. Q. *18*, 189–192.

Hathaway, R., Nagda, B., Gregerman, S. (2002). The relationship of undergraduate research participation to graduate and professional educational pursuit: an empirical study. J. Coll. Stud. Dev. *43*, 614–631.

Ishiyama, J. (2001). Undergraduate research and the success of first generation, low income college students. Counc. Undergrad. Res. Q. *22*, 36–41.

Jonides, J. (1995). Evaluation and Dissemination of an Undergraduate Program To Improve Retention of At-Risk Students. Michigan, Ann Arbor, MI: College of Literature, Science, and the Arts, University of Michigan.

Kardash, C. M. (2000). Evaluation of an undergraduate research experience: perceptions of undergraduate interns and their faculty mentors. J. Educ. Psychol. *92*, 191–201.

Knox, D. L., DePasquale, P. J., and Pulimood, S. M. (2006). A model for summer undergraduate research experiences in emerging technologies. Houston, TX: Special Interest Group on Computer Science Education.

Kremer, J. F., and Bringle, R. G. (1990). The effects of an intensive research experience on the careers of talented undergraduates. J. Res. Dev. Educ. *24*, 1–5.

Long, J. S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage Publications.

Lopatto, D. (2004a). What undergraduate research can tell us about research on learning. Paper presented at the Council on Undergraduate Research Annual Meeting, LaCrosse, WI. Vol. 4, What Works, What Matters, and What Lasts, Washington, DC: National Center for Educational Research and Development.

Lopatto, D. (2004b). Survey of undergraduate research experiences (SURE): First findings. Cell Biol. Educ. *3*, 270–277.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. CBE Life Sci. Educ. *6*, 297–306.

Mabrouk, P. A., and Peters, K. (2000). Student perspectives on undergraduate research experiences in chemistry and biology. Counc. Undergrad. Res. Q. *21*, 25–33.

Malanga, C. J. (1988). The annual pharmacy undergraduate research seminar at West Virginia University School of Pharmacy. Am. J. Pharm. Educ. *52*, 393–396.

Manduca, C. (1997). Broadly defined goals for undergraduate research projects: a basis for program evaluation. Counc. Undergrad. Res. Q. *18*, 64–69.

Matsui, J., Liu, R., and Kane, C. M. (2003). Evaluating a science diversity program at UC Berkeley: more questions than answers. Cell Biol. Educ. *2*, 117–121.

Mauger, K. (1990). An evaluation of the impact of an undergraduate research seminar. Am. J. Pharm. Educ. *54*, 43–45.

McGee, R., and Keller, J. L. (2007). Identifying future scientists: predicting persistence into research training. CBE Life Sci. Educ. *6*, 316–331.

Merkel, C. A. (2001). Undergraduate Research at Six Research Universities, Pasadena, CA: California Institute of Technology.

Merkel, C. A. (2003). Undergraduate research at the research universities. New Directions Teach. Learn. *93*, 39–53.

Mervis, J. (2001a). Liberal arts schools pass science checkup. Science *293*, 193.

Mervis, J. (2001b). Student research: what is it good for? Science *293*, 1614–1615.

Millspaugh, J. J., and Millenbah, K. F. (2004). Value and structure of research experiences for undergraduate wildlife students. Wildl. Soc. Bull. *32*, 1185–1194.

Nagda, B. A., Gregerman, S., Jonides, J., von Hippel, W., and Lerner, J. (1998). Undergraduate student-faculty research partnerships affect student retention. Rev. High. Educ. *22*, 55–72.

Okorodudu, C. (1970). The Development and Evaluation of Undergraduate and Graduate Research Curricula; Section III, Part IIa: Instrument Development. Paper presented at American Educational Research Association Annual Meeting, Minneapolis, MN, 1970.

Pace, C. R. (1979). Measuring Outcomes of College: Fifty Years of Findings and Recommendations for the Future, San Francisco, CA: Jossey-Bass.

Page, M. C., Abramson, C. I., and Jacobs-Lawson, J. M. (2004). The National Science Foundation Research Experiences for Undergraduates Program: experiences and recommendations. Teach. Psychol. *31*, 241–247.

Palladino, J. J. (1982). Undergraduate research in psychology: assessment and directions. Teach. Psychol. *9*, 72–74.

Peppas, N. A. (1981). Student preparation for graduate school through undergraduate research. Chem. Eng. Educ. *15*, 135–137.

Prentice-Dunn, S., and Roberts, M. C. (1985). A summer internship in psychological research: preparation of minority undergraduates for graduate study. Teach. Psychol. *12*, 142–145.

Russell, S. H., Hancock, M. P., and McCullough, J. (2006). Evaluation of NSF support for undergraduate research opportunities: synthesis report, Arlington, VA: National Science Foundation.

Russell, S. H., Hancock, M. P., and McCullough, J. (2007). The pipeline: benefits of undergraduate research experiences. Science *316*, 548–549.

Schowen, B. (1998). Research as a critical component of the undergraduate educational experience. In: Assessing the Value of Research in the Chemical Sciences (workshop report). Washington, DC: National Academies Press.

Schowen, B. (2002). Value and impact of undergraduate research in chemistry: lessons gained from 45 years of experience at the University of Kansas. ConfChem Conference on Undergraduate Research, Lawrence, KS, April 17–18, 2002.

Seymour, E., Hunter, A.-B., Laursen, S. L., and DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. Sci. Educ. *88*, 493–534.

Shachter, A. M. (2003). Integrating ethics in science into a summer undergraduate research program. J. Chem. Educ. *80*, 507–512.

Spilich, G. (1997). Does undergraduate research pay off? Counc. Undergrad. Res. Q. *18*, 57–59.

Summers M., and Hrabowski, F. (2006). Preparing minority scientists and engineers. Science *311*, 1870–1871.

Sweeting, L. M. (1999). Ethics in science for undergraduate students. J. Chem. Educ. *76*, 369–372.

Villarejo, M., Barlow, A. E. L., Kogan, D., Veazey B. D., and Sweeney, J. K. (2008). Encouraging minority undergraduates to choose science careers: career paths survey results. CBE Life Sci. Educ. *7*, 394–409.

Weaver, G. C., Russell, C. B., and Wink, D. J. (2008). 2008. Inquiry-based and research-based laboratory pedagogies in undergraduate science. Nat. Chem. Biol. *4*, 577–580.

Ward, C., Bennett, J. S., and Bauer, K. W. (2002) Content Analysis of Undergraduate Research Student Evaluations. www.udel.edu/RAIRE/Content.pdf (accessed 5 August 2009).

Winthrop University Undergraduate Research Advisory Committee (2006). Undergraduate Research as a Deep Learning Experience, Rock Hill, SC: Winthrop University.

Zydney, A. L., Bennett, J. S., Shahid, A., and Bauer, K. W. (2002). Impact of undergraduate research in engineering. J. Eng. Educ. *91*, 151–157.