Undergraduate research programs build skills for diverse students

C. Brooke Bruthers,1 Emma L. Hedman,1 and Marsha Lakes Matyas2
1American Physiological Society, Rockville, Maryland and 2Evaluation for Excellence, Olney, Maryland

Abstract

While many professional societies, colleges, and universities offer undergraduate summer research experience (URE) programs for students, few have systematically evaluated their programs for impacts on the fellows. The American Physiological Society (APS) developed and administered multiple UREs with varying target groups: students with and without prior research experiences and students from disadvantaged groups, including underrepresented racial/ethnic minorities (URM), persons with disabilities, first generation college students, and persons with financial or social disadvantages. Each program had specific goals and measurable objectives. To assess the impact of these programs, APS both documented student completion of program tasks (e.g., designing experiments, analyzing data, writing abstracts) and developed reliable and valid survey instruments to quantify students’ self-ratings on a variety of research and career planning skills related to the program objectives. Results indicate that fellows as a whole and for most individual programs gained skills and knowledge in numerous areas: experimental design, data management, lab safety, statistical analysis, data presentation, scientific writing, scientific presentation, professional networking, professional networking at scientific meetings, authorship attribution, animal use in research, human subjects in research, roles of lab mates and mentors, and research career training and planning. Furthermore, there were few differences within the diversity comparison groups (women vs. men, URM fellows vs. non-URM fellows, etc.). Suggestions for improvement of URE programs are proposed.

disability; minorities; research careers; research skills; undergraduate research

INTRODUCTION

For more than a decade, undergraduate research experiences (UREs) have been hailed as an important tool in increasing undergraduate student interest in pursuing advanced science, technology, engineering, and mathematics (STEM) degrees and selecting research careers. While most data are not from randomized, controlled studies, there is a growing body of evidence to indicate that UREs positively influence STEM retention and graduation rates as well as student confidence in their STEM content knowledge and their understanding of experimental design and data analysis (1–3). UREs also support the development of the students’ science “identity,” that is, their feeling of belonging in the world of research (4, 5). These impacts apply to diverse students. The 2017 National Academy of Sciences (NAS) report concluded that, for students from underrepresented (UR) groups, UREs increase persistence in STEM and validate their science identities but we do not have a clear understanding of how UREs accomplish these outcomes (6). The analysis of Krim et al. (7) of 70+ recent URE studies found that only 1% focused on UR students and only 18% identified UR student outcomes in the evaluation data. The NAS report called for studies to explore “…how UREs work for different students, why they work, and how to evaluate the reported outcomes for URE fellows” (6) (p. 4). They specifically call for URE programs to assure that they are not only accessible but also welcoming to those who are, for example, students with disabilities and first-generation college students.

What constitutes a URE? “[T]he student is doing the type of work that STEM researchers would typically do; that is, the student is engaging in discovery and innovation, iteration, and collaboration as the student learns STEM disciplinary knowledge and practices while working on a topic that has relevance beyond the course. A URE is structured and guided by a mentor; the students are intellectually engaged and assume increasing ownership of some aspects of the project over time” (6) (p. 22). Increasingly, colleges and universities are looking for practical models for class-based undergraduate research experiences (CURES). However, our understanding of URE impacts has been provided primarily by evaluation of supplementary programs, including summer research fellowship programs sponsored by colleges/universities, federal agencies, professional societies, foundations, or other support (7, 8). Evaluation of programs targeted at underrepresented racial/ethnic minority (URM) students have provided evidence of URE programs’ broader impacts (6).

While not recognized in recent studies (7–9), many professional societies have URE programs to encourage students to enter their disciplines. A longitudinal study of 29 life science societies found that more than half (52%) of the societies...
surveyed offered undergraduate research fellowships in 2014, and nearly a third (32%) had fellowships specifically for underrepresented students (10). Furthermore, the percentage of life science societies offering URE fellowships rose significantly from 2008 to 2014 as did the percentage offering travel awards for undergraduate students to attend their scientific meetings. Interviews with past URE fellows who were from disadvantaged groups suggest that the combination of a summer research experience and subsequent attendance at a large scientific conference to present one’s work is a powerful contributor to one’s science identity (11). However, few undergraduate programs offered by professional societies have specific goals and measurable objectives and, in terms of evaluation, nearly two-thirds (62%) use only an exit survey to get general feedback from fellows (10). Therefore, their program impacts are neither documented nor shared. The American Physiological Society (APS) was one of the few societies in the study that routinely establishes goals and measurable objectives for its fellowship programs, tracks fellows’ completion of fellowship activities, and regularly uses entry/exit measures of fellow self-reports of behavioral and perceptual changes resulting from the fellowship.

The current study reports on the summative evaluation data from 2016 to 2018 for four of APS’ undergraduate summer research programs:

- **UGSRF**: APS Undergraduate Summer Research Fellowship (UGSRF) for students with little or no previous research experience (9 mo or less) to conduct research in any area of physiology. Funded by APS, this program offered 24 fellowships per year during the study period.

- **UGREF**: APS Undergraduate Research Excellence Fellowship (UGREF) for students with previous research experience (more than 9 mo) to work in any area of physiology. APS provided funding for up to six students per year during the study period.

- **STRIDE**: Short-Term Research Education Program to Increase Diversity in Health-Related Research Fellowships (STRIDE) for underrepresented students with or without previous research experience. Funding was provided by the National Heart, Lung, and Blood Institute (NHLBI) for up to 25 students annually to work in NHLBI-related research areas.

- **IOSP**: Integrative Organismal Systems Physiology (IOSP) Fellowships for underrepresented students with or without previous research experience. Funding was provided by the National Science Foundation Integrative Organismal Systems Biology Division for up to eight students annually to work in IOSP-related research areas.

STRIDE and IOSP were specifically focused on supporting students from groups underrepresented in STEM: URM students, students with disabilities, and students from disadvantaged backgrounds (e.g., first generation college, financially disadvantaged, or socially disadvantaged students). Disadvantaged students were also eligible for and encouraged to apply to the UGSRF and UGREF programs.

The four programs overlapped substantially in their goals and measurable objectives. In all four programs, fellows were to develop their scientific research and communication skills: develop and revise a working hypothesis; conduct experiments; analyze and report data; write research summaries in report and abstract forms; prepare a scientific poster; present data and results orally; and network effectively with other fellows and APS members.

All four programs included a career development focus. For UGSRF and UGREF, the objectives were to increase fellows’ commitment to a research career, including plans to enroll in a graduate or graduate/professional program and pursue a career in basic research. Because the STRIDE and IOSP programs were funded for a limited number of years (thus precluding federally funded long-term tracking of fellows), their career development focused more on exploration, with fellows seeking to investigate career options and needed skills and reflect on their strengths for and interest in specific careers. Fellows in the UGREF program and those receiving a second year of STRIDE funding were expected to write a draft manuscript of their research and create an Individual Development Plan (12) to outline their career goals and plans to achieve them.

The current study focuses on evaluation of program impacts on undergraduate students who received fellowships in the UGSRF, UGREF, IOSP, and/or STRIDE programs. Specifically, we sought to determine whether participation in the APS summer research programs:

1. Increased fellows’ scientific research and communication skills as measured by structured self-ratings.
2. Increased fellows’ understanding of and interest in research careers as measured by self-reports; and
3. Was equally effective for diverse students (women, URM students, persons with disabilities, and persons from disadvantaged backgrounds).

## METHODS

### Fellows

The large majority of those who received APS summer research fellowships between 2016 and 2018 completed both their program assignments and evaluation surveys. Overall, complete data were received from 146 summer research fellows. As shown in Fig. 1A, female students comprised more than two-thirds of the fellows (68%). Most fellows self-identified their race as Caucasian (62%), Asian (20%), or Black/African American (10%) (Fig. 1B) and 16% self-identified their ethnicity as Hispanic or Latino (Fig. 1C). Each of the APS programs included fellows from low-income families and disadvantaged backgrounds (Fig. 1D). At least 4% (n = 6) of the fellows were persons with disabilities. The actual percentage may be higher; some of the past STRIDE fellows participating in an interview study disclosed that they had physical disabilities that were not reported on their application forms or surveys (11).

Fellows also reported on their previous research experiences. This varied considerably by program, partially because some programs had specific application requirements concerning previous research experience. For example, UGSRF applicants could have no more than 9 mo of previous research experience in a physiology laboratory while UGREF applicants had to have at least 9 mo of previous research experience. These requirements are reflected in fellows’ self-reports of their years of previous research experience (Fig. 2).

---

Notes:
- DOI: 10.1152/advan.00165.2020
- http://advan.physiology.org
Program Activities

All APS summer research fellows were asked to complete program evaluation surveys at the start and end of their fellowship. Fellowships applications were reviewed early in the calendar year, acceptances were sent in early spring, and the fellowship activities began in late spring. Fellows conducted research for a minimum of 10 wk in the laboratory of an APS member (research host). They received a $4,000 stipend for their work and their research host received either a $500 reimbursement of expenses related to the fellow’s project (IOSP and STRIDE) or $300 in unrestricted funds (UGSRF and UGREF). IOSP fellows also were eligible for housing support during their research experience. Each fellow completed nine online professional development activities on topics related to the fellowship’s measurable objectives: networking, research design, hypothesis development, structured career comparisons, ethical issues in research and publication, how to ask for letters of recommendation, and development of both personal career plans and personal statements for graduate school applications. In the fall after their summer research experience, fellows and their research hosts were encouraged to submit abstracts to the APS annual meeting, Experimental Biology (EB). Fellows received financial support to attend EB the following spring, and nearly all fellows presented a poster at the meeting during the regular poster session and/or at a special APS undergraduate poster session. The latter included a presession exhibit by graduate school representatives so fellows could learn about graduate programs in physiology. Some fellows competed for undergraduate abstract and presentation awards. Fellows were encouraged to attend an undergraduate orientation session for all undergraduates who submitted first author abstracts to APS and a diversity networking/career development event. Fellows received an invitation to complete the postfellowship survey at the end of EB.

Documentation Evidence

Primary evidence of accomplishment of the measurable objectives outlined above was through documentation of fellows’ work, including hypothesis development, experimental results, writing research summaries (abstracts, blogs, posters), affirmation of completion of responsible conduct of research (RCR) training, and presenting at EB. Additional
documentation was gathered through fellows’ completion of the online professional development activities. These were tracked by staff and research hosts; all fellows completed these major projects to receive their program completion certificate and letter.

Survey Instruments

APS developed survey instruments to assess student perceptions of what they gained through the fellowships specific to the measurable objectives. Initial surveys used a five-point Likert-type rating scale (“Strongly Agree” to “Strongly Disagree”). However, for the high-achieving undergraduate students in these programs, this rating scale did not provide adequate breadth of response; fellows tended to overestimate their skills and knowledge on the entry survey, leaving little room to measure improvement. In 2016, the rating scale was revised to a seven-point scale asking respondents to compare their level of knowledge/skill to that of a novice or expert:

- Expert (7): Think of your summer research host... An “Expert” is an experienced researcher who can do work independently without assistance or help and has the knowledge to guide/teach these skills to others.
- Novice (1): Think of a “first-timer” in the lab... a “Novice” has very limited lab experience, needs assistance or guidance in doing his/her work, and does not yet have the knowledge to guide/teach these skills to others.

Survey instructions reminded fellows that there are no right or wrong answers and encouraged fellows to answer thoughtfully and honestly. The respondents were told, “It is OK to be a novice at this stage of your career! We do not expect our undergraduate fellows to be proficient in all of these areas! If you find yourself marking ‘Expert’ on many or most questions, please read the definitions above carefully.” Fellows were asked to mark that they read the instructions and would follow them in answering the questions. The survey also included attention check questions (e.g., “Please mark this question with a ‘5’”) on several subscales to assure that fellows were reading and responding to the questions (13). If a fellow incorrectly answered an attention check question or marked the same answer or the same pattern of answers for most questions, the fellow was asked to review their answers on the survey (e.g., “Preliminary analysis of your entry survey responses suggests that you may not have thoroughly read and/or understood all of the questions in the survey... You are not required to change any answers.”). The small number of fellows who were asked to do so reviewed their survey responses for correctness.

Each survey included confidentiality assurances, a description of how the fellow’s name, institution, and response data would be used and a contact point for the fellow to ask questions about the program, survey and/or evaluation process. The evaluation plan was reviewed by the Federation of American Societies for Experimental Biology (FASEB) Institutional Review Board (IRB) Exemption Review Board and deemed eligible for educational exemption.

The survey modifications described above resulted in robust data to serve as the basis for the program analysis for years 2016–2018, reported below. The survey subscales (Table 1) were analyzed for both reliability (Cronbach’s alpha ranged from 0.78 to 0.97) and validity (confirmatory factor analysis factor loadings ranged from 0.50 to 0.95) using the entry survey data. Each subscale proved both reliable and valid for the constructs measured. Detailed information for each subscale, including items, are available in Supplemental Table S1 (https://doi.org/10.6084/m9.gshare.13670302.v1), and copies of the survey instruments are available online (https://doi.org/10.6084/m9.figshare.13344881.v1; and https://doi.org/10.6084/m9.figshare.13344962.v1).

The survey also included pre- and post-questions on fellows’ next steps in education and career choices. For these questions, respondents were asked to rate on a 7-point scale where 7 = “My top choice—I’m committed to pursuing this” to 1 = “Not at all interested in pursuing.” They also could select a “Don’t know much about it” option that was unscored. The exit survey also included open-ended questions on whether and how their career plans were changed by the fellowship and feedback on program aspects which they felt were especially good or needed work. The latter information was used as formative feedback for program improvement.

Comparison Groups

In addition to looking at changes on self-ratings for the total sample, we looked for differences in impacts by subgroups. For all comparisons, we excluded those who did not provide the demographic information on the survey or in their program application, therefore, total samples for the comparisons may be less than the total sample size of n = 146. Separate analyses were run by the following:

- **Fellowship program:** STRIDE (n = 45), UGSRF (n = 69), UGREF (n = 18), and IOSP (n = 14)
- **Gender:** Male (n = 46) and Female (n = 99)
- **URM status:** Underrepresented racial/ethnic minority (URM) (n = 39) fellows included those who self-identified as Hispanic/Latino, Black/African American, Native American/American Indian, and/or Native Hawaiian/...
Pacific Islander. The non-URM group included 102 fellows.

- **URP status:** Underrepresented persons (URP) \( n = 67 \) included URMs, persons with disabilities, and persons from disadvantaged backgrounds (economic or social). The non-URP group included 79 fellows.

- **Previous research experience:** Fellows who had 0 or <1 yr of any type of previous research experience were coded as “Less Research Experience” (LRE) \( n = 105 \). Those who had both 1+ years of previous research experience AND who had previously worked in a lab on their own project were coded as “More Research Experience” (MRE) \( n = 26 \).

### Data Analysis

Entry and exit scores were calculated for each fellow by totaling the fellow’s responses on the questions comprising each scale. For example, the Lab Safety scale has five questions, each asking the respondent to select a value from 1 to 7 on the Likert-type scale described above. A fellow’s score would be the sum of his/her five responses. Fellow’s entry survey scale scores were compared with their exit survey scale scores via a paired samples t test. Significance was set at \( P < 0.05 \). Reported means are an average of all fellows’ scores on the scale. Mean scores are described in the results text based on the percentage of maximum scale score: High (85%), moderately high (70%), moderate (50%), relatively low (30%), and low (15%). For example, a mean of 30.0 on a subscale where the maximum possible score is 35 would be described as “high.” Effect sizes were calculated via Cohen’s \( d \) and were rated as “small” \( (d = 0.2) \), medium \( (d = 0.5) \), or large \( (d = 0.8) \) \( (14) \).

The mean difference and the gain score on each scale was also calculated for each fellow. The gain score was the average of the exit score minus the entry score for each fellow.

The mean difference was the average of the absolute value of the exit score minus the absolute value of the entry score for each fellow, that is, the mean of the differences in before-and-after scores in paired observations. To determine whether there were differences in the comparison groups (women, URM, URP, and MRE/LRE), mean differences were compared (e.g., URM versus non-URM fellows) via one-way ANOVA with a significance level of \( P < 0.05 \). This allowed us to determine whether each comparison subgroup expressed significant increases in a skill over the course of the fellowship and whether the amount of increase differed between the subgroups (e.g., whether both men and women increased their general networking skills ratings and whether both men and women increased their ratings by similar amounts).

### RESULTS

Fellows’ responses on the entry and exit surveys are summarized in Table 2–5. Table 2 describes impacts on skills development on 15 scales in 6 areas: research, data analysis and presentation, scientific communication, professional networking, scientific ethics/responsible conduct of research (RCR), and laboratory personnel/research career training.

### Research Skills

Fellows rated their skills in designing experiments, data management and storage, and laboratory safety (Table 2, “Research Skills”). Entry ratings were moderate for Experimental Design and moderately high on the other two scales. On all three scales, fellows rated their skills higher on the exit survey compared with the entry survey. Effect sizes were high for Experimental Design and Data Management and Storage and moderate for Lab Safety.

---

**Table 1. Survey scale details on items, scoring, validity, and reliability by major topic**

| Major Topic/Scale Name | #Items | Minimum Score | Maximum Score | Entry Mean | Entry SD | Factor Loading Range | Cronbach’s Alpha |
|------------------------|--------|----------------|----------------|------------|----------|----------------------|-----------------|
| **Research Skills Understanding** | | | | | | | |
| Experimental Design | 5 | 5 | 35 | 20.2 | 6.1 | 0.83–0.88 | 0.92 |
| Lab Safety | 5 | 5 | 35 | 26.4 | 5.2 | 0.73–0.80 | 0.83 |
| Data Management | 5 | 5 | 35 | 24.0 | 5.6 | 0.66–0.89 | 0.83 |
| **Data Analysis Skills** | | | | | | | |
| Statistics | 6 | 6 | 42 | 21.8 | 7.9 | 0.71–0.88 | 0.89 |
| Data Presentation | 6 | 6 | 42 | 22.2 | 7.6 | 0.77–0.84 | 0.89 |
| **Scientific Communication** | | | | | | | |
| Scientific Writing | 11 | 11 | 77 | 34.3 | 15.3 | 0.74–0.86 | 0.95 |
| Oral Presentation | 10 | 10 | 70 | 35.7 | 15.8 | 0.82–0.90 | 0.97 |
| **Professional Networking** | | | | | | | |
| General Networking | 6 | 6 | 42 | 26.3 | 8.4 | 0.80–0.90 | 0.94 |
| Scientific Meeting Networking | 14 | 14 | 98 | 42.3 | 19.6 | 0.70–0.87 | 0.96 |
| **Scientific Ethics/RCR** | | | | | | | |
| Authorship Attribution | 7 | 7 | 49 | 20.1 | 9.8 | 0.72–0.88 | 0.91 |
| Using Animals in Research | 2 | 2 | 14 | 7.2 | 4.0 | 0.98–0.98 | 0.95 |
| Working with Human Subjects in Research | 2 | 2 | 14 | 4.9 | 3.2 | 0.91–0.91 | 0.78 |
| **Scientific Career Understanding and Planning** | | | | | | | |
| Scientific Career Understanding and Planning | 11 | 11 | 77 | 44.3 | 13.4 | 0.60–0.87 | 0.93 |
| Interactions with Research Mentor and Lab Group | 8 | 8 | 56 | 32.3 | 12.6 | 0.83–0.91 | 0.94 |

Entry data were used for determining reliability (Cronbach’s alpha) and validity (factor analysis) for each major topic area. Factor analysis data guided the division of major areas into scales. Cronbach alpha was deemed acceptable at 0.7 and factor loadings at 0.5. Additional details, including lists of items for each scale, are included in Supplemental Table S1. Survey copies are available for download in Supplemental materials. RCR, responsible conduct of research.
Table 2. Impacts on self-ratings of skills and knowledge

| Skills/Knowledge Scales (minimum/maximum score) | Entry Mean (SD) | Exit Mean (SD) | Mean Difference (SD) | Sig | Cohen’s d |
|-----------------------------------------------|-----------------|----------------|----------------------|-----|-----------|
| Research Skills                               |                 |                |                      |     |           |
| Experimental Design (5/35)                    | 20.2 (6.1)      | 25.7 (5.3)     | 5.1 (4.4)            | <0.001 | 0.96     |
| Data Management and Storage (5/35)            | 24.0 (5.6)      | 28.3 (4.8)     | 4.3 (4.8)            | <0.001 | 0.82     |
| Lab Safety (5/35)                             | 26.4 (5.2)      | 29.1 (4.8)     | 2.7 (3.8)            | <0.001 | 0.54     |
| Data Analysis and Presentation                |                 |                |                      |     |           |
| Statistical Analysis (6/42)                   | 21.8 (7.9)      | 29.3 (7.6)     | 9.2 (7.1)            | <0.001 | 0.97     |
| Data Presentation (6/42)                      | 22.2 (7.6)      | 30.9 (6.9)     | 9.7 (6.8)            | <0.001 | 1.20     |
| Scientific Communication                      |                 |                |                      |     |           |
| Scientific Writing Skills (11/77)             | 34.3 (15.3)     | 55.5 (13.0)    | 21.2 (14.0)          | <0.001 | 1.49     |
| Oral Presentation Skills (10/70)              | 35.7 (15.8)     | 53.2 (11.0)    | 19.0 (13.5)          | <0.001 | 1.28     |
| Professional Networking                       |                 |                |                      |     |           |
| General Networking (6/42)                     | 26.3 (8.4)      | 32.6 (5.7)     | 6.2 (6.6)            | <0.001 | 0.88     |
| Scientific Meeting Networking (14/98)         | 42.3 (19.6)     | 72.1 (14.9)    | 31.2 (17.7)          | <0.001 | 1.71     |
| Scientific Ethics/RCR                         |                 |                |                      |     |           |
| Authorship Attribution (7/49)                 | 20.1 (9.8)      | 34.5 (8.4)     | 14.6 (9.6)           | <0.001 | 1.58     |
| Use of Animals in Research (2/14)             | 7.2 (4.0)       | 9.8 (3.7)      | 2.6 (3.1)            | <0.001 | 0.67     |
| Human Subjects in Research (2/14)             | 4.9 (3.2)       | 7.1 (4.1)      | 2.2 (3.0)            | <0.001 | 0.60     |
| Lab Personnel/Research Career Training        |                 |                |                      |     |           |
| Lab Mates and Mentors (8/56)                  | 32.3 (12.6)     | 44.9 (8.7)     | 12.1 (10.9)          | <0.001 | 1.16     |
| Research Career Training (10/77)              | 44.3 (13.4)     | 59.2 (11.0)    | 14.7 (12.1)          | <0.001 | 1.22     |

For Scientific Writing Skills, n = 98 students; for all other scales, n = 146 students. Minimum and maximum score is listed for each scale. The same items are included in the entry and exit surveys. Participants rate their skills on each item on a scale from 1 to 7 where 7 = expert (“think of your summer research host...”); n = novice (“think of a first-timer in the lab...”). RCR, responsible conduct of research. Cohen’s d provides an estimate of the effect size: small (0.2), medium (0.5), and large (0.8).

Data Analysis and Presentation Skills

Fellows rated their skills/knowledge on statistical analysis of data and data presentation (Table 2, “Data Analysis and Presentation”). Entry ratings were moderate for both scales. Ratings significantly increased from entry to exit survey, and effect sizes were large for both scales (0.97 and 1.20, respectively).

Scientific Communication

The Scientific Writing and Oral Presentation scales asked fellows to rate their communication skills before and after the fellowship (Table 2, “Scientific Communication”). The Scientific Writing questions were added to the survey in 2017; therefore, the data presented includes only data from 2017 and 2018 cohorts (n = 98). Scientific Writing self-ratings increased significantly during the fellowship, and the effect size was large (1.27). For Oral Presentation skills, self-ratings increased and, again, the effect size was large (1.28).

Professional Networking

Networking skills were divided into a General Networking scale and a Scientific Meeting Networking scale (Table 2, “Professional Networking”). Self-ratings increased for both scales and the effect sizes were large, with Scientific Meeting Networking having the largest effect size of all the scales (1.71).

Scientific Ethics and Responsible Conduct of Research

Fellows rated their skills in three areas related to scientific ethics and RCR: Authorship Attribution, Use of Animals in Research, and Human Subjects in Research (Table 2, “Scientific Ethics/RCR”). Self-ratings increased significantly on all three scales from entry to exit surveys. Effect size was high on the Authorship Attribution scale (1.58) and moderate on the Use of Animals in Research (0.67) and Human Subjects in Research (0.60) scales.

Lab Personnel and Research Career Training

Fellows rated their knowledge of the roles they were expected to play in the research process in their laboratory as well as the roles of their fellow researchers (students, postdocs, technologists, etc.) and their research mentor (Table 2, “Lab Personnel/Research Career Training”). They also rated their skills/knowledge on how to find career information, plan one’s training and career, and write personal statements for graduate school applications. For both scales, self-ratings increased significantly from entry to exit survey for the total sample and effect sizes were large.

In summary, fellows in the APS summer research programs felt their skills and knowledge levels increased significantly in all areas assessed: research, data analysis and presentation, scientific communication, professional networking, scientific ethics, and working with laboratory personnel and research career training.

While all fellows in these programs experienced the same program components, they represented a diversity of demographic characteristics and previous research experiences. We determined whether there were differences in fellows’ responses depending on the fellowship program in which they participated, their gender, race/ethnicity, and disadvantaged status and whether they had more or less previous research experience. As described in METHODS, we compared their entry versus exit scores using a repeated measures t
test and analyzed the mean difference of their scores using ANOVA (e.g., male vs. female). Both tests used a significance cutoff of $P < 0.05$. Post hoc analyses (Scheffé’s test, $P < 0.05$) were done for the comparisons by program since there were more than two programs. Overall, there were very few significant differences. Results are summarized below by comparison group.

Program

For all survey scales, participants in all programs reported significantly higher skills self-ratings on the exit survey compared with the entry survey. Also, the size of the self-ratings increase was similar for all programs with four exceptions: ANOVAs indicated that ratings increases were different among the four programs for Experimental Design ($F = 3.515$, df = 3, $P = 0.17$), Lab Safety ($F = 18.656$, df = 3, $P = 0.029$), Data Presentation ($F = 4.631$, df = 3, $P = 0.004$), and Oral Presentation ($F = 91.175$, df = 3, $P = 0.003$). To identify the specific differences, we used Scheffé’s post hoc analysis since the groups were not of even sizes and set significance at $P < 0.05$. On the Experimental Design scale, UGREF fellows reported a larger increase than did UGSRF fellows (UGREF mean difference 5.8; UGSRF mean difference 4.6, $P = 0.034$). No other differences reached significance. On Lab Safety, UGSRF and IOSP fellows reported the largest increases in skills (mean difference was 4.9 for each) compared with STRIDE and UGREF fellows (mean differences were 4.3 and 4.2, respectively) but the Scheffé’s test was not significant ($P = 0.310$). For Data Presentation, IOSP fellows reported greater gains than did UGREF fellows (IOSP mean difference 8.5, UGREF mean difference 4.9) but the Scheffé’s test significance level was marginal ($P = 0.053$). On the Oral Presentation scale, UGSRF students reported the highest skills increases (mean difference = 21.9) and UGREF the lowest (mean difference = 12.4) but the Scheffé’s test was not significant ($P = 0.060$).

Gender

Like the total sample, both male and female fellows significantly increased their self-ratings of skills/knowledge from the entry to the exit survey on all scales listed in Table 2. When comparing the ratings of male versus female fellows, there was only one significant difference. Male fellows felt their skills increased more than did female fellows on the Human Subjects in Research Scale (male fellow mean gain, 3.5; female fellow mean gain, 1.6, $F = 8.678$, df = 1, $P = 0.004$).

URM Status

Both URM and non-URM fellows significantly increased their self-ratings on all the Table 2 scales over the course of the program. Likewise, there were no differences between URM and non-URM fellows in their estimates of how much their skills/knowledge had improved.

URP Status

The results for URP status, including URM students, students with disabilities, and students from disadvantaged backgrounds, were the same as for URM status. URP and non-URP fellows increased their ratings on all scales and by similar amounts.

Previous Research Experience

There were more differences between MRE and LRE fellows than for any other comparison group. On all but four scales, both LREs and MREs significantly increased their self-ratings over the course of the fellowship. On the remaining four scales, LREs significantly increased their self-ratings while MREs did not: Data Management and Storage (LRE mean difference 5.2, SD 6.0, $P < 0.001$; MRE mean difference 2.0, SD 6.0, $P = 0.099$); Laboratory Safety (LRE mean difference 3.1, SD 5.2, $P < 0.001$; MRE mean difference 1.2, SD 5.4, $P = 0.286$); Use of Animals in Research (LRE mean difference 3.2, SD 4.1, $P < 0.001$; MRE mean difference 0.8, SD 2.5, $P = 0.132$); and Human Subjects in Research (LRE mean difference 2.2, SD 3.4, $P < 0.001$; MRE mean difference 1.3, SD 4.4, $P = 0.158$).

LRE fellows’ self-ratings increased more than did those of MRE fellows on several scales:

- Experimental Design (LRE mean difference 6.4; MRE mean difference 2.9; $F = 6.963$, df = 1, $P = 0.009$);
- Data Management and Storage (LRE mean difference 5.2; MRE mean difference 2.0; $F = 5.955$, df = 1, $P = 0.016$);
- Data Presentation (LRE mean difference 10.1; MRE mean difference 3.6; $F = 15.585$, df = 1, $P < 0.001$);
- Scientific Writing (LRE mean difference 24.3; MRE mean difference 12.3; $F = 9.692$, df = 1, $P = 0.003$); Oral Presentation (LRE mean difference 19.7; MRE mean difference 8.3; $F = 11.886$, df = 1, $P = 0.001$);
- Scientific Meeting Networking (LRE mean difference 31.8; MRE mean difference 20.5; $F = 7.123$, df = 1, $P = 0.009$);
- Use of Animals in Research Scale (LRE mean difference 3.2; MRE mean difference 0.8; $F = 8.811$, df = 1, $P = 0.004$);
- Authorship Attribution (LRE mean difference, 15.8; MRE mean difference 7.6; $F = 14.957$, df = 1, $P < 0.001$);
- and
- Lab Mates and Mentors (LRE mean difference 14.4; MRE mean difference 5.3; $F = 11.685$, df = 1, $P = 0.001$).

On all other scales where both LREs and MREs increased their self-ratings, the gains were similar for both groups.

Career Goal Impacts

On both the entry and exit surveys, fellows indicated their current plans after completing their undergraduate degrees (Table 3). Graduate school in science and medical school remained the most common goal. On most items, there was no change in fellows’ interests over the course of the fellowship. However, there was decreased interest in entering combined degree programs such as MD/PhD and there was increased interest in finding a job in science or other fields. Fellows also indicated their interest in several research-related and nonresearch related career options on both the entry and exit surveys (Table 4). Research careers were rated highly while non-science careers were rated lowest. On the exit survey, fellows were asked whether the fellowship activities changed their career plans and, if so, how. Nearly a third (31%, $n = 45$) described how their career goals were changed or impacted by the fellowship (Table 5). Most of
undergraduate research builds skills for diverse students

Table 3. Few career goals changed during fellowship

| Career Goal                                      | Entry Mean (SD) | Exit Mean (SD) | Mean Difference (SD) | Sig   |
|-------------------------------------------------|-----------------|----------------|----------------------|-------|
| Attend graduate school in science.              | 5.2 (1.6)       | 5.2 (2.1)      | 1.4 (1.5)            | 0.780 |
| Attend medical school (e.g., MD, DO).           | 5.0 (2.1)       | 4.8 (2.4)      | 0.9 (1.3)            | 0.164 |
| Attend combined degree program (e.g., MD/Ph.D.).| 4.6 (1.9)       | 3.8 (2.2)      | 1.8 (1.6)            | <0.001|
| Find a job after graduation in science.         | 4.2 (2.2)       | 4.9 (2.3)      | 2.0 (1.8)            | 0.001 |
| Attend other professional school in science (e.g., DVM, DDS, PA, PT). | 3.0 (2.1)       | 2.7 (2.2)      | 1.3 (1.5)            | 0.155 |
| Find a job after graduation, but not in science.| 1.5 (1.0)       | 1.5 (1.3)      | 0.8 (1.1)            | 0.011 |
| Attend professional school, but not in science (e.g., JD, MBA). | 1.5 (1.1)       | 1.9 (1.6)      | 1.0 (1.3)            | 0.764 |
| Attend graduate school, but not in science.     | 1.4 (1.1)       | 1.7 (1.6)      | 0.3 (1.4)            | 0.069 |

Career goals questions included 8 items that participants answered on a scale with a minimum score of 1 and a maximum score of 7 where 7 is “My top choice(s)–I am committed to pursuing this” and 1 is “not at all interested in pursuing.” The same items are included in the entry and exit surveys. SD, estimated standard deviation of the sample mean; mean difference, average difference between individual entry and exit scores; Sig, two-tailed P value for paired observations of entry and exit scores computed using the t distribution. P < 0.05 is considered significant. Only 3 item means changed significantly; participants expressed decreased interest in entering combined degree programs and increased interest in getting a job after graduation.

their responses (76%) indicate a movement toward inclusion of research in their future careers. Nearly a quarter planned to include clinical research as part of their medical careers and 22% switched from other degrees to a PhD or MD/PhD.

**DISCUSSION**

We sought to evaluate the APS undergraduate summer research fellowship programs for their impacts on scientific research and communication skills and understanding of and interest in research careers. Each of these were primary goals for all four programs and encompassed a sizeable group of objectives. In addition, we sought to assess the effectiveness of the programs for diverse students.

Self-ratings clearly indicate that the fellows believed they not only accomplished the measurable objectives of the programs but gained additional research and professional skills in areas of professional ethics. In each area measured, fellows rated their knowledge/skills significantly higher after they completed the fellowship. Effect sizes (Cohen’s d) were large for nearly all measured impacts on skills and knowledge (78%). The three remaining scales (Lab Safety, Use of Animals in Research, and Human Subjects in Research) had moderate effect sizes.

The fellowships had an especially strong impact on fellows who had little or no previous research experience. They made significant gains in every area and, in a few cases, made greater gains than did the fellows who started the fellowship with more research experience. Thiery et al. (15) also noted the steep learning curve experienced by undergraduate research “novices.” Since the current study focused primarily on evaluating objectives designed for novice research fellows, there was less focus on more advanced skills such as manuscript development, responding to reviews, giving seminars, and making oral presentations.

The results also suggest that the programs were effective for a very diverse population of undergraduate students, including women, underrepresented minorities, persons with disabilities, and persons from disadvantaged backgrounds (e.g., first generation college, financial and/or social disadvantage). Our interview study with past STRIDE fellows had similar findings (11).

Professional societies are uniquely positioned to facilitate the development of scientific trainees’ professional networks and provide opportunities for pivotal events such as early career presentations, publications, awards, membership, and engagement in society leadership opportunities. While they cannot fill the critical one-on-one mentoring roles that mentors have with their students, scientific societies make important contributions to both the psychosocial and career mentoring activities in which a trainee engages. Psychosocial mentoring activities enhance the student’s “sense of self and

Table 4. Research career goals remained steady during the fellowship

| Career Goal                                      | Entry Mean (SD) | Exit Mean (SD) | Mean Difference (SD) | Sig   |
|-------------------------------------------------|-----------------|----------------|----------------------|-------|
| A career in physiology research.                | 5.0 (1.4)       | 4.7 (1.8)      | 1.4 (1.4)            | 0.124 |
| A career in biomedical research.                | 4.9 (1.7)       | 4.7 (2.0)      | 1.3 (1.4)            | 0.323 |
| A career as a MD with a clinical research component. | 4.9 (2.1)       | 4.8 (2.2)      | 1.1 (1.3)            | 0.720 |
| A career in cardiovascular research.            | 4.3 (1.7)       | 4.1 (2.0)      | 1.3 (1.2)            | 0.172 |
| A career in digestive, diabetes, or kidney research. | 4.3 (1.9)       | 4.0 (1.9)      | 1.4 (1.3)            | 0.630 |
| A career in comparative physiology research.    | 3.4 (1.8)       | 3.2 (2.0)      | 1.2 (1.5)            | 0.590 |
| A career as a MD not involving research.        | 3.1 (1.8)       | 3.2 (2.0)      | 1.2 (1.2)            | 0.634 |
| A career in another science-based profession (e.g., DDS, PA, DO). | 3.1 (1.9)       | 3.0 (2.2)      | 1.4 (1.5)            | 0.571 |
| Graduating and going directly into the workforce. | 2.5 (1.6)       | 2.2 (1.7)      | 1.2 (1.4)            | 0.052 |
| A career in a non-science-based profession (e.g., law or business). | 1.4 (1.0)       | 1.7 (1.3)      | 0.7 (1.1)            | 0.046 |

Research career goals questions included 10 items that participants answered on a scale with a minimum score of 1 and a maximum score of 7 where 7 is “My top choice(s)–I am committed to pursuing this” and 1 is “not at all interested in pursuing.” The same items are included in the entry and exit surveys. SD, estimated standard deviation of the sample mean; mean difference, average difference between individual entry and exit scores; Sig, two-tailed P value for paired observations of entry and exit scores computed using the t distribution. P < 0.05 is considered significant. Only 2 item means changed significantly; Participants expressed decreased interest in getting a job immediately after graduation and expressed greater interest in nonscience careers.
Table 5. Planned changes in career goals after fellowship

| Planned Change                      | Number of Responses | Percentage of Responses |
|-------------------------------------|---------------------|-------------------------|
| Add clinical research to MD career* | 11                  | 24%                     |
| Increased interest in research career*| 7                   | 16%                     |
| Validated pursuit of research career*| 6                   | 13%                     |
| Changed career plans: MD to MD/PhD* | 4                   | 9%                      |
| Changed career plans MD to PhD*     | 3                   | 7%                      |
| Switch to a new research area       | 3                   | 7%                      |
| Get another degree before continuing research career training | 3 | 7% |
| Decreased interest in research career | 3                | 7%                      |
| Moved from MD/PhD to PhD*           | 2                   | 4%                      |
| Moved from DPT to PhD*              | 1                   | 2%                      |
| Expanded my career options          | 1                   | 2%                      |
| Became undecided about career goal   | 1                   | 2%                      |

Planned changes in career goals were tabulated from optional open-ended responses (n = 45; 31% of total respondents). “These changes indicate strengthened research career goals.

their social relation with their environment” (16) (p. S51).

Fellowship programs such as undergraduate research and travel award programs contribute to this aspect of mentoring of students, especially underrepresented students (5). Career mentoring activities offer access to professional society career services, enhance skills, and build professional contacts (16). Scientific societies contribute through not only their scientific meetings and publications but through targeted activities such as in-person and virtual diversity networking sessions at meetings, professional development workshops, inventories of appropriate skills for trainees, and directories of research opportunities.

For underrepresented students, specific efforts that incorporate cultural aspects are critical. These aspects are often not recognized and not addressed by individual advisors and mentors (see Ref. 16 for review of numerous studies). The success of the APS programs with diverse participants may reflect the professional socialization activities incorporated into each of the programs. These included online discussions with APS members and EB activities designed to help undergraduates network with established researchers (orientation session, undergraduate poster session and graduate school tables, diversity networking breakfast, etc.). Development of one’s science identity and integration into the professional community are especially important for underrepresented STEM students (see review in Ref. 5). Activities that enrich opportunities for networking, especially with diverse fellows, should be included in not only undergraduate but graduate and postdoctoral programs to continue to strengthen science identity for persons underrepresented in STEM. The American Association for the Advancement of Science’s 1984 study, “Equity and Excellence: Compatible Goals,” found that “Exemplary programs are sensitive to the group or groups they are intended to serve and address these audiences’ fundamental needs...” (17) (p. vii). The success of the APS programs reflects the intentional design of these programs to address the needs and interests of diverse students, including early professional socialization.

In terms of career choice, APS program activities focused on broadening fellows’ understanding of what research is, how it is done, and who does it. They also engaged students in exploring careers that were not their top choice to broaden their perspectives on possible career options. The focus was, as noted by Lopatto (1), on empowering students to make a variety of career choices. Survey responses showed that those who received APS research fellowships entered the programs primarily focused on working toward careers in research or medicine, with their top goals being graduate or medical school. As they left the program, their focus was essentially unchanged. Did the fellowship program significantly impact their career plans in any way? Open-ended responses (Table 5) suggest that the program gave fellows an important opportunity to find out if research was their passion. Forty-six fellows (32% of total surveyed) described planned changes in their career goals after the fellowship. Fifteen fellows (10% of the total sample) who planned to enter medical school decided they wanted to continue their research by adding clinical research to their MD career or switching from an MD to a MD/PhD degree. The STRIDE interview study found similar results (11). An additional six fellows (4%) switched their plans from earning other degrees (MD, MD/PhD, DPT) to seeking a PhD, and six more (4%) said the program experience validated their pursuit of a research career. Finally, seven fellows (5%) gained insights into career planning that led them to new research areas, new career options, and plans for additional degrees in conjunction with plans to pursue a PhD. Overall, 45 of 147 program fellows (30%) described a direct impact on their career plans and only three of those responding (2%) said their interest in a research career decreased as a result of the program. In the future, APS programs should ask additional direct questions about career choices and program impacts on careers.

Two significant changes in fellows’ career goals should be noted. There was a significant decrease in fellows’ plans to enter an MD/PhD program (Table 3). Although there were no comments in the open-ended questions on the reasons for this change, in the STRIDE interview study, some fellows noted that the length of time needed to complete both degrees before taking a full-time position was intimidating (11). It is not clear whether the students in the current study understood that MD/PhD programs typically offer financial aid to help offset the cost of medical schools (18). However, the numerous students whose degree goals moved between PhD, MD, and MD/PhD suggests that the fellows are seeking more information about what these programs entail, financial options for each, and how each degree can include a strong focus on research (basic or clinical). Programs should provide a balanced view of each, outlining pros and cons and offering perspectives on research careers from persons with those degrees.

Second, fellows’ plans to find a science job immediately after graduation increased significantly over the course of the program. This may reflect the growing trend of taking a gap year to prepare medical or graduate school applications or to work at a postbaccalaureate program to improve academic records (e.g., missing courses and/or poor grades) (19, 20). Several STRIDE fellows in the interview studies secured jobs as technicians in the laboratories where they did research as undergraduates to have time to prepare graduate/medical school applications (11).
Finally, the current study provided an opportunity to develop a set of reliable and valid measures of undergraduate students' self-assessments of their research skills. It is hoped that these will provide other professional societies as well as campus-based programs additional evaluation tools for measuring program impacts.

**ACKNOWLEDGEMENTS**

The authors are deeply grateful to the undergraduate students who thoughtfully completed the evaluation surveys and to the American Physiological Society (APS) members who served on the Career Opportunities in Physiology Committee and the Short-Term Research Education Program to Increase Diversity in Health-Related Research Fellowships (STRIDE) and Integrative Organismal Systems Physiology (IOSP) Advisory Boards. They also thank the APS Council and executive directors Martin Frank and Scott Steen for providing leadership and funding, and former APS staff members Melinda Lowy and Allison Hood who worked on these programs and their evaluations.

**GRANTS**

The programs and study were funded by the American Physiological Society, National Heart, Lung, and Blood Institute Grant R25-HL-115473, and National Science Foundation Division of Integrative Organismal Systems Grant 1238831.

**DISCLOSURES**

C. B. Bruthers and E.L. Hedman are employees of APS. M. Matyas is a retired APS employee and currently serves as a consultant to APS.

**AUTHOR CONTRIBUTIONS**

C.B.B., E.L.H., and M.L.M. conceived and designed research; C.B.B., E.L.H., and M.L.M. performed experiments; E.L.H. and M.L.M. analyzed data; C.B.B., E.L.H., and M.L.M. interpreted results of experiments; M.L.M. prepared figures; C.B.B. and M.L.M. drafted manuscript; C.B.B., E.L.H., and M.L.M. edited and revised manuscript; C.B.B., E.L.H., and M.L.M. approved final version of manuscript.

**REFERENCES**

1. Lopatto D. *Science in Solution: The Impact of Undergraduate Research on Student Learning*. Tucson, AZ: Research Corporation for Science Advancement, 2009.
2. *National Academy of Sciences (NAS). Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report*. Washington, DC: The National Academies Press, 2017.
3. *President’s Council of Advisors on Science and Technology (PCAST). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering and Mathematics*. Washington, DC: Executive Office of the President of the United States, 2012.
4. Estrada M, Burnett M, Campbell AG, Campbell PB, Denetclaw WF, Gutierrez CG, Hurtado S, John GH, Matsui J, McGee R, Okpudo CM, Robinson TJ, Summers MF, Werner-Washburne M, Zavala M. Improving underrepresented minority student persistence in STEM. CBE Life Sci Educ 15:e55, 2016. doi:10.1187/cbe.16-01-0038.
5. Hurtado S, Cabrera NL, Lin MH, Arellano L, Espinosa LL. Diversifying science: underrepresented student experiences in structured research programs. Res High Educ 50: 189–214, 2009. doi:10.1007/s11162-008-9114-7.
6. *National Academy of Sciences (NAS). Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities*. Washington, DC: The National Academies Press, 2017.
7. Krim JS, Cote LE, Schwartz RS, Stone EM, Cleeves JJ, Barry KJ, Burgess W, Buxner SR, Gerton JM, Horvath L, Keller JM, Lee SC, Locke SM, Rebar BM. Models and impacts of science research experiences: a review of the literature of CUREs, UREs, and TRES. CBE Life Sci Educ 18: ar65, 2019. doi:10.1187/cbe.19-03-0069.
8. Linn MC, Palmer E, Baranger A, Gerard E, Stone E. Undergraduate research experiences: impacts and opportunities. Science 347: 1261757, 2015. doi:10.1126/science.1261757.
9. Vispoel WP, Austin JR. Success and failure in junior high school: a critical incident approach to understanding students’ attributioinal beliefs. Am Educ Res J 32: 377–412, 1995. doi:10.3102/00346543071004549.
10. Matyas ML, Ruedi EA, Engen K, Chang AL. Life science professional societies expand undergraduate education efforts. CBE Life Sci Educ 16:ar5, 2017. doi:10.1187/cbe.16-01-0019.
11. Bruthers CB, Matyas ML. Undergraduates from underrepresented groups gain research skills and career aspirations through summer research fellowship. Adv Physiol Educ 44: 525–539, 2020. doi:10.1152/advan.00014.2020.
12. Fuhrmann CN, Hobin JA, Lindstaedt B, Clifford PS. MyIDP–Individual Development Plan. Washington, DC: American Association for the Advancement of Science, 2019.
13. Kung FY, Kwok N, Brown DJ. Are attention check questions a threat to scale validity? *Appl Psychol 67*: 264–283, 2018. doi:10.1111/apps.12108.
14. McLeod S. What does effect size tell you? *Simply Psychology*. https://www.simplypsychology.org/effect-size.html. [2019].
15. Thiry H, Weston TJ, Laursen SL, Hunter AB. The benefits of multi-year research experiences: differences in novice and experienced students’ reported gains from undergraduate research. CBE Life Sci Educ 11: 260–272, 2012. doi:10.1187/cbe.11-11-0098.
16. Davidson MN, Foster-Johnson L. Mentoring in the preparation of graduate researchers of color. *Rev Educ Res 71*: 549–574, 2001. doi:10.3102/00346543071004549.
17. Malcom SM. Equity and Excellence: Compatible Goals. Washington, DC: American Association for the Advancement of Science, 1983.
18. *Association of American Colleges (AAMC). Financial Support for MD-PhD Trainees*. Washington DC: AAMC, 2021.
19. Bodine PS. A Guide to Postbac Premed Programs: Another Route to Med School. Noodle. https://www.noodle.com/home. [2015].
20. Digirolama M. Taking a Gap Year before Graduate School. GoGrad.org. https://www.gograd.org/. [2020].