SBE Office of Multi-Disciplinary Activities (SMA)
SBE Office of the Assistant Director
Directorate for Social, Behavioral, and Economic Sciences (SBE)
National Science Foundation

Committee of Visitors
July 13-15, 2020

Introduction

The Committee of visitors (COV) for the SMA met from July13-15, 2020 in the first ever virtual meeting of a COV. The charge to the committee was as follows:

> "NSF relies on the judgment of external experts to maintain high standards of program management, to provide advice for continuous improvement of NSF performance, and to ensure openness in the research and education community served by the Foundation. The Committee of Visitors (COV) report provides NSF with the judgment of external efforts in two primary areas: (A) the integrity and efficiency of the **processes** related to proposal review; and (B) the quality of the **results** of NSF's investments in the form of outputs and outcomes that appear over time. The COV also explores the relationships between award decisions and program/NSF-wide goals in order to determine the likelihood that the portfolio will lead to the desired results in the future".

Four programs were reviewed:

- Resource Implementations for Data Intensive Research in the Social, Behavioral and Economic Sciences (RIDIR)
- Science of Science and Innovation Policy (SciSIP)
- Research Experiences for Undergraduates Sites (REU)
- SBE Postdoctoral Research Fellowships (SPRF)

These programs are all multidisciplinary, cutting across the disciplines of SBE, which is why they are housed in SMA. RIDIR and SciSIP are research-building programs whereas REU and SPRF are capacity building (pipeline) programs. The goal of **RIDIR** is to support the creation of large-scale state-of-the-art data resources (e.g., databases) and relevant analytic tools to enhance data-intensive research in the SBE sciences. The goal of **SciSIP** is to create new models, datasets, and analytic tools that can help the SBE sciences better determine how investment in research and development can contribute to important outcomes and science policy decision making. The goal of **REU** is to support research in the SBE sciences that actively engages undergraduates. And the goal of **SPRF** is to support postdoctoral training in one of two tracks: Fundamental Research in the SBE sciences or Broadening Participation in the SBE sciences.

These four programs were last reviewed in 2015. Our review covered the 5-year period from 2015 to 2019. We note that RIDIR has recently been replaced by Human Networks and Data Sciences (HNDS): HNDS-I on the development of infrastructure and HNDS-R on core research.

The COV was comprised of the chair and eight members, two of whom were assigned to each program under review. Each pair was given a detailed Self-Study Guide and electronic access to program-specific materials, including a random sample of about 5 percent of awarded and declined proposals and review-related documents as well as administrative and financial data on the MyNSF COV Dashboard. COV members also had opportunities to meet with program officers and SBE staff.

The COV met virtually for three 5-hour days (11am-5pm EDT). Each day was divided into individual meetings of COV pairs to complete the self-study report on their assigned program, meetings with program officers and NSF staff, and whole-group plenary sessions at the beginning and end of each day to craft general recommendations that cut across all four programs. Those general recommendations are discussed below, followed by program-specific reports adhering to the NSF template provided in the self-study guide. Both our COV recommendations and the program-specific reports were framed around two broad issues: (1) merit review criteria, including whether the process of evaluating proposals in each program adequately addresses both intellectual merit and broader impacts, reviewer selection, and quality of the reviews; and (2) management of the program under review, including program planning and priorities, appropriate portfolio balance, and responsiveness to emerging research and education priorities.

We would like to start by acknowledging and thanking NSF staff (particularly John Garneski) who were extremely supportive and helpful throughout the review process. As the first-ever virtual COV meeting, all of us approached this 3-day gathering with some degree of trepidation. COV members had to weave in and out of break-out rooms, navigate screen sharing, engage in group discussions with colleagues they never actually saw in person, and negotiate the vast amount of electronic data, while at the same time writing individual program reports. The technology for a virtual meeting worked incredibly well – there were very few glitches and these were solved quite quickly. The NSF staff was technologically very well prepared to facilitate the virtual meeting in addition to being willing to share their knowledge and gather new data on the spot about the various programs.

COV2020 Assessment and General Recommendations

In general, COV members were very positive about the programs they reviewed. Program officers, both permanent and rotating, were perceived as conscientious, hard-working, and committed to the mission of SBE as enacted in the specific programs they headed. Working within the constraints of limited funding, the allocation of awards seemed thoughtful and balanced with the need to be equitable. Given the differences across programs, core questions, concerns, and specific recommendations are addressed in the individual program reviews. In our general discussions, however, a number of concerns and themes that cut across programs emerged and resulted in the recommendations detailed below. We offer 10 recommendations that cover the merit review process, quality of reviews within the structure of a program panel system, issues of broadening participation, and SMA management.

Merit Review Process

While the panel review mechanism works well in SBE and in these four programs, there are continuing challenges associated with quality of panel reviews and summaries, creating panels that have the needed disciplinary, geographic, and demographic diversity, and sustained engagement of panel members. There are also administrative burdens and costs of convening panels twice a year, which is standard practice for SBE. COV teams also commented that regardless of the size of the panel, only the three participants who actually provided written evaluations and scores of a given proposal participated in the discussion, leading to questions about the contribution of panel members as a group to decisions about a proposal. The practice of requiring a panel member (scribe) to write the panel summary further detracted from member engagement. In general, panel summaries were viewed as needing serious re-thinking. With all the challenges of convening and administering panels, some COV members wondered how, in practice, funding decisions based on the panel model differ from a simple average of ad hoc reviewer scores. SBE seems to have adopted a particular model for proposal review that generally works well but could benefit from some improvements. In the absence of systematic research on different panel structures, we cannot know what those improvements might entail.

> **Recommendation #1**: We recommend that SBE undertake a more systematic analysis of the panel review process with the goal of identifying what works well and what would benefit from improvement. Some practices such as who prepares panel summaries or the number of ad hoc reviews about which we have impressions of effectiveness (or not) would benefit from more rigorous analysis. At present, panel-and review-related data are not as robust as needed to study effectiveness of the review process as it currently exists and to consider alternative models.

Insufficient attention to measuring long-term outcomes and research impact of funded proposals is a concern raised in previous COV reviews and one which our COV also detected. For example, PIs submit annual and final reports of proposal outcomes that are stored somewhere and it is unclear what happens to those reports. Individual programs have program-specific mechanisms to evaluate funded proposals in the short-and long-term, but there is no division-wide infrastructure to organize these efforts. We should look toward other government funding agencies that do a good job of tracking long-term outcomes. The Office of Portfolio Analysis (OPA) at NIH and its *iCite* tool is a good example of modern methods for documenting program impact.

> **Recommendation #2**: We recommend that SBE capitalize on the expertise of RIDIR and SciSIP in particular to create a data infrastructure to track program outcomes over time using state-of-the-art methods. Such an infrastructure would also facilitate our efforts to ask the appropriate questions and keep track of long-term outcomes without having to add additional work for PIs after the research support has ended.

Quality of Proposal Reviews

The varying quality of reviews by panel members and ad hoc reviewers across programs remains a concern. Some reviews tend to be long and detailed whereas others are quite cursory. Some reviewers disproportionately highlight weaknesses with very little attention to strengths. Some proposals have no ad hoc reviews whereas others have many (11 in one case) and we have no idea of the impact on scoring of multiple reviews. There is also great variance in the extent to which the Broader Impact criteria are considered. In a noticeable number of cases proposal scores did not closely match reviews and reviews were not well reflected in panel summaries. This kind of mismatch may be more likely to occur in multidisciplinary and capacity building (pipeline) programs because of the diverse backgrounds of panel members and ad hoc reviewers across fields and institutions. In addition to undermining the overall integrity of reviews, we think these slippages pose particular challenges for PIs of declined proposals who look to reviews to guide their decisions about whether to re-submit a proposal. We offer four specific recommendations aimed at improving the overall quality of reviews, scoring, and feedback to PIs.

Recommendation #3: We recommend that there be better training of reviewers – both panel members and ad hoc – in how to write an effective review (e.g., highlighting strengths as well as weaknesses in a proposal).

Recommendation #4: Scoring of proposals is the lever for funding. There appears to be considerable variability in scores within a proposal and such variability has unknown impact on whether a proposal achieves a fundable score. We recommend that SBE consider implementing calibration exercises during panel meetings to gain more insight into the challenges of scoring variability.

Recommendation #5: Scoring impacts which proposals get triaged (i.e., not discussed) and there are many unknowns about the triage process. For example, do triaged proposals have consistently low scores across all reviews? Are they more likely to come from members of underrepresented groups or minority serving institutions? We recommend that SBE systematically analyze the process by which proposals are triaged, the characteristics of such proposals, and the kind of feedback given to PIs of these unsuccessful proposals that do not have the benefit of a panel summary.

Recommendation #6: Writing good reviews presents workload issues for panel members. We are concerned that some panel members are overburdened with the required number of proposals to review and how this varies across programs. We recommend that SBE analyze how workloads (number of proposals to review) varies across programs and address question such as whether there is an association between number and quality of reviews from panel members. We also recommend that panelist compensation adjustments may need to be considered for time commitment and to emphasize the importance and significance of their reviews. Workload expectations should clearly be articulated to reviewers to ensure sufficient time for thorough reviews.

Broadening participation

Fundamental to the mission of NSF is to broaden participation of diverse populations, including members of underrepresented groups, in the science enterprise. The nation's population is rapidly changing as are the research needs of a society that must keep pace with this dramatic population shift. A core value of NSF is to "support outstanding researchers and innovative thinkers from across our Nation's diversity of regions, types of organizations, and demographic groups". The starting point to addressing this mission is to know who submits and who gets funded. Yet NSF, including SBE, lacks a mechanism to gather reliable data on the diversity of PIs and Co-PIs of submitted and funded projects. This seems to be a persistent problem across SBE given regulations that providing demographic information is purely voluntary.

> **Recommendation #7**: We recommend that SBE consider innovative ways to increase the willingness of applicants for funding to report critical demographic information such as gender and race/ethnicity. If reporting of such information remains voluntary, consider situating requests with explanations of why such information is important and how it will be used. It could be that the way NSF currently collects demographic data creates uncertainty or disincentives to respond. NSF should update systems, such as Fastlane, to increase reporting of demographic data. This includes question ordering and sequence, clear disclaimers, annual updates, etc. The COV strongly believes that accurate tracking of the diversity of applicants is essential.

All of the COV teams noted that members of underrepresented groups and MSIs were not well represented in their program portfolios. In many cases, race/ethnicity of PIs and undergrads in the REU program was not available, which only compounded the problem of knowing who applied and who got funded. In SPRF, COV members were concerned that there were few applicants to the Broadening Participation track compared to the Fundamental Research track. Because of the big data infrastructure nature of the RIDIR and SciSIP programs, there were few (if any) applicants from members of underrepresented groups or MSIs. In other words, broadening participation was a concern in all four programs reviewed by the COV. The COV concluded that collaborations between established researchers and scholars from MSIs and/or members of underrepresented groups remains a challenge. This lack of collaboration is particularly noteworthy and potentially problematic in multidisciplinary programs such as those in SMA.

Recommendation #8: We recommend that SBE engage in more systematic efforts to work with the HBCU program, the new Building and Broadening initiative, and NSF INCLUDES) as good mechanisms for identifying promising collaborators from underrepresented groups. Individual program solicitations should highlight the importance of collaborations and how they might be initiated. Regular availability of webinars for grant preparation is advisable, as well as outreach to relevant college and university graduate programs (e.g., Alliances for Graduate Education and the Professoriate, AGEP). We further recommend outreach not just to potential applicants, but also to potential faculty mentors as well. For example, faculty can be encouraged to identify individuals who might consider a postdoctoral fellowship, and then encourage and support the proposal development.

Even if we master the challenge of broadening participation in the submission process, SBE programs continue to confront the fact that applicants from underrepresented groups are still less likely to have fundable proposals. Some of this disparity is likely due to biases – both conscious and unconscious – that enter the review process. For example, reviewers may be unknowingly biased for or against a proposal based on the gender, race/ethnicity, or institutional affiliation of the applicant. We believe that it is critically important where possible to decouple investigator characteristics from evaluations of the merit of specific proposals.

> **Recommendation #9**: We recommend that SBE conduct systematic studies (e.g., double blind review) appropriate for specific programs to evaluate whether and how proposal evaluation is influenced by knowledge of PI demographic characteristics. If it is not, then knowing PI and Co-PI demographics can be an asset rather than deterrent to broadening participation.

SMA management

The COV was overall very positive about the management of the four programs included in this review. Programs officers were seen as conscientious and committed, even when they were responsible for more than one program. They made good efforts toward balancing their portfolios between, for example, early career and more established researchers and securing co-funding, which is critical for these multidisciplinary programs. COV members did express some frustration because critical data were missing from the individual proposals (e.g., PI demographics, geographical region) to fully evaluate portfolio balance.

One issue that all program reviews noted was the balance between permanent and rotating program officers. On the one hand, rotators are especially valuable for keeping fresh ideas on the table and for facilitating responsivity to new directions in a field. On the other hand, because they are only at NSF two years, rotators take up valuable time of staff who must train new recruits on a regular basis. Permanent program officers have unique value because of their institutional memory and the need for co-funding. Permanent staff are more likely to have established relationships with other programs in SBE and staff in other divisions, which are critical for co-funding agreements. But it is difficult to keep up this pace for permanent officers who manage more than one program. We note that RIDIR and SciSIP have never had their own permanent program officer and REU and SPRF share one officer who appears to give 100 percent to each program.

> **Recommendation #10:** We recommend that each program in SMA have at least one permanent member. When program officers are responsible for more than one program, we recommend that each managed program has a rotator to assist the permanent officer as needed. The ideal staffing composition is a dedicated, permanent PO and a rotator per SMA program.

Program-specific issues

Each program review described the strengths and weakness of that particular program using the criteria in the self-study guide. These are outlined in the individual program reviews that follow this section on overall recommendations. Two issues not addressed in our 10 recommendations

are noteworthy. First, all of the programs struggled with underfunding; most budgets have not increased in several years. Thus many meritorious proposals had to be denied simply because there was not enough money. A big data multidisciplinary program like RIDIR is particularly hard hit because it competes with multidisciplinary efforts in the Computer and Information Sciences and Engineering (CISE) and Engineering (ENG) Directorates that connect data sciences with the social sciences and are much better funded. Second, the multidisciplinary breadth of some programs was not as robust as it could be. SPRF seemed to still favor psychology and SciSIP is underrepresented in proposals by political scientists and sociologists studying political and social institutions. SciSIP also disproportionately concentrated on creating datasets in lieu of transformative research on investments in science and science policy. It is evident that all of the programs need more money, avenues for collaboration with other directorates, and help with fulfilling their broad interdisciplinary mandate.

Progress since 2015

The previous COV met in 2015 to review these programs from 2010-2015. The COV review was very thorough as was the response from SBE leadership. Some of the issues addressed in our report were present in 2015 and continue as challenges. NSF still does not have a rigorous system for gathering demographic data on proposal applicants. This weakness significantly undermines all programs' efforts to track broadening participation of members of underrepresented groups at both the proposal application and award stages. Concerns were raised in the prior review, as in ours, about the quality of reviews. Some training opportunities were implemented, but this needs more careful study. Calibration exercises for reviewers were also recommended before, but it is not clear that they were implemented. Variance in the quality of reviews and the panel structure itself need systematic evaluation. Concerns were raised in 2015 about the need to better track outcomes of funded projects. The response suggested that NSF was setting up a new Evaluation and Assessment Capability (EAC) section in the Office of Integrative Activities. No mention was made of that entity in any of the discussions or reviews.

On the more positive side, SBE and NSF more broadly have been proactive in fostering collaborations with MSIs and members of underrepresented groups. We have the HBCU program, the new Build and Broaden initiative, and NSF INCLUDES; all are strong mechanisms for identifying promising collaborators from underrepresented groups. SPRF now has two tracks -- FR and BP which replaced the Minority Fellowship Program. The current COV viewed this as a good change, although the BP still lags behind in applications from members of underrepresented groups.

COV Membership

Appointment Type	Name	Affiliation
COV Chair	Sandra Graham	University of California, Los Angeles
COV Members:	Tucker Balch	Georgia Tech on sabbatical to JP Morgan Chase
	Margrit Betke	Boston University
	Margo Monteith	Purdue University
	Lisa Green	University of Massachusetts - Amherst
	Adam Jaffe	Brandies University and Motu Research
	Mary Frank Fox	Georgia Tech
	Bethany Usher	George Mason University
	Jerry Mitchell	University of South Carolina

INDIVIDUAL PROGRAM RECOMMENDATIONS

Research Experience for Undergraduates

Date of COV: July 13-15, 2020

Program/Cluster/Section: Research Experiences for Undergraduates Sites (REU Sites)

Divisions: SBE Office of Multidisciplinary Activities (SMA)

Directorate: Directorate for Social, Behavioral and Economic Sciences

Number of actions reviewed:

Awards: 4

Declinations: 10

Other:

Total number of actions within Program/Cluster/Division during period under review (SBE and SMA):

 Awards: 60

 Declinations: 164

Other:29 (Includes the non-merit reviewed actions of supplements of existing awards and forward funding actions)

Manner in which reviewed actions were selected:

A randomized sample of approximately 5 percent of proposals were selected from each program to be reviewed by the COV. First, proposals were sorted by fiscal year and then award number. Second, a randomized number generator tool[1] was utilized to select approximately 5 percent of proposals by fiscal year. Non-merit reviewed proposals and those returned without review were excluded for review and only the lead proposal for projects within a collaborative project proposal were included in the sampling exercise.

[1] **Random.org website**

STUDY QUESTIONS

INTEGRITY AND EFFICIENCY OF THE PROGRAM'S PROCESSES AND MANAGEMENT

Briefly discuss and provide comments for *each* relevant aspect of the program's review process and management. Comments should be based on a review of proposal actions (awards, declinations, and withdrawals) that were *completed within the past five fiscal years*. Provide comments for *each* program being reviewed and for those questions that are relevant to the program(s) under review. Quantitative information may be required for some questions. Constructive comments noting areas in need of improvement are encouraged.

I. Questions about the quality and effectiveness of the program's use of merit review process. Please answer the following questions about the effectiveness of the merit review process and provide comments or concerns in the space below the question.

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Are the review methods (for example, panel, ad hoc, site visits) appropriate? 1. The review methods are appropriate; however, we make the following recommendations to strengthen the process. 2. Recommendations: 3. The full panel is often not represented in the summary review of the proposals. Typically, only the three reviewers participate in the discussion. The process would be stronger if all members of the panel participated in the discussion and contributed to the summary and placement of the proposal. This change in process could increase the workload and require smaller, more focused panels, or a longer time to support the reviews. 4. Schedule site visits for the Program Officer and staff. These visits would give insight for the Program Officer and for the staff into the value of programs and their best practices. Including staff here may also promote ownership in the program (e.g. the ability to see the outcomes for their efforts supporting the program's administration).	YES

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
5. Sample reviews from previous panels are sent to new reviewers as a guide. High-quality panel summaries should be included for all potential panel scribes. 6. Assign the scribe role to a panelist who was not a reviewer to include more voices and develop a stronger consensus from a larger set of the panel. 7. *Ad hoc/mail reviewers should be used when deemed necessary by the Program Officer. These reviews should be complete prior to the panel review so the ad hoc review can be part of the complete review package.*	

2. Are both merit review criteria addressed:
 	
The merit review criteria are especially appropriate for the NSF REU program. REUs balance intellectual merit and the broader impacts (especially for the students). The best REUs have strong research projects that transform science, and provide an opportunity to recruit and train a diverse new generation of students.
 	
The reviews have gotten stronger over time, with reviewers addressing the additional REU review criteria. This provides better feedback to the PIs.
 	
   a) In individual reviews?
 	
For the most part, yes, but there is significant variance in the quality and level of detail in the individual reviews.
 	
For example, this is a typical sentence in many reviews: "On a level of broader impact, the project has the potential to be transformative for participants." In what way can the project be transformative? The proposer would have very little guidance here on how to improve their project.
 	
   b) In panel summaries?
 	
The panel summaries are often weak on details. Greater specificity on ways to improve proposals would make these summaries more likely to help PIs have successful future proposals.
 	
   c) In Program Officer review analyses?
 	
The Program Officer Review Analysis was often stronger and provided support for the funding rationale/decision.
	YES

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
Recommendation: The Program Officer Review Analysis is well-written and should be shared as Program Officer notes to the PIs. Sharing this information only with NSF directors is a poor use of this valuable resource.	
3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals? We have examined the reviews from the perspective of PIs, either who are going forward with a project but still will benefit from constructive comments from the panel, and one who was rejected and was interested in strengthening the proposal for a future re-submission. We also looked as outsiders who wanted a summary of the proposal itself, and the reasons for supporting or not supporting the proposal. There is significant variation in the quality of the individual reviews. Many of the reviews give comments about the strengths and weaknesses of the proposal, but far fewer give substantive support for future improvement. The Program Office has instituted individual training for new reviewers and panelist that has increased the quality of the individual reviews. In the REU program, there are additional criteria, and more of these criteria have been specifically addressed individual reviews recently because of the PO has emphasized giving this feedback. Recommendation: The program officer should provide additional training to ensure that individual reviews, including sharing model panel summaries contain not only critiques and praise, but offer suggestions for improvement.	YES
4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)? The panel summaries often do not consistently articulate why the panel reached consensus, although they do address situations where there are outliers in the scoring.	NO
5. Does the documentation in the jacket provide the rationale for the award/decline decision?	YES

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
The jacket documentation provides a clearer rationale for award decisions as the program officer review analysis is available. This often gives the clearest explanation as to why a proposal was funded or not.	
6. Does the documentation to the PI provide the rationale for the award/decline decision? The PI documentation usually includes a context statement, individual reviews, and a panel summary. The program officer may provide additional information to explain the basis for declining a proposal. The COV reviewers note that not all reviews contain information about how to improve proposals to be competitive in future submissions or how to improve the research and student experience for on-going projects not supported by NSF. Recommendation: Include the program officer review analysis to help the PI improve their work.	YES
1) Additional comments on the quality and effectiveness of the program's use of merit review process: Overall, the reviewers and panelists are understanding and addressing the Merit Review criteria. The highest-quality proposals that meet both criteria and have the highest opportunity for success are funded.	

II.   **Questions concerning the selection of reviewers. Please answer the following questions about the selection of reviewers and provide comments or concerns in the space below the question.**

SELECTION OF REVIEWERS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Did the program make use of reviewers having appropriate expertise and/or qualifications? Most proposals have at least two reviewers with subject area expertise in the field of the proposer or a closely allied field. Panels also appear to include at least one reviewer with experience with REU sites as an investigator and/or with undergraduate research. The PO is thoughtful in her selection of ad hoc reviewers and panelists, and tailors the panel to the portfolio of proposals. The REUs projects are often interdisciplinary, and support the broad inclusion of students. The panelists appear to be diverse, by field and background, and type of institution. The panel's demographic diversity (gender, ethnicity, etc.) is not clearly known.	YES
2. Did the program recognize and resolve conflicts of interest when appropriate? Sufficient efforts have been made to identify COIs and resolve them at multiple stages in the review process. These efforts include not only institutional and financial conflicts, but also personal (former student, antagonisms, etc.) where known.	YES
3. Additional comments on reviewer selection: None	

**III.   Questions concerning the management of the program under review.
Please comment on the following:**

MANAGEMENT OF THE PROGRAM UNDER REVIEW
1. Management of the program.
The REU program appears to be thoughtfully run and focused on a diversity (demographic, geographic) of reviewers and awards. The program budget has remained largely the same and is thus strained as this does not recognize the increased costs PIs have had over time to conduct high-quality REU programming.
The current leadership is invested in seeing that the program is successful despite having a high workload by also managing another program. The program officer has created outreach programs to increase the number of proposals from minority-serving institutions and 4-year colleges.
Recommendations:
1. Enhancing the outreach budget to encourage quality submissions from institutions that have not submitted previously.
2. Provide additional administrative support to grow the number of REU sites supported.
Responsiveness of the program to emerging research and education opportunities.
Over the past 20 years the number of students encouraged to participate in undergraduate research has grown with a concomitant increase in the number of students applying to REU programs.
Recommendations:
1. A funding increase to the REU program to increase the number of student spots available and to grow the number of REU sites.
2. Program planning and prioritization process (internal and external) that guided the development of the portfolio.
The program officer successfully meshes the goals of the REU program with that of the SBE Office of Multidisciplinary Activities to create a portfolio that reflects the breadth of the disciplines and promotes multidisciplinary programs. The program officer prioritizes inclusiveness and innovation in building the portfolio.
3. Responsiveness of program to previous COV comments and recommendations.
While we can comment on NSF responsiveness to some of the previous recommendations, NSF did not share equivalent data to respond fully here.
Recommendations:
1. The 2015 COV recommendation for reviewer training has been partially addressed. The program officer has implemented new reviewer training and shares previous reviews with panelists to highlight appropriate review

MANAGEMENT OF THE PROGRAM UNDER REVIEW
expectations. We continue to support the earlier recommendation for formalized reviewer training. 2. The 2015 COV report stated, "Reviewer and PI data should be better collected or reported. For example, currently the information provided only presents primary PI and not coPI characteristics." We concur, and would encourage NSF to provide more information about the representativeness of Co-PIs and the reviewers. 3. The 2015 COV report stated, "More systematic data should be collected and provided on program outcomes and participants productivity." We enthusiastically concur and address below.

IV. **Questions about Portfolio. Please answer the following about the portfolio of awards made by the program under review.**

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity? This COV review group did not see enough information to assess this. The eJacket contained only the awards made in the REU program, and not the full list of submitted proposals. The proposals are not tagged by discipline for analysis. Simply pulling out the home department of the PIs would be helpful. A non-systematic perusal appears to show that the proposals are distributed across SBE disciplines, and has a significant number of inter-disciplinary projects (both within SBE and across directorates, as seen in the co-funding table). Recommendation: 1. To foster multi-disciplinary projects, each program officer should have a budget set-aside for partnered work. Current practice is for program offers to make requests of other programs with a less than certain outcome for joint support. This uncertainty could be lessened with a common and dedicated set-aside for collaborations.	DATA NOT AVAILABLE
2. Are awards appropriate in size and duration for the scope of the projects? The solicitation indicates that most awards are for 3 years (up to five years are possible) so the duration is in line with that recommendation. $80-130K is the typical annual funding amount, so it is not surprising that $300K is the average. The program officer indicates that the demand for the program, especially from students, is increasing. The average size of the awards has not increased over the 5 years of review. The new REU Solicitation will increase the size of each award (which is needed to keep up with the cost of living), but the budget of the overall REU program has not increased. This will decrease the number of projects that can be funded with the current budget. We recommend: 1. Increase the budget of the SBE REU program to offer additional awards and to meet the needs of student costs.	APPROPRIATE

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
2. Does the program portfolio include awards for projects that are innovative or potentially transformative? The reviews of the programs show that the Intellectual Merit of the funded proposals is universally excellent, with average scores above 4.0 every year. The proposals are exciting and expose students to authentic research each year.	APPROPRIATE
1) Does the program portfolio include inter- and multi-disciplinary projects? See our note in IV 1 above. The projects span SBE disciplines, and the co-funding between SBE REU and other directorates demonstrate a commitment to interdisciplinarity. However, SBE REU preferentially partners with a few directorates, and there could be an effort to span other disciplines.	APPROPRIATE
4. Does the program portfolio have an appropriate geographical distribution of Principal Investigators? The map needs to be completely redone, as it is unclear what it is trying to convey. The legends to do not match what is visualized. Maps with proportional symbols for each state's data are more appropriate. Here is a choropleth map version that includes data ranges too diverse to be meaningful (there are errors, too). Although there has been at least 1 submission from every state in the past 5 years, there are 15 states that have not had awards. Given that these states vary from Massachusetts to Mississippi, there are probably several different reasons they are not represented (not all are EPSCoR states). The program could conduct outreach to encourage participation from institutions in these states. The COV reviewers note that we are only seeing data on the PI, and not the Co-PIs. Given the potential for collaboration across institutions, this information could not be representative of the geographical impact of the projects. The bigger question, though, is whether the SBE REU program offers research opportunities to students from all geographical areas. In other words, one state may have an award, but students from other states are also participating in that program. These data are not provided. Recommendations:	NOT APPROPRIATE

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
1. Include the Co-PI's geographic information. 2. Collect location data on program participants. 3. Recreate the maps with the assistance of a geographer to more explicitly highlight spatial patterns of interest in the data.	
5. Does the program portfolio have an appropriate balance of awards to different types of institutions? There are very few submissions from public 4-year institutions. The percentage of submissions and awards (51%, 13% respectively) is highest from public PhD research institutions. To assess broadening participation, the reviewers looked for information about awards to Minority Serving Institutions, but these data are not separately available. The bigger question, however, is whether these sites have student participants from institutions that have not traditionally had research experiences available. Recommendations: 1. The REU program must collect data on the home institution for each student participant. 2. NSF and the REU program should collect data on the number and success rates of proposals from MSIs.	APPROPRIATE
6. Does the program portfolio have an appropriate balance of awards to new and early-career investigators? Early career PIs were funded 21% of the time, while experienced PIs were funded 35% of the time. There is no information about whether this is the first NSF support for these PIs. Over 50% of the proposals were from early career faculty. Recommendation: 1. The REU program can also encourage submissions from more experienced PIs. 2. The SBE REU program could run (or collaborate with organizations like the Council on Undergraduate Research) proposal writing workshops for new and early-career investigators.	APPROPRIATE

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
7. Does the program portfolio include projects that integrate research and education? The COV reviewers looked at this from two perspectives. The first is that the REU program itself is designed to include undergraduate students in authentic scientific research, and the program successfully funds proposals that are balanced in Intellectual Merit and Broader Impacts. In the second perspective, it is less clear that the SBE REU programs are providing education to the public as a broader impact specifically. For example, a REU site could produce K-12 curriculum materials as part of its deliverables. That effort engages research and education to a much different degree than "education by exposure" as this question seems to indicate. Some proposals included teachers as participants through the RET program, but that data is not quantified.	APPROPRIATE
8. Does the program portfolio have appropriate participation of underrepresented groups[1]? In the first three years of review, all identified minority PI proposals were declined. In 2018, 75% were awarded (3/4) and in 2019, 50% (1/2) were awarded. There were no PIs who identified themselves as disabled. We also note that 30% of females were funded compared to 31% of males. Although the minority award success rate is a promising trend, the numbers of proposals with identified minority PIs is too small to draw significant conclusions. In addition, up to 30% of the proposals per year do not identify the minority status of the PI, making analysis more difficult. The bigger question, though, is whether the SBE REU program offers research opportunities to underrepresented students. These data are not provided. NSF cannot run programs meant to broaden the impacts and access to scientific research without collecting the necessary data to determine which initiatives support these goals. Recommendation: 1. We strongly encourage or require PIs/Co-PIs to identify individual characteristics. NSF can use the Fastlane interface to	APPROPRIATE

[1] NSF does not have the legal authority to require principal investigators or reviewers to provide demographic data. Since provision of such data is voluntary, the demographic data available are incomplete. This may make it difficult to answer this question for small programs. However, experience suggests that even with the limited data available, COVs are able to provide a meaningful response to this question for most programs.

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
make it more difficult to avoid answering the questions, while staying within the legal requirements. 2. The REU program must collect data on the gender and minority status of student applicants and participants.	
9. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports. National priorities can be fluid, but SBE and the SBE REU program stay focused on the NSF mission of "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense..." (NSF At a Glance on NSF.gov). The SBE REU program clearly has a commitment to the NSF's 10 Big Ideas with interdisciplinary projects.	APPROPRIATE
10. Additional comments on the quality of the projects or the balance of the portfolio: The SBE REU has been intentional in funding a portfolio of high-quality projects, although additional data could help NSF understand how to strengthen the programs in the future.	

Summary recommendation: The Research Experience for Undergraduates program should be conducting research on the effectiveness of the program in creating the next generation of scientists.

> "Research experience is one of the most effective avenues for attracting students to and retaining them in science and engineering, and for preparing them for careers in these fields. The REU program, through both Sites and Supplements, aims to provide appropriate and valuable educational experiences for undergraduate students through participation in research. REU projects involve students in meaningful ways in ongoing research programs or in research projects specifically designed for the REU program. REU projects feature high-quality interaction of students with faculty and/or other research mentors and access to appropriate facilities and professional development opportunities.

> REU projects offer an opportunity to tap the nation's diverse student talent pool and broaden participation in science and engineering. NSF is particularly interested in increasing the numbers of women, underrepresented minorities, and persons with disabilities in research."

> (Research Experiences for Undergraduates, Sites and Supplements Program Solicitation, NSF 13-542, retrieved July 15, 2020).

To determine the effectiveness of the REU program in preparing the next generation of scientists, and to improve the program to help meet this goal, the NSF REU program is encouraged to create a program-wide robust assessment initiative. The National Academy of Sciences has recently demonstrated that there has not been systematic assessment of the value of undergraduate research (Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities), and NSF is positioned to study the effectiveness of the premier national program. This assessment should include two types of data collection and analysis:

1. NSF should collect data about the characteristics of the students who apply to and participate in REU programs, including gender, minority-status, home institution, discipline, etc., to determine if the program is encouraging participation by diverse and under-represented students.

NSF REU should include student outcomes assessment, including student self-assessment (such as the Evaluate UR program from Buffalo State, which was developed with NSF support, http://serc.carleton.edu/).

Resource Implementations for Data Intensive Research in the Social, Behavioral and Economic Sciences (RIDIR).

Date of COV: July 13-15, 2020

Program/Cluster/Section: RIDIR

Divisions: SBE Office of Multidisciplinary Activities (SMA)

Directorate: Directorate for Social, Behavioral and Economic Sciences

Number of actions reviewed:

Awards: 3

Declinations: 13

Other:

Total number of actions within Program/Cluster/Division during period under review (SBE and SMA):

 Awards: 32

 Declinations: 296

Other: (Includes the non-merit reviewed actions of supplements of existing awards and forward funding actions)

Manner in which reviewed actions were selected:

A randomized sample of approximately 5 percent of proposals were selected from each program to be reviewed by the COV. First, proposals were sorted by fiscal year and then award number. Second, a randomized number generator tool[1] was utilized to select approximately 5 percent of proposals by fiscal year. Non-merit reviewed proposals and those returned without review were excluded for review and only the lead proposal for projects within a collaborative project proposal were included in the sampling exercise.

[1] www.random.org

V. I. Questions about the quality and effectiveness of the program's use of merit review process. Please answer the following questions about the

effectiveness of the merit review process and provide comments or concerns in the space below the question.

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?	Yes

2. Are both merit review criteria addressed
 	
    a) In individual reviews?
 	
    b) In panel summaries?
 	
    c) In Program Officer review analyses?
	Yes

3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals?	Yes
4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?	Yes
5. Does the documentation in the jacket provide the rationale for the award/decline decision?	Yes

6. Does the documentation to the PI provide the rationale for the award/decline decision?	Yes
7. Additional comments on the quality and effectiveness of the program's use of merit review process: The COV discussed panel versus ad hoc reviews and felt the choice for panels made sense for the RIDIR program. However, there are special considerations with respect to a multidisciplinary program (see below in Program Management section III 2.).	

VI. **II. Questions concerning the selection of reviewers. Please answer the following questions about the selection of reviewers and provide comments or concerns in the space below the question.**

SELECTION OF REVIEWERS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Did the program make use of reviewers having appropriate expertise and/or qualifications?	Yes (but see below item 3)
2. Did the program recognize and resolve conflicts of interest when appropriate?	Data not available
3. Additional comments on reviewer selection: There is not always a domain expert on the panel. However, we acknowledge that panels cannot handle every domain combination possible. Program managers described the difficulty of finding appropriate reviewers for the panels.	

VII.   **III.** **Questions concerning the management of the program under review. Please comment on the following:**

MANAGEMENT OF THE PROGRAM UNDER REVIEW

1. Management of the program.

The program may have suffered because it did not have a dedicated manager responsible for the program over an extended period of time. Instead, several program managers had to split their time with this program and other programs they were responsible for. Because this program and its follow-up replacement versions (HNDS-I and HNDS-R) are very important not just for SBE but NSF's mission overall, the COV applauds that the past management approach has recently changed with the hiring of a new program manager whose main responsibility will be the multidisciplinary programs that support research in social, behavioral, and economic sciences by building infrastructure and by providing with tools from data science and infrastructures.

2. Responsiveness of the program to emerging research and education opportunities.

The COV notes that there were many highly competitive proposals submitted to RIDIR that were not funded due to funding limitations. The RIDIR program had (and now HNDS has) a very limited budget (~$4.5 million annually) compared to programs that support multidisciplinary efforts in the CISE or ENG Directorates that connect data science and engineering with social sciences. The COV recommends that NSF works out better structures across Directorates to support multidisciplinary research.

There are many emerging research and education opportunities in the social, behavioral, and economic sciences due to recent technical developments that enable unprecedented database creations and data analytics. These include new technical tools from (1) machine learning and AI, and (2) data management and privacy.   To be able to fund such opportunities, the RIDIR program needs more funding.

Another issue is bringing competitive proposals to a fundable level.   It is important that truly interdisciplinary, high-quality research proposals that involve both social & data/computing sciences are not rejected because reviewers with social science expertise expect fundamental research outcomes in the social sciences, and reviewers with a computing background expect fundamental research outcomes in computing. There must be a place where interdisciplinary research that involves applied computing and social sciences can be funded.

3. Program planning and prioritization process (internal and external) that guided the development of the portfolio.

The RIDIR program had (and now HNDS has) a very limited budget (~$4.5 million annually), compared to programs that support multidisciplinary efforts in the CISE or ENG Directorates that connect data science with social sciences.

4. Responsiveness of program to previous COV comments and recommendations.

We concur with the previous COV recommendation "best to avoid ad hoc reviewers who won't be part of the panel conversation, though of course that may be necessary at times." The program managers for RIDIR have followed this recommendation.

The previous COV mentioned that "drop-in proposals" were funded, which the committee defined as proposals by scholars in another area who see a potential fit for their work in the program. The current COV did not see such drop-in proposals that were funded. There were some proposals by PIs with expertise outside the social sciences that were declined because of weaknesses in the intellectual merits, in particular, a lack of depth with regards to the research questions in the social sciences.

VIII. **IV. Questions about Portfolio. Please answer the following about the portfolio of awards made by the program under review.**

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity? Yes, but the co-funding levels between SBE and ENG, and SBE and CISE are not appropriate. The funding rate for the new HNDS program could likely be increased if co-funding was actively sought.	Yes
2. Are awards appropriate in size and duration for the scope of the projects? However, we recommend considering a mix of award sizes for the new HNDS program	Yes

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
3. Does the program portfolio include awards for projects that are innovative or potentially transformative? We point to the proposal "1539129: Collaborative Research: Enabling Access to and Analysis of Shared Daylong Child and Family Audio Data"	Yes
4. Does the program portfolio include inter- and multi-disciplinary projects?	Yes
5. Does the program portfolio have an appropriate geographical distribution of Principal Investigators?	Yes
6. Does the program portfolio have an appropriate balance of awards to different types of institutions? No, however, the kind of research supported by the RIDIR program is difficult to conduct by higher education institutions that are not R1 institutes. Perhaps community building efforts could enable smaller universities to become involved in team building earlier so they could join larger proposals.	No.

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
7. Does the program portfolio have an appropriate balance of awards to new and early-career investigators? The rate new/prior proposers are 37%/63%. The funding rate for new PIs is 8.9% vs. 10% for prior PIs.	Yes.
8. Does the program portfolio include projects that integrate research and education? Education is typically integrated by training graduate and undergraduate research assistants.	Yes.
9. Does the program portfolio have appropriate participation of underrepresented groups? Efforts should be made to solicit proposals by members of underrepresented groups. One way to do that is to support team building through workshops in advance of program calls. And also, to more aggressively involve underrepresented groups in the review process.	No.
10. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.	Yes.

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
11. Additional comments on the quality of the projects or the balance of the portfolio: The quality of the proposals was very high, and it is important for SBE to reach out to other Directorates, who have multidisciplinary programs for data infrastructure building, for co-funding (CISE & ENG).	

Summary of recommendations:

Stronger collaboration between directorates (SBE / CISE, SBE / ENG) in order to fund more projects. Many very competitive proposals are not funded.

More assertive measures to involve HBCUs and other MSIs. Pre-solicitation workshops (e.g., multidisciplinary NSF ENG Future of Work Workshop organized by Laurell Smith-Doerr)?

The use of a mix of award sizes might be beneficial for the new HNDS program.

Having a dedicated Program Manager is important.

SBE Post-Doctoral Fellowship Program (SPRF)

Date of COV: July 13-15, 2020

Program/Cluster/Section: SBE Postdoctoral Research Fellowships (SPRF)

Division: SBE Office of Multidisciplinary Activities (SMA)

Directorate: Directorate for Social, Behavioral and Economics Sciences

Number of actions reviewed:

Awards: 4

Declinations: 24

Other

Total number of actions within Program/Cluster/Division during period under review:

Awards: 128

 Declinations: 411

Other: 13 (Includes the non-merit reviewed actions of supplements of existing awards and forward funding actions)

Manner in which reviewed actions were selected:

A randomized sample of approximately 5 percent of proposals were selected from each

program to be reviewed by the COV. First, proposals were organized by fiscal year and then award number. Second, a randomized number generator tool[1] was utilized to select approximately 5 percent of proposals by fiscal year. Non-competitive proposals and those returned without reviews were excluded for review and only the lead proposals for projects within a collaborative project proposal were included in the sampling exercise.

[1] **www.random.org**

I. **Questions about the quality and effectiveness of the program's use of merit review process. Please answer the following questions about the effectiveness of the merit review process and provide comments or concerns in the space below the question.**

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
I. Are the review methods (for example, panel, ad hoc, site visits) appropriate?	YES
I. Are both merit review criteria addressed d) In individual reviews? e) In panel summaries? **f)** In Program Officer review analyses?	YES
I. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals? Comments: Occasionally reviews lack enough detail, providing insufficient explicit discussion or mention of intellectual merit, broader impacts, or Track-relevant evaluation. Reviewers sometimes noted certain broader impacts that the researcher did not happen to address, even though other broader impacts were adequately addressed in the proposal. Given proposals need not address all broader impacts in all possible ways, whether the broader impacts were actually insufficiently addressed was sometimes not clear. Summary statements sometimes do not seem to match the rating (e.g., rating good when no weaknesses are noted in the summary statement) or do not state the recommendation specifically. We **recommend** additional reviewer training to increase the likelihood that these problems are corrected.	YES

7. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)? Comments: Sometimes even after panel discussion the ratings ranged across the board (from F to E). We noticed inconsistencies in panel summary (e.g., one case broader impacts were discussed in one part, then in the same panel statement, the summary statement noted that "broader impacts were strong"). We **recommend** relevant training for preparing the panel summaries and consideration of who should prepare the panel summaries.	YES
5. Does the documentation in the jacket provide the rationale for the award/decline decision? Comments: Documentation in the jacket usually includes a context statement, individual reviews, panel summary (if applicable), site visit reports (if applicable), program officer review analysis, and staff diary notes.]	YES
6. Does the documentation to the PI provide the rationale for the award/decline decision? Comments: Documentation to PI usually includes context statement, individual reviews, panel summary (if applicable), site visit reports (if applicable), and, if not otherwise provided in the panel summary, an explanation from the program officer (written in the PO Comments field or emailed with a copy in the jacket, or telephoned with a diary note in the jacket) of the basis for a declination.]	
Comments: In the BP program, compared to the FR track, there was inconsistency across reviews in the level of detail regarding strengths and weaknesses. Sometimes there was no explicit evaluation provided with respect to Broadening Participation. In some cases, the Broader Impacts and Broadening Participation categories were lumped into one. It might be the case that there is overlap between the two; however, this is not always the case and each should be evaluated individually. For instance, if the PI is from an underrepresented group, that should be sufficient for BP. We **recommend** that reviewer training clearly distinguish between broader impacts and broadening participation, and that reviewers should address each individually.	

II. Questions concerning the selection of reviewers. Please answer the following questions about the selection of reviewers and provide comments or concerns in the space below the question**.**

SELECTION OF REVIEWERS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Did the program make use of reviewers having appropriate expertise and/or qualifications? Comments: In additional to panel reviewers, the Program Officer regularly solicits and obtains ad hoc reviewers to ensure that the appropriate expertise is always available. This is a strength. The PO also makes efforts to recruit panelists and ad hoc reviewers from minority-serving institutions.	YES
2. Did the program recognize and resolve conflicts of interest when appropriate?	YES
Additional Comments: We appreciated the PO's self-reported efforts to achieve breadth across institutions in the representation of the reviewers (e.g., small, large, PWI, HBCU). We **recommend** that data be collected and provided to document the extent of breadth.	

III. Questions concerning the management of the program under review. Please comment on the following:

MANAGEMENT OF THE PROGRAM UNDER REVIEW
1. Management of the program. Although relevant data were not provided the COV, the PO's description during the COV meeting suggested an impressive job of managing the panels and ad hoc reviewers for the broad range of disciplines, attending not only to disciplinary expertise but also to institutional and demographic diversity. We **recommend** that that the COV be provided with relevant summary data for all proposals. In order to fund high quality proposals, starting in 2018 the PO has established alliances with other NSF directorates and programs to seek co-funding. This endeavor appears to have been highly successful in engaging other disciplines and programs with SBE. Judged by the above, the COV feels positively about the current management of the program.
2. Responsiveness of the program to emerging research and education opportunities. In 2017, the IBSS track was discontinued and replaced with the SPRF-FR and SPRF-BP programs. Given IBSS was not attracting many applicants and exactly what constituted interdisciplinary research was often unclear, this change is very positive and captures the goals of the SPRF program well. The BP track continues to focus on high quality research that supports diversity among researchers and research that focuses on broadening participation of underrepresented groups. The FR track supports research to build knowledge about human behavior in the SBE sciences. We conclude that the program is adequately comprehensive and responsive to emerging research and education opportunities. However, we are concerned with the relatively low (compared to FR) number of applications for the BP program. We **recommend** increased efforts at outreach at minority-serving institutions, professional conferences, societies and professional organizations. Regular availability of webinars for grant preparation is advisable, as well as outreach to relevant college and university graduate programs (e.g., Alliances for Graduate Education and the Professoriate, AGEP). We **recommend** outreach to not just potential applicants, but also to faculty mentors. Faculty can be encouraged to identify individuals who might consider a postdoctoral fellowship, and then encourage and support the proposal development.
3. Program planning and prioritization process (internal and external) that guided the development of the portfolio. The changes implemented in 2017 continue to capture the aim of the program: to prepare new PHDs for careers in the sciences. The program has been structured to keep the

components of interdisciplinarity given the engagement from other disciplines/directorates. We view this change positively.

The BP track continues to provide opportunities for minority fellowship applicants in particular, while increasing the scope of the track beyond the prior exclusive focus on minority applicants (i.e., in the previous Minority Postdoctoral Research Fellowship program) to include research that may not be conducted by minority scientists but that is relevant to broadening participation of underrepresented groups. We view this change positively, although we caution that close attention should be paid to whether those from underrepresented groups are receiving fellowship in sufficient numbers.

4. Responsiveness of program to previous COV comments and recommendations.

A Committee of Visitors (COV) convened on August 27-28, 2015, at the National ScienceFoundation to review the programs in the SBE Office of Multidisciplinary Activities (SMA).

The previous COV for SPRF recommended the following in connection with the BP track:

"We therefore recommend that the SPRF program officer or her designee extend the program's outreach activities to minority serving institutions to include grant writing workshops. One possible opportunity is to host regional grant writing training events targeted to institutions with graduate and postdoctoral training programs in SBE disciplines that are developed in collaboration with former awardees, postdoctoral program offices, and/or professional societies, such as AAAS or the National Postdoctoral Association. This would provide the training in the preparation of competitive proposals under the guidelines of this program and decrease a perceived barrier to submission."

We do not have information about follow-up for this recommendation, but we note that applications to the BP program continue to be lower than to the FR program. We **recommend** the activities identified by the previous COV, along with this COV's relevant recommendation (see section III, #2 above).

I. **Questions about Portfolio.** Please answer the following about the portfolio of awards made by the program under review.

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity? Comments: Although we indicated appropriate, sufficient data for evaluating this comprehensively across all submitted proposals were not available to us. We **recommend** that the COV be provided a table that summarizes discipline breakdown for all submitted proposals. We also **recommend** that efforts be made to increase the number of proposals from a variety of disciplines for the BP track.	Appropriate
2. Are awards appropriate in size and duration for the scope of the projects?	Appropriate
3. Does the program portfolio include awards for projects that are innovative or potentially transformative?	Appropriate
4. Does the program portfolio include inter- and multi-disciplinary projects?	Appropriate
5. Does the program portfolio have an appropriate geographical distribution of Principal Investigators?	Appropriate
6. Does the program portfolio have an appropriate balance of awards to different types of institutions? Comment: Institutional data are not available for 2018 and 2019 because of the change to providing awards to individuals rather than to institutions. We **recommend** providing information about the institutions where applicants would be pursuing research, and institutional data indicating where applicants received their PhDs.	Appropriate

7. Does the program portfolio have an appropriate balance of awards to new and early-career investigators? Comments: NOTE: A new investigator is an individual who has not served as the PI or Co-PI on any award from NSF (with the exception of doctoral dissertation awards, graduate or post-doctoral fellowships, research planning grants, or conferences, symposia and workshop grants.) An early-career investigator is defined as someone within seven years of receiving his or her last degree at the time of the award. Comments: Almost all investigators are new for these postdoctoral fellowship applications.	Appropriate
8. Does the program portfolio include projects that integrate research and education? (to extent that we can tell based on proposals in our jackets)	Appropriate
9. Does the program portfolio have appropriate participation of underrepresented groups[2]? We see some participation of underrepresented groups. However, given the number of applicants reporting "unknown" and the lack of clarity in tables provided to us (e.g., years are not specified for Tables 11 and 12), gauging the appropriateness of participation is difficult. We **recommend** taking steps to increase PIs' reporting of gender and race/ethnicity in their FastLane profile by reminding PIs to check that their profiles are up-to-date, emphasizing how the data will be used, emphasizing the importance of having this demographic information, and indicating that the information will not be linked to any specific proposals PIs may submit during the evaluation and award process.	Data not sufficient
10. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports. Comments: Relevant to national priorities and agency mission, the program seeks to train future STEM scientists, with particular attention to broadening participation of underrepresented minorities and women in the BP track. Many STEM disciplines and industries recognize the need for increased representation of underrepresented groups and women, so the BP track in particular provides opportunities for building pipelines and increasing representation. In addition, the program recognizes the importance of mentoring.	Appropriate
11. Additional comments on the quality of the projects or the balance of the portfolio:	

Science of Science and Innovation Policy (SciSIP)

Date of COV: July 13-15, 2020
Program/Cluster/Section: Science of Science and Innovation Policy (SciSIP)
Divisions: SBE Office of Multidisciplinary Activities (SMA)
Directorate: Directorate for Social, Behavioral and Economic Sciences
Number of actions reviewed: 24 **Awards: 3** **Declinations: 21** **Other:**
Total number of actions within Program/Cluster/Division during period under review (SBE and SMA):   **Awards: 189**   **Declinations: 338** **Other: 23 (Includes the non-merit reviewed actions of supplements of existing awards and forward funding actions)**
Manner in which reviewed actions were selected: **A randomized sample of approximately 5 percent of proposals were selected from each program to be reviewed by the COV. First, proposals were sorted by fiscal year and then award number. Second, a randomized number generator tool[1] was utilized to select approximately 5 percent of proposals by fiscal year. Non-merit reviewed proposals and those returned without review were excluded for review and only the lead proposal for projects within a collaborative project proposal were included in the sampling exercise.** **[1] www.random.org**

I. Questions about the quality and effectiveness of the program's use of merit review process. Please answer the following questions about the effectiveness of the merit review process and provide comments or concerns in the space below the question.

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?	Overall, yes, but see answer to Question 7.
2. Are both merit review criteria addressed a) In individual reviews? b) In panel summaries? **c)** In Program Officer review analyses?	We note that scores appear to be less connected to broader implications than to intellectual merit.
3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals? Scope of comments varies quite a bit; some reviews are quite cursory. Overall, a fair degree of convergence exists in issues addressed. Comments do not always address consistently the policy implications.	
4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)? Panel comments add no apparent value. For proposals that were not discussed at the Panel, there are no Panel comments. When the number of reviews is very large (as many as 11 in one case we saw), there is almost inevitably a range of evaluations from the reviewers. It was not clear how the Panel dealt with this wide range.	

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
5. Does the documentation in the jacket provide the rationale for the award/decline decision? Comments: Documentation in the jacket usually includes a context statement, individual reviews, panel summary (if applicable), site visit reports (if applicable), program officer review analysis, and staff diary notes.] In those cases where the proposal was not discussed by the panel, the justification for non-funding is largely implicit in the relatively low scores of the reviews.I	
6. Does the documentation to the PI provide the rationale for the award/decline decision? Comments: Documentation to PI usually includes context statement, individual reviews, panel summary (if applicable), site visit reports (if applicable), and, if not otherwise provided in the panel summary, an explanation from the program officer (written in the PO Comments field or emailed with a copy in the jacket, or telephoned with a diary note in the jacket) of the basis for a declination.]	
7. Additional comments on the quality and effectiveness of the program's use of merit review process: Comment 1 There are several dimensions across which a surprisingly large number of the proposals were judged to be weak. These include: • Failure to clearly state the research goals; • Lack of congruence between research goals and research methods; • Absence of serious consideration of how claimed 'broader impacts' will actually be brought about. Recommendation: We respectfully suggest that these issues are sufficiently prevalent that the Program should consider whether it could strengthen its communications to potential applicants to make the need for these clearer. We understand that this program does not have a program-specific solicitation. Perhaps such a solicitation could be used to lay out guidelines covering these	

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE

common deficiencies. Better guidance would make the process more satisfying for applicants, reviewers, and the Program.

Comment 2

There is an inherent problem with having a few proposals with a large number of reviews. Empirically, proposals with more reviews will be less likely to be funded, simply because the chances of having one or more very negative reviews rises with the number of reviews. The Program should consider giving Panels some systematic instruction as to how to handle these situations to mitigate this bias.

Comment 3

The value-added of the Panel in these processes is far from clear. More generally, while participants in these programs often have strong beliefs as to the importance of the panels in the review process, no systematic empirical evidence exists on the panels' value-added. Further, research on committees suggests that panels may exacerbate implicit biases brought to the process by all participants. Finally, panels are costly and so their inclusion in the process can be justified if they result in decisions that are materially better than the decisions that would be taken without them.

Recommendation: we respectfully suggest that NSF should undertake an analysis of how panels operate, identifying systematically the cases where panels recommended funding decisions that differed from what would have been suggested by (e.g.) simple averaging of external scores, and then evaluating those differences in terms of gender and other biases, and ultimate outcomes.

Comment 4

The SciSIP Program has the epistemological distinction of being simultaneously a research funder, and a leading exemplar of the institutions that the program studies. One way it could advance the science of science and innovation would be to experiment as a program.

Recommendation: The program should consider varying key elements of its processes, and then retain and make available for research (subject to appropriate privacy protections) the outcomes of different approaches. Such experiments could include, for example, efforts to address the issues raised

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
above, by trying different forms of program guidance and/or different panel structures and then evaluating the outcomes.	

II. **II. Questions concerning the selection of reviewers. Please answer the following questions about the selection of reviewers and provide comments or concerns in the space below the question.**

SELECTION OF REVIEWERS	YES, NO, DATA NOT AVAILABLE, or NOT APPLICABLE
1. Did the program make use of reviewers having appropriate expertise and/or qualifications?	Reasonably well matched
2. Did the program recognize and resolve conflicts of interest when appropriate?	Yes
3. Additional comments on reviewer selection:	See comments at the end of Section I.

III.   III. Questions concerning the management of the program under review. Please comment on the following:

MANAGEMENT OF THE PROGRAM UNDER REVIEW
1. Management of the program.
We note that the program has never had a permanent Director. This creates issues for any program, as the operation and direction of the program suffer as a new Director gets up to speed. We believe that this problem is particularly acute for this program, because it relies significantly on co-funding of proposals. Co-funding is inevitably somewhat dependent on personal relationships with other Program Directors. Such relationships take time to establish, so that a series of rotating Directors cannot be fully effective. We also realize benefits to rotating Directors (see next question).   Recommendation: We respectfully suggest that SMA consider having both a rotating and a permanent program officer for this program.
2. Responsiveness of the program to emerging research and education opportunities.
The potential scope of this program is broad. Historically, it seems that new and emerging opportunities have been brought into the program, in part, by the rotating in of new directors.
3. Program planning and prioritization process (internal and external) that guided the development of the portfolio.
The program is intrinsically broad and somewhat ill-defined. It is not clear to us by what process differing topic areas have been emphasized over time.   One issue is that proposals concentrate, disproportionately, on creating datasets.   Recommendation: extended support of transformative research on complex processes that address understandings of investments in science, engineering, and technology—key to science policy.

4.   Responsiveness of program to previous COV comments and recommendations.
 	
     a. Demographic info appears to have been included.
     b. It does not appear that the program has made significant progress with respect to development of systems for eventually obtaining and tracking long-term outcome measures. This is a hugely important, ongoing issue.
 	
 Recommendation: As the premier U.S. science funding agency, it is incumbent on NSF to lead the way in developing a system for obtaining, tracking, and eventually evaluating long-term outcomes and impacts. Fundamentally, the 'Science of Science' cannot be undertaken without such a system, and this program should take the lead in developing it.
 	
     c. Systematic guidelines for reviewers: This appears to be a continuing issue.
     d. Infrastructure for large-scale data science: Not clear to us if/how this applies to this program
     e. Outreach for proposals from minority-serving institutions: Based on our small sample, the number of such proposals remains small, but we do not know what kind of outreach was undertaken.

f. Capacity and community building: Community building is a challenge given "no dominant disciplinary base, central conferences, and professional society" to advocate for the area (as the COV Study Guide points out, p. 56) However, such decentralization also affords the opportunity for breadth of participation and impacts—without disciplinary constraints on topics and publication venues, for example.

IV.   IV. Questions about Portfolio. Please answer the following about the portfolio of awards made by the program under review.

RESULTING PORTFOLIO OF AWARDS	APPROPRIATE, NOT APPROPRIATE, OR DATA NOT AVAILABLE
1) Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity? Within our small samples of proposals, there appears to be under representation of studies of political institutions, social institutions, and organizations. Recommendation: These would be expanded though wider participation of political scientists and sociologists. Business schools are well-represented.	
2) Are awards appropriate in size and duration for the scope of the projects? There is a significant increase in the average award size in the last two years, which we understand is the result of phasing out of small policy study awards that had previously been made. It appears that the typical award is now about $150K/year for 2 years. We have no reason to believe that this is not appropriate.	
3) Does the program portfolio include awards for projects that are innovative or potentially transformative? In our sample of projects, we note a significant proportion of data projects, which do not seem to be innovative or transformative. While we do not dispute the value of data construction, we note the tradeoff in funding such efforts, since they reduce the funding available for innovative and transformative research.	
4) Does the program portfolio include inter- and multi-disciplinary projects?	yes
5) Does the program portfolio have an appropriate geographical distribution of Principal Investigators?	Yes, across the span of regions
6) Does the program portfolio have an appropriate balance of awards to different types of institutions?	

Note: Awards occur mostly to doctoral granting institutions, which we think is expected. Public doctoral granting institutions seem to be somewhat more successful. None of this is surprising.	
7) Does the program portfolio have an appropriate balance of awards to new and early-career investigators? Comments: NOTE: A new investigator is an individual who has not served as the PI or Co-PI on any award from NSF (with the exception of doctoral dissertation awards, graduate or post-doctoral fellowships, research planning grants, or conferences, symposia and workshop grants.) An early-career investigator is defined as someone within seven years of receiving his or her last degree at the time of the award.	Approximately 40% new PIs seems reasonable.
8) Does the program portfolio include projects that integrate research and education? Several of the sample project proposals included undergraduate research participation.	
9) Does the program portfolio have appropriate participation of underrepresented groups[3]? Men submit more proposals than women. Women are slightly more successful in obtaining awards. In light of research on gender and performance in science, this is unsurprising. In the most recent years, the total number of awards made to men and women have been roughly equal. Minority applicants very few. No evidence this is program-specific.	
10) Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports. Examples are topics (independent of merit-scores) that include, but are not limited to: biomedical innovation; contribution of universities to commercial innovation; the effect of State-level education investment on research quality; distributional impacts of open data policies; space transportation systems; and transition to renewable energy.	Yes.
11) Additional comments on the quality of the projects or the balance of the portfolio:	

[3] NSF does not have the legal authority to require principal investigators or reviewers to provide demographic data. Since provision of such data is voluntary, the demographic data available are incomplete. This may make it difficult to answer this question for small programs. However, experience suggests that even with the limited data available, COVs are able to provide a meaningful response to this question for most programs.