SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT
SUBMISSION
NATIONAL SCIENCE FOUNDATION
EDUCATION AND TRAINING APPLICATION PILOT
OMB Clearance Number 3145-0248

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent universe and sampling methods

- ETAP data system

Universe of ETAP respondents:
  - Individual students (mostly undergraduate and graduate students) interested in applying for opportunities available through ETAP. These individuals will submit background information described above in Part A as part of their application.
  - Principal investigators (PIs) or their designees (such as co-PIs or administrators) and reference writers. These individuals will be asked to provide useful information. For example, PIs will be asked to provide information on the opportunities for prospective applicants and record admissions decisions and participation information (who actually participated in the program or opportunity).

Sampling methods. There is no sampling in ETAP as PIs need information from every applicant. Consequently, the system will collect basic information from the universe of individuals who submit applications. A justification for the estimated numbers of applicants is provided in section 12 of Part A of this request. Note that collecting this information from the universe of applicants does not add cost or burden to the government or respondents. Indeed, it should reduce cost and burden (on both PIs and applicants) by providing an online, centralized alternative to the applications currently offered by individual awards (see section 3 of Part A of this request for more details).

- **National student clearinghouse (NSC)**

Universe. We will obtain information on educational outcomes (enrollment and graduation) from the NSC for a former cohort of REU participants who applied through the REU data system, was admitted, and participated in the 2019 REU program at a pilot Site (N=471). Most of these students are expected to have graduated by 2023, when NSC data will be obtained.

Sampling methods. We will not use sampling given that the information we seek is available in a centralized location (the NSC) at zero burden for the participants and at low cost ($1 per student for the first 1,000 students, with marginal cost for additional records decreasing with volume). We also want to ensure adequate representation of groups that are traditionally underrepresented in STEM (minorities, students with disabilities, and others).

- **Employment Survey**

Universe. 471 undergraduates in the pilot who participated in the REU program for the 2019 program cycle.

Sampling methods. We will not use sampling because (1) this is a relatively small population and (2) we seek to obtain an accurate measure of the expect response rate for the future. Note that the logic of this pilot is to test a system that can be easily implemented by NSF in the future. The information collected—including a non-response bias analysis comparing survey data to NSC and registration/common application data—will enable the study team to make a recommendation to NSF regarding whether and how to sample in future data collection rounds, if this approach to data collection is adopted.

2. **Procedures for the collection of information**
   a. **Statistical methodology for stratification and sample selection**

The approach to scaling up the test of the system will be two-prong. First, we will target all NSF training programs offering research experiences, to enable the Agency to produce accurate estimates in one area of its education and training portfolio of investments as quickly as possible. Second, we will consult with NSF program officers to devise a plan for recruiting other types of training programs to test the system. We will assess programs against a number of criteria, including whether the program (1) has a data collection system in place (to prioritize those that do not), (2) is large in terms of the number of training opportunities funded (to have enough voluntary participation within a program to offer insights into how the system serves program needs), and (3) are representative of a line of training funding or o type of training recipient (to offer informative findings that would be useful for scaling the system to others). Within a program, the pilot will seek volunteers among PIs with active awards, collaborating with NSF program officers to devise the recruitment strategy.

   b. **Data collection**

- **ETAP data system.** This is a web-based system to which users will have access for different purposes and at different points in time. Applicants will use ETAP to complete their registrations or applications. They will learn about ETAP through the NSF website, PI websites (such as those set up for REU Sites), and/or guidance sent to prospective applicants by PIs. Data collection will begin in the Fall of 2021 to coincide with the earliest opening of applications for opportunities, as determined by PIs. Applications will be accepted within the dates specified by PIs.

- **NSC.** We will prepare and submit to the NSC a request for data on former participants.

- **Publications and Patents.** We will use NSF's Apache-Solr enterprise search engine to match participants to Web of Science, Scopus, and USPTO databases. These databases have already been ingested into NSF systems and are frequently updated.

- **Employment Survey.** In alignment with current best practices in data collection, we will use a multi-mode contact protocol to guide respondents to a web-based employment survey. We will attempt to reach participants via email, text (if they consent) and, if needed, postal mail to take them to a survey instrument that will be dynamically optimized for a range of screen sizes and devices (Dillman 2017; Finamore and Dillman 2013).

3. **Methods to maximize response rates and deal with nonresponse**

   The pilot is designed to determine the unit and item response rates expected using low-cost outreach strategies.

   - **ETAP data system.** Non-response is not a concern for the common application, as prospective applicants will be asked by PIs to submit their applications through the system. Non-response could be a concern, however, among noncompetitive Sites/Projects who will be asking prospective participants (undergraduate students, graduate students, and post-docs) to register in the system even though they have already been "accepted" (this would be the case for PIs who do not run a competition for admissions but rather invite individuals to participate, as is often the case for REU supplements). To encourage timely reporting, we will have in place the following mechanisms: 1) PIs cannot "certify" that the list of participants is complete if the student has not completed the registration, 2) PIs will receive automatic reminders through ETAP to provide missing participation information, and 3) NSF program officers will have the ability to check status of participant reporting. The pilot will test the success of this approach and elicit feedback to consider improvements.

   - **NSC.** Non-response is not a concern, but non-matches is. The pilot will expand testing of the feasibility of obtaining educational outcomes information on NSF participants through the NSC.

   - **Employment Survey.** Non-response is a concern for this survey, particularly because it will be administered several years after participation. As described earlier, this pilot will including testing of alternative low-cost strategies that show promise in the survey literature.

4. **Tests of Procedures or Methods to be Undertaken**

   - **ETAP data system.** For new system development, we will conduct two types of pre-tests—alpha testing to ensure the system is working properly, and beta- or usability-testing to ensure the content works as intended, that is, items are eliciting the responses we seek. Specifically:

     Alpha-testing. The study team will conduct testing of the web-based system to ensure that 1) the site functions properly, 2) the user interface is friendly and clear, and 3) the business logic aligns with specifications. Site functionality testing will involve testing the flow of the site between pages, verifying the ability to enter and retrieve appropriate information based on user role, and ensuring that the data collection capabilities of the site work properly and securely with website's attached database. Site user interface testing ensures

that the site's buttons, tabs, scroll bars, text boxes and other inputs are laid out in an intuitive way and function properly. Business logic verification ensures that all text content and status reports are aligned with the business needs for the instrument and the specifications which document those business needs. The goal of these tests is to ensure that the application works properly before we move to the stage of beta-testing.

Beta- or usability-testing. We will pre-test aspects of the application that have not previously been tested with the precursor system (such as a significantly revised order of items and user flow of the applicant module, and enhancements to the PI module to access data from prior years). Changes are modeled after the NSF GRFP common application (OMB control number 3145-0023) and industry leading applications such as the Commonapp for college, and based on feedback gathered through the help desk and other interactions with stakeholders (such as presentation at PI conferences and during system demonstrations webinars).

- **NSC, Web of Science, Scopus, and USPTO.** Not applicable as, for many years, these providers have been collecting and providing the data needed.

- **Employment survey**. Pretesting the web-based employment survey is vital to the integrity of the data collection. Alpha testing will help ensure that the survey behaves as intended. Beta-testing will enable us to assess understanding of the terms and questions presented in the survey, accuracy and relevance of the questions, whether important questions are missing or existing ones should be revised, and the length of time the survey takes to complete.

5. **Individuals consulted on statistical aspects of the design**

Andrew Weiss, vice president and director of Mathematica's Survey Operations Center, is an expert in data collection systems, research operations, and survey methodology.

Dr. Michael Sinclair, a statistician who was formerly with the U.S. Department of Justice and Census Bureau and has extensive expertise in areas such as sampling methodology, survey weighting and imputation, and data linkage, oversaw the original design for the REU data system pilot that is the foundation for the ETAP pilot.

Dr. John Deke, a senior fellow and methodological advisor at Mathematica, oversaw the experimental design to evaluate alternative strategies to increase response rates in online surveys, and conducted quality assurance of the approach.

REFERENCES

Dillman, D. "The Promise and Challenge of Pushing Respondents to the Web in Mixed-Mode Surveys." *Survey Methodology*, vol. 43, no. 1, 2017, pp. 3–30.

Finamore, J., and D.A. Dillman. "How Mode Sequence Affects Responses by Internet, Mail, and Telephone in the National Survey of College Graduates." Presentation to the European Survey Research Association, Ljubljana, Slovenia, July 18, 2013.

Galesic, M., and M. Bosnjak. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly*, vol. 73, no. 2, 2009, pp. 349–360.

Lopatto, David. Survey of Undergraduate Research Experiences (SURE): First Findings. *Cell Biology Education*, Vol. 3, 270–277, Winter 2004

Lopatto, David. Undergraduate Research Experiences Support Science Career Decisions and Active Learning. *CBE—Life Sciences Education*, Vol. 6, 297–306, Winter 2007

Mathematica. "Research Experiences for Undergraduates (REU) Data System Pilot Analyses: REU Participant Exit Survey." Findings submitted to NSF. Washington, DC: Mathematica, January 2020.

Zuckerman, B., J. Doyle, A. Mudd, T. Jones, and G. Davis. "Assessment of the Feasibility of Tracking Participants from the National Science Foundation's Research Experiences for Undergraduates (REU) Sites Program." Final report. Washington, DC: STPI, 2016.