SECTION B: Collection of Information Employing Statistical Methods

Introduction

The National Science Foundation (NSF) requests that the Office of Management and Budget (OMB) approve, under the Paperwork Reduction Act of 1995, a three-year clearance for original data collection to be used in the evaluation of the Scholarships in Science, Technology, Engineering, and Mathematics (S-STEM) Program. The new data collections includes an S-STEM Recipient Survey (Appendix A); an S-STEM Principal Investigator Survey (Appendix C); and Site Visit Interview and Focus Group Protocols (Appendix D).

The S-STEM program, which operates within NSF's Division of Undergraduate Education, awards grants to a geographically diverse set of two- and four-year institutions of higher education (IHEs) that then provide scholarships for academically talented students, in science and engineering disciplines, who have demonstrated financial need. Individuals may be granted S-STEM scholarships for up to five years and receive up to $10,000 per year depending on financial need. The institutions also provide resources and support services (e.g. academic support, career counseling, recruitment, research opportunities) to students to support them in becoming and/or remaining engaged in science and engineering through successful pursuit of associate, baccalaureate, or graduate-level degrees. Institutions are not required to provide any specific type of resources or support services (e.g., faculty advisors, peer tutoring, career counseling) beyond student scholarships, so part of the evaluation will be to gather data on the services provided (see the proposed Principal Investigator Survey, Appendix C, Module D for specific survey items addressing the non-financial support offered to S-STEM scholarship recipients at S-STEM grantee institutions).

S-STEM awards to institutions may last up to five years. The maximum S-STEM request is normally not to exceed $600,000 in total direct costs; annual budgets are limited to $225,000 direct costs. As part of the direct costs, institutions may request funds up to 5 percent of the total scholarship amount for expenses related to program administration, and up to 10 percent of the total scholarship amount for student support services.

The goals of S-STEM are to: 1) improve educational opportunities for students; 2) increase retention and degree attainment; 3) improve student support programs at institutions of higher education; and 4) increase the number of well-educated and skilled employees in technical areas of national need.[8] Successful outcomes of the program include graduation with a STEM major, transfer of STEM students from two-year to four-year colleges, pursuit of STEM graduate degree studies, and employment in the STEM workforce. Funding for S-STEM comes from H-1B VISAs, funding which was reauthorized in FY 2005 through Public Law 108-447. NSF receives 40 percent of the H-1B funding, and the agency uses 75 percent of its portion of these funds for the S-STEM program (NSF, 2011). In 2006, S-STEM expanded to include technology and science fields beyond the original computer science, engineering, and mathematics fields included in its precursor program—the Computer Science, Engineering, and Mathematics

[8] National Science Foundation. NSF Scholarships in Science, Technology, Engineering, and Mathematics (S-STEM) Program Solicitation (NSF 12-529). Retrieved 5/21/13 from http://www.nsf.gov/pubs/2012/nsf12529/nsf12529.htm

Scholarship (CSEMS) Program, targeting instead the development of the broader STEM workforce.

The evaluation of S-STEM will explore the practices (academic and student support services), and characteristics of the implementation of S-STEM projects; the educational and career outcomes of the S-STEM scholarship recipients; the student or programmatic outcomes associated with the receipt of an S-STEM award; the relationship between specific project features or practices and student outcomes; and any promising practices or lessons learned about implementation of S-STEM projects. The evaluation will draw on extant data as well as require new data collection efforts. This package seeks OMB approval for the new data collection efforts, which include a survey of S-STEM principal investigators, a survey of S-STEM scholarship recipients, and site visit interview and focus group protocols. Although approval is sought only for the new data that will be collected for the study, our description of the evaluation below includes both the extant and original data sources that will serve the evaluation.

Given the nature of the S-STEM program and the type of information being sought, a mixed-methods evaluation design will be employed. The mixed methods approach to the evaluation integrates both quantitative and qualitative data analyses and methods. Specifically, the evaluation will consist of the following descriptive, relational, benchmarking, and quasi-experimental study components:

- A descriptive implementation study that describes the ways in which S-STEM projects (i.e., grantee institutions) recruit and retain students in STEM fields, allocate scholarship funds, and provide educational and support programming for scholarship recipients;
- A relational study of associations between project characteristics and practices and recipient outcomes;
- A benchmarking comparison of recipients' educational and academic support experiences to national trends; and
- A comparative, quasi-experimental study using Propensity Score Matching (PSM) at the individual level to compare the educational and career outcomes of S-STEM scholarship recipients to a matched sample of respondents from a national survey of postsecondary students.

The data sources to be used in the evaluation include both new data collections and extant sources (see Section A.1 for more information about the circumstances requiring these data sources):

New Data Collections

- Survey data from S-STEM Principal Investigators to understand how the S-STEM projects operate and how the S-STEM awards are implemented;
- Survey data from S-STEM scholarship recipients to: (a) compare their educational and career outcomes to a matched comparison group of respondents to Beginning Postsecondary Students Longitudinal Study (BPS) surveys, an extant data source; and (b) benchmark (contextualize) their undergraduate experiences and academic engagement to a reference group of respondents to the National Survey of Student Engagement (NSSE);[9]

[9] The NSSE is administered once annually at participating institutions of higher education (IHE) that award a four- or five-year Bachelor's degree to a random sample of currently-enrolled students within each institution.

- Site visits to a purposive sample of S-STEM projects at up to six institutions of higher education per year to gather in-depth information through interviews and focus groups, about project implementation in particular contexts.[10]

Extant Sources

- Extant program data from projects' annual reports to NSF and the S-STEM Monitoring System to examine the components and practices employed by S-STEM projects;
- Extant institutional data maintained by the Integrated Postsecondary Education Data System (IPEDS; U.S. Department of Education), to which institutions of higher education report annual data on a wide range of institutional characteristics including types of degrees granted; public versus private control; non-profit or for-profit status; demographic characteristics of student body (e.g., race, ethnicity, gender); undergraduate and graduate student enrollment; retention and degree completion rates; percentage of students majoring in academic fields; tuition and cost-of-attendance; percentage of students receiving financial aid; expenditures on instruction, research and development, operations; etc.;
- Extant data from a matched comparison group of respondents to the Beginning Postsecondary Students Longitudinal Study (BPS) surveys, part of a longitudinal, nationally-representative study conducted by the U.S. Department of Education to examine rates of college enrollment, degree attainment, and how students pay for college;[11]
- Extant data from a comparison group of respondents to the National Survey of Student Engagement (NSSE), an annual survey of undergraduate students at participating four-year institutions to examine students' engagement in learning and co-curricular campus activities, and students' perception of the degree of faculty and staff support for their educational and career goals.[12]

While approval is sought only for the new data that will be collected for the study, both the extant and original data sources that will serve the evaluation are described in Supporting Statement A.

S-STEM recipients will be matched on student level characteristics (see Exhibit B.1 below) to BPS survey respondents. Details of the approach to matching are contained in Appendix E.

---

10 In selecting sites, the study will be purposeful in ensuring a diversity of institutional types and contexts, with visits to two-year schools, minority serving institutions, public research universities, and private liberal arts colleges..

11 This longitudinal study of first-time beginning postsecondary students included three waves of data collection beginning with the National Postsecondary Student Aid Study (NPSAS04), with two follow-up waves: the Beginning Postsecondary Student surveys in 2006 and 2009 (BPS:04/06 and BPS:04/09).

12 To determine the feasibility of the matching, NSF's evaluation contractor has obtained access to the extant datasets (IPEDS, BPS, NSSE), processed the data, and identified the samples.

Exhibit B.1: Individual characteristics to be included in matching	
Financial aid	
	Received Federal Stafford Loan
	Received Pell Grant
	Received school grant/scholarship
	Received State grant/scholarship
	Received any other financial aid for education
Academic information	
	SAT I math score
	SAT I verbal score
	ACT composite score
	Cumulative GPA (or an estimate of GPA) through the end of the first school year
Demographic characteristics	
	Gender
	Age
	Race and Ethnicity
	Citizenship
	Disability
Other characteristics	
	Type of degree (Associates or Bachelor's degree)
	Major (Current field of degree)
	Full-time enrollment status

NSF has contracted with Abt Associates to conduct an evaluation of S-STEM that will explore the strategies, practices, and characteristics of implementation of exemplary S-STEM awardees; investigate S-STEM-related outcomes among recipients; and investigate institutional-related outcomes of S-STEM grantees. There were 513 S-STEM projects awarded from 2006 to 2010. The evaluation will include the 494 S-STEM projects that provided scholarships to undergraduate recipients, the PIs of these projects, and a sample of S-STEM scholarship recipients who were (or are currently) supported by these projects. The study will answer the following overarching questions:

1. How do individual awardees implement their S-STEM projects?
2. What are the educational and career outcomes of the S-STEM scholarship recipients? How do outcomes of S-STEM recipients compare to an appropriate comparison group or national trends?
3. Are there student or programmatic outcomes associated with receipt of an S-STEM award?
4. What is the relationship between specific project characteristics or practices and student outcomes? Are there promising practices or lessons learned about implementation of S-STEM projects?

Exhibit B.2 below maps the research questions to the sources of data, and the proposed analyses. (The analyses are described in Section B.3.) Following Exhibit B.2, we discuss the proposed sampling methods.

Exhibit B.2: Map of Research Questions to Data Sources and Analyses

Research Questions	Data Sources							Analyses
	New			Extant				
	PI survey	Recipient survey	Site visit interview/focus groups	Annual Reports/ Monitoring System	IPEDS	BPS data	NSSE data	
1. How do individual awardees implement their S-STEM projects?								
a. What are promising practices in key areas (e.g. recruitment, academic support, retention) of and/or lessons learned from highly effective and successful S-STEM projects?	X		X	X				Descriptive, Relational
2. What are the educational and career outcomes of the S-STEM scholarship recipients? How do outcomes of S-STEM recipients compare to an appropriate comparison group or national trends?								
a. What is the effect of the S-STEM program scholarships on recipients?		X	X				X	Descriptive, Benchmarking
b. To what extent do S-STEM scholarship recipients transfer to a four-year program as compared to an appropriate comparison or national trends?		X			X	X		Comparative
c. To what extent do S-STEM scholarship recipients join the STEM workforce after graduation as compared to an appropriate comparison or national trends?		X			X	X		Comparative
d. To what extent do S-STEM scholarship recipients apply for and attend a STEM graduate program as compared to an appropriate comparison group or national trends?		X			X	X		Comparative
e. How effectively does the program meet the needs of academically talented financially needy STEM students as compared to other need-based opportunities and/or mechanisms?		X	X		X	X		Comparative
3. Are there student or programmatic outcomes associated with receipt of an S-STEM award?								
a. What is the effect of the S-STEM program on student outcomes (e.g. recruitment, retention) in STEM at institutions that have received an award compared to an appropriate comparison or national trends?	X	X			X	X		Comparative
b. What is the added value of the S-STEM program on institutions that receive awards?			X					Descriptive
c. What is the added value of the S-STEM program on student support and educational opportunities for recipients in institutions receiving S-STEM program funding?	X		X				X	Descriptive, Benchmarking
d. What is the added value of the S-STEM program on STEM programs, STEM departments, and/or IHEs that have received S-STEM funding?	X		X					Descriptive
4. What is the relationship between specific project features or practices and student outcomes? Are there promising practices or lessons learned about implementation of S-STEM projects?								
a. Are there any unintended positive and negative consequences/outcomes that can inform project and program management and design?	X	X	X	X				Descriptive, Relational
b. To what extent can outcomes be attributed to components supported by the S-STEM program?		X		X				Descriptive
c. Are there other NSF-funded education-oriented projects at the institution?	X							Descriptive
d. What is the relationship among these NSF-funded education-oriented efforts at the institution?	X		X					Descriptive

B.1. Respondent Universe and Sampling Methods
From 2006 to 2010, the S-STEM program granted 513 S-STEM awards. Of these, 19 provide S-STEM scholarships only to graduate students and are excluded from the sampling frame. The evaluation will examine the remaining 494 S-STEM awards that provide scholarships to undergraduate recipients (see Exhibit B.3). Given that earlier studies of NSF STEM scholarships and fellowship programs, conducted by the contractor, have achieved response rates of 80%,[13] this study has set a target response rate of 80%. This target response rate corresponds to the threshold below which OMB guidance requires a nonresponse bias analysis,[14] thus we include a plan for nonresponse bias analysis (see section B.4 and Appendix G), in the event that an 80% response rate is not achieved.

PI Survey Sample
Of the 513 S-STEM awards made from 2006 through 2010, 19 are excluded because they give scholarships to graduate students only. This leaves 494 eligible awards in the sampling frame; among these, there are 483 unique PIs (11 had more than one S-STEM award). We propose to survey the census of 483 unique PIs.

We propose a census of PIs because extant data are not available on characteristics or models of the S-STEM projects, which vary with respect to recruitment and selection strategies, and educational opportunities and support services available to scholarship recipients. The lack of data makes it difficult to divide this population into homogenous subgroups to obtain reasonable strata from which to sample, and a simple random sample could potentially leave out programs that are unique in nature and not provide precise estimates of the population. Because we propose to survey the census of PIs, we do not present a sampling plan. All 483 unique PIs will be invited to participate in the PI survey.

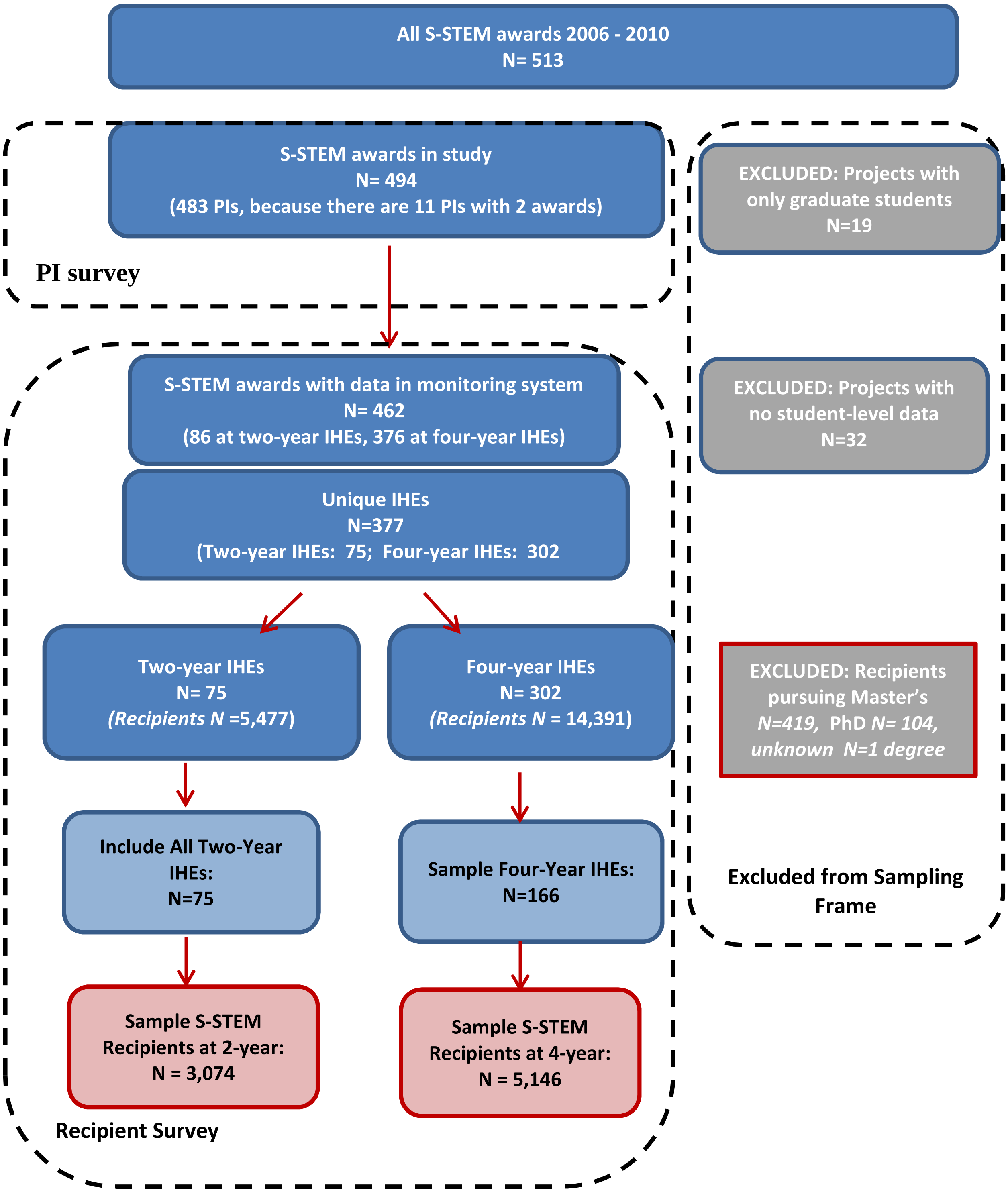
S-STEM Scholarship Recipient Sample
The target population for the Recipient Survey includes S-STEM scholarship recipients enrolled at S-STEM grantee institutions between 2006-2010. Because the S-STEM monitoring system collects information about scholarship recipients from PIs, the monitoring system serves as the source for identifying recipients to be included in the study. However, student-level information is only available for 462 awards, and the 32 awards with no student-level information are excluded from the sampling frame for the recipient survey (Exhibit B.3).

Because the comparison group will be matched to recipients within IHEs the number of unique IHEs will serve as the initial unit for sampling. In analyses, we will compare recipients selected from within an IHE to a matched comparison group of students who attended the same IHE and were participants in the BPS:04/09 survey – for which we will use extant data. The differences in outcomes between S-STEM recipients and BPS respondents will be calculated by averaging across IHEs.

---

13 Abt Associates Inc., CAREER, GK-12, and NSF-International studies.
14 See Guideline 1.3.4 in the *Office of Management and Budget Standards and Guidelines for Statistical Surveys*, September 2006. http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Exhibit B.3: Study Samples for PI and S-STEM Recipients for Surveys



The 462 awards were made to a total of 377 unique IHEs (n=75 two-year IHEs and n=302 four-year IHEs). The S-STEM projects in these two-year IHEs had a population of 5,477 undergraduate S-STEM recipients. The four-year IHEs had a population of 14,391 undergraduate S-STEM recipients (these numbers include those who are currently or were formerly enrolled). The sum yields a total of 19,868 eligible recipients in the sampling frame (see Exhibit B.3).

We will sample from two independent populations of scholarship recipients, classified by the type of degree program in which they were enrolled during the first year they received S-STEM support. We will select the census of two-year awardee IHEs (n=75). Within these 75 IHEs we will select a sample of 3,074 S-STEM scholarship recipients. From four-year awardee IHEs we will select a sample of 166 IHEs. Within these we will select a sample of 5,146 S-STEM scholarship recipients. The total number of S-STEM recipients who will be invited to complete the recipient survey is 8,220.

The analytic sample size estimates for the two- and four-year IHE recipient samples are based on a desired minimum detectable effect size (MDE) of 0.075 for continuous outcomes (such as time to degree), and corresponding minimum detectable differences (MDDs) of between 2.3 and 3.8 percentage points for dichotomous outcomes (such as “earned degree” versus “did not earn degree”).[15] Previous literature relevant to this study has shown that the typical effect size for similar continuous outcomes ranges from 0.075 to 0.2 and from 5 to 20 percentage points for dichotomous outcomes(e.g., Crisp et al., 2009; Eagan et al, 2010; Dowd & Coury, 2006; Ishitani, 2012; Melguizo & Dowd, 2006). Based on this literature, the proposed evaluation is designed to detect MDEs of 0.075 which corresponds to MDDs of between 2.3 and 3.8 percentage points. Sampling calculations are based on the following assumptions:

- Significance level (alpha) = 0.05;
- Power = 80 percent
- The variance of effect size of the outcome across S-STEM awardee institutions is zero (this assumption is consistent with a fixed effects model for the treatment variable).
- The proportion of variation in outcomes explained by institutional-level covariates (reported symbolically as B) is approximately 0.1[16] and that the proportion of variation explained by individual-level covariates (R-squared) is 0.2.[17]
- The number of units in the constructed comparison group will equal the number of units in the “treatment” group, namely the S-STEM scholarship recipients.

To achieve the desired MDEs, the target sample sizes reported here have been increased to account for the following:

- Outdated or inaccurate contact information resulting in an estimated non-location of 20 percent of scholarship recipients named in the S-STEM Monitoring System;

[15] The MDE (used for continuous variables) is expressed as a percent of the standard deviation of the outcome, and the MDD (used for dichotomous variables) is expressed as a percentage point difference in the mean value of the outcome.

[16] Dowd & Coury, 2006; Melguizo & Dowd, 2006 show that selectivity of school explains 10 percent of persistence and graduation rates.

[17] Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor’s degree attainment. Washington, DC: U.S. Department of Education. Retrieved September 20, 2012 from http://www2.ed.gov/pubs/Toolbox/Title.html.

- An estimated non-response rate of 20 percent for those scholarship recipients who can be accurately located;
- An estimated loss of 20 percent of respondents during the propensity score matching (PSM) of recipients to a comparison group of non-recipients (see Appendix E for details of the PSM); and
- An increase in the total number of IHEs in case no recipients from a selected S-STEM awardee respond (non-response at the IHE level).

To create a balance across award cohort (i.e., cohort of awards made in 2006-07; in 2007-08; etc.) and recipient characteristics of interest, within each IHE we will select a stratified systematic sample of recipients from each of four strata, representing cross-classifications of gender (male, female) and minority status (underrepresented minority, non-underrepresented). Recipients will be selected systematically using Lahari's circular method from each stratum. Prior to sampling the recipients will be sorted by cohort and discipline. Systematic sampling after sorting by cohort and discipline will increase the likelihood of having a wide distribution of recipients across the grant projects, the cohorts, and various disciplines in the selected sample. We will over-sample recipients from strata with small cell counts to ensure adequate representation.

See Appendix E for details on the sample selection and power analyses.

Sampling IHEs for Site Visits

A purposive sample of six S-STEM awardee institutions per year for three years will be selected for site visits. Site selection requires variation on the dimensions of theoretical interest (Seawright & Gerring, 2008), thus sites will be selected to represent the variety of institutional contexts in which S-STEM awards operate (e.g., institution type, size, location, STEM fields supported). The set of sites selected will include community colleges, 4-year institutions, and universities, and will represent the geographic diversity of S-STEM institutions.

B.2. Information Collection Procedures

Primary data collection activities include a web-based survey of S-STEM Principal Investigators; a web-based survey of current and former scholarship recipients at 2006-2010 S-STEM awardee IHEs; and interviews and focus groups conducted during site visits to selected S-STEM institutions. We describe the data collection procedures for each activity below.

PI and Recipient Surveys

To locate survey respondents, researchers will construct a database of current contact information using information available through the S-STEM Monitoring System. The evaluation contractor, Abt Associates (Abt), will also use other location methods such as general web searches and the fee-based electronic database Accurint. Abt will then contact S-STEM PIs to request contact information for individuals in the sample for which this information is missing. (See Appendix F for the invitations.)

S-STEM recipients and PIs will receive an email inviting them to participate in an online survey (see Appendix F for email text). This email will explain the purpose of the study and a link to the web-survey. Each survey link will be unique to the selected respondent. This survey link will launch the respondent's web browser and open the survey to an introductory "landing page" that will describe the purpose of the survey, its expected length, and instructions for navigating through the survey. Links to a "frequently-asked questions" page will be included along with information about how the respondent's privacy will be safeguarded. The landing page will clearly display the OMB control number and expiration date, along with confirmation that the study has received IRB approval and contact information for potential survey respondents to use if they have questions about the study. Respondents who consent to participate in the survey will be asked to click on a button to launch the survey.

During the one-month survey field period, up to three email reminders and three telephone reminders will be used to encourage survey completion. (See Appendix F for text of reminders.) If desired response rates have not been achieved at that time, the contractor may extend the survey deadline by one to two weeks. Throughout the data collection cycle, a toll-free study telephone number and e-mail address will be available to allow potential respondents to easily obtain answers to questions or concerns about the study. In their survey invitation, S-STEM PIs will also be informed of the timeline for fielding of the Recipient Survey so that they may encourage scholarship recipients to participate in the survey.

Once approval is received from OMB, the web-based surveys will be programmed for online data collection. The study team will test the programmed surveys to ensure functionality and accuracy of data capture. (See Appendices A and C for copies of the survey instruments.) The Recipient Survey proposed for this study has been developed using identical survey items from the BPS:04/09 survey along with items taken directly from the earlier-administered NPSAS:04 survey (see Appendix B for sources of specific items) to permit:

- matching of S-STEM recipients to BPS respondents on a shared set of characteristics;
- comparison of outcomes using a shared set of measures (i.e., identical survey items and response categories); and
- inclusion of covariates based on a common set of measures (identical or nearly-identical survey items and response categories).

Site Visit Interviews and Focus Groups

During each site visit, S-STEM PIs, other STEM faculty, senior college/university administrators, and currently-enrolled S-STEM scholarship recipients will be invited to participate in interviews and focus groups. Prior to each site visit, each PI will receive an email from NSF introducing the study, why the respective institution was selected for site visit, and why site visits are an important component of evaluating federally-funded STEM higher education programs. (See Appendix F for email text.) The research team will then send an introductory email to begin communication with the PI and other relevant individuals to address details of the site visit, including logistics and expectations (See Appendix F for email text). Next, the research team will contact currently-enrolled S-STEM scholarship recipients, faculty in STEM departments, and university administrators via email to invite these respondents to participate in an interview (or focus group) and to explain the purposes of the site visit. (See Appendix D for the site visit interview protocols and Appendix F for introductory letters and emails).

In preparation for each site visit, a detailed background study will be conducted using available sources (S-STEM institutional websites, annual project reports) to allow the team to identify institution-specific probes during focus groups and interviews (e.g., use and helpfulness of S-STEM offerings by students).

Once on-site, a team of three to six site visitors will collect in-depth qualitative data will be collected following structured interview protocols (Appendix D) keyed to each respondent type. Interviews and focus groups will be audio-recorded so that notes can be captured and analyzed using a qualitative software package (see Section B.3 for approach to analyses). Site visit reports will be prepared for each site and shared with the S-STEM project. The project will have the opportunity to review and comment on the site visit report before it is submitted to NSF.

Data Security and Privacy Protection

Abt Associates, the contractor that will conduct the proposed data collection activities, has conducted numerous studies involving sensitive and non-sensitive information. All project staff members employ both electronic and physical safeguards to protect data from unauthorized access. Electronic project directories, files, and databases are accessible to project staff only and are protected by discretionary access control lists (ACLs), group memberships, passwords, and locking workstations. Access to the data processing area and database servers is limited to authorized personnel. Building security staffs all sites 24 hours, 7 days per week. To protect against data loss, Abt also uses automated, redundant backup procedures and file management techniques to ensure that files are not inadvertently lost or damaged. All data, including the web-based survey data, will be maintained on a secure server with appropriate levels of password protection. After the conclusion of each site visit, interview notes and audio recordings will be stored in locked storage locations within the contractor’s offices. Respondent names will be replaced with pseudonyms in NVivo and other analysis files created.

B.3. Estimation Procedures

The research team will conduct descriptive, comparative, and relational analyses of extant and primary data collected from interviews and surveys to address the four primary research questions as depicted in Exhibit B.2 above.

Descriptive Analyses

Descriptive statistical procedures will include the calculation of means and standard deviations for continuous variables (e.g., mean amount of scholarship funding per student) and percentages for categorical variables (e.g., the percentage of S-STEM IHEs that provide learning support centers for math or science coursework). Where applicable, the number of missing responses will also be reported. Where nationally-representative data or data representative of students at particular institutions are available, these will be presented as a context in which to interpret descriptive data on recipients or S-STEM awardee institutions. Site visit interview and focus group data will be analyzed to identify patterns of similar and contrasting responses from respondents. Evaluators will begin by developing an understanding of the S-STEM project in its institutional context, including the characteristics of affiliated departments, academic support services at the institution, particularly those related to STEM courses, and accessibility of project activities and support services. Analyses will focus on identifying responses keyed to the evaluation indicators, and on identifying patterns both within and across institutions, categories, and emergent themes.

Relational Analyses

To explore relationships between S-STEM project characteristics (e.g., scholarship amount per student, types of student support services offered) and recipient outcomes of interest, analyses will include multivariate, ordinary-least-square regression models for continuous outcomes, and logistic regression models for dichotomous outcomes (including covariates, where applicable for institutional (e.g., public vs. privately controlled, Carnegie classification) and recipient characteristics (age, race/ethnicity, gender). To make optimal use of data in exploring relationships between recipient characteristics and outcomes, evaluators will use hierarchical linear modeling to derive appropriate standard errors for multi-level data (i.e., data that reflect both intra- and inter-individual change over time). Where corresponding nationally-representative data about relationships between institutional characteristics and student outcomes are available, reports will include such data as relevant benchmarks against which to compare the strength of association for such characteristics and outcomes among S-STEM participant institutions and recipients.

Quasi-experimental Comparative Analyses of Recipient Survey Data

We propose a quasi-experimental design (QED) using propensity score matching (PSM) to create matched treatment and comparison groups. (The advantages and limitations of this approach are described at the end of this section.) The comparison group will consist of participants in the Beginning Postsecondary Students (BPS) study (BPS:04-09 and NPSAS04 survey data); the treatment group will consist of S-STEM recipients within each awardee IHE. These two groups will be matched based on IHE attended and on individual-level characteristics observed prior to receipt of S-STEM. Such characteristics include, for example: college admission test scores (SAT or ACT), first year college GPA, financial aid received, and demographic information such as marital status, income, number of dependent children, race, gender, disability, and ethnicity. Differences in outcomes (and standard errors of those differences) for S-STEM recipients and the matched comparison group will be estimated using multi-level linear and logistic regression models that incorporate the propensity scores and control for other covariates shown to be associated with these outcomes. Differences will be tested for statistical significance (for details see Appendix E).

Analyses comparing S-STEM recipients' outcomes to those of a nationally-representative comparison group of BPS respondents will be conducted separately for two- and four- year IHEs. The first step of the analysis is to create a matched comparison group of students using the respondents of the BPS survey. Matching will be done within each IHE from which Recipient Survey data are collected. The next section details the propensity score matching process; this is followed by an explanation of procedures to estimate differences from this matched sample.

Propensity Score Matching

A quasi-experimental statistical method termed propensity score matching (PSM) will be used to match S-STEM recipients with BPS respondents.[18] Propensity score matching allows S-STEM

[18] Rosenbaum, P.R. and Rubin, D.B. (1984) Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association* 79*:* 516-524.

recipients to be matched to BPS respondents (control group) *based on their similarity*.[19] With this approach, the S-STEM recipients would be compared to BPS respondents who are as similar as possible to them in terms of observable characteristics, allowing us to estimate what the S-STEM recipients' outcomes would have been had they not received the S-STEM award, had other characteristics been equivalent.[20] PSM analysis will be performed using the following four steps:

Step 1: We will identify a set of characteristics, measured prior to the treatment group's receipt of S-STEM scholarship funding (i.e., called pre-treatment characteristics) that will be used in the propensity score model to match S-STEM recipients to BPS respondents. These characteristics include variables that likely are associated both with the likelihood of receiving an S-STEM scholarship (e.g., financial aid received for the first year of enrollment; SAT or ACT college admissions test scores) and with the outcomes of interest (e.g., persistence to degree attainment). S-STEM scholarship recipients selected by the awardee institution must be US citizens or permanent residents who are enrolled full time in a program leading to an associate or baccalaureate degree in a STEM discipline;[21] selected students must demonstrate financial need and academic potential or ability. Therefore, pre-treatment characteristics such as SAT/ACT scores, types of financial aid received, college credit for high school coursework, and first year GPA (if prior to receipt of an S-STEM scholarship), will be used as matching variables. These data will be obtained from survey data (the Recipient Survey and BPS extant survey data).

Step 2: Using these pre-treatment characteristics, we will fit a logistic regression model that predicts the probability of being awarded a STEM scholarship. We will then use the coefficients from this model to estimate, for each individual-- including each BPS respondent--a "propensity score," which represents the probability of receiving an S-STEM scholarship. Next, **within each IHE**, we will identify and exclude from further analyses those individuals for which no credible match from the other group can be found (that is, any S-STEM recipient for whom there are no credible matches in the BPS respondent group within that IHE will be excluded from analysis; and vice versa, any BPS respondent for which there are no credible S-STEM recipient matches will be dropped).[22]

---

[19] Angrist, J.D. (1998) Estimating the labor market impact of voluntary military service using social security data on military applicants, *Econometrica*, 66: 249-288. 1998; Heckman, J. Ichimura, H. Smith, J. and Todd, P. (1998). Characterizing selection bias using experimental data, *Econometrica*, 66: 1017-1098.

[20] The comparison groups will be constructed separately for recipients who were pursuing an Associates degree and those who were pursuing a bachelor's degree at the time they first received an S-STEM scholarship, and propensity score models will be fit separately for the two types of scholarship recipients. All comparison analyses will be conducted separately for the each.

[21] Students enrolled for a graduate degree in a STEM discipline are also eligible for an S-STEM scholarship but are not included in the proposed evaluation.

[22] More technically, those individual who fall outside of the "area of common support," the range of common propensity scores across S-STEM recipients and BPS respondents within that IHE will be excluded from analysis. Enforcing the criterion of common support is important to ensure the similarity of the matched recipients to non-recipients (Rosenbaum and Rubin, 1983; Caliendo and Kopeinig, 2008).

Step 3: **Within each IHE** we will use the estimated propensity scores to create matched sets of S-STEM recipients and BPS respondents. If the matching is not possible within an awardee institution, we will match students from institutions with similar institutional characteristics using data from the Integrated Postsecondary Education Data System (IPEDS). [23] There are a variety of techniques available for using propensity scores to create such matched sets; we will use stratification to place matched sets of treatment and comparison students into five subgroups, or strata within each IHE (for more detail, see Appendix E). [24]

Step 4: Finally, we will test whether there are any differences between the S-STEM recipients and corresponding "matched" BPS respondents within each propensity score strata **for each IHE**. Once the stratification is "balanced," propensity scores of the treatment and comparison group members within each stratum are statistically equivalent (see Appendix E for additional details of this step in the PSM procedure). If we find that statistical balance is not achieved across treatment and comparison groups in each stratum for each IHE, we will modify the logistic model used in Step 2 by including interactions and higher-order terms of the unbalanced characteristics and repeat Steps 2 through 4 until satisfactory balance is achieved.

Estimation of Differences

Following the propensity score matching, we will estimate the effect of the S-STEM program separately for recipients in two- year and four-year institutions by comparing S-STEM recipients' outcomes to those of their comparison group using multivariate regression models for each outcome of interest (see Exhibit B.4). For each outcome, the regression model will include a dichotomous indicator or "dummy" variable to indicate whether each student included in the model is an S-STEM recipient or not; the model will employ a number of matching characteristics and other control variables that are hypothesized to affect the outcomes of interest as covariates. The inclusion of the matching characteristics in this model will give us the chance to get a "doubly-robust" estimate since they will have been used twice: both in the propensity score model and in the estimation of effect sizes.[25] The regression model produces separate estimates of the effect of S-STEM scholarship receipt for each propensity score stratum; these estimates for each propensity score stratum will be aggregated based on the relative proportion of treatment and comparison group members in each stratum (see Appendix E for details). Estimated coefficients from the regression model and the overall treatment effect estimates will be presented along with corresponding standard errors and p-values. Hence, for dichotomous outcomes, estimates will be presented in the form of percentage points, whereas for continuous

---

23 Examples of these characteristics are: geographic location of institution, sector of institution (public, private, non-profit), level of institution (2 - year vs. 4 year), historically black college/university, degree of urbanization, Carnegie classification, cost of attendance, selectivity of the institution enrollment size, and other enrollment characteristics.

24 Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica*, 71(4): 1161-89; Morgan S.L. and Harding D. J. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research*, 35(1), 3–60; and Abadie, A., & Imbens, G. W. (2009). Matching on the Estimated Propensity Score. NBER Working Paper.

25 Ho D.E., Imai K., King G., and Stuart E. A. (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*.; 15: 199–236.; Morgan S.L. and Harding D. J. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research*, 35(1), 3–60.

outcomes, overall estimates in "effect size" units (e.g., Hedges' g [Hedges & Olkin, 1985]) will be presented.

Exhibit B.4: Key S-STEM scholarship recipient outcomes of interest	
Primary S-STEM recipient outcomes	**Description**
Degree earned (Bachelor's, Associates)	Was the expected degree earned
Major field (primary and secondary if double major) of degree	Was the degree earned in STEM field
Time to degree	Time elapsed from date of enrollment to date degree earned
Time enrolled to degree	Number of months of enrollment required until degree earned
Date degree expected	If currently enrolled, date degree is expected
Transfer to Bachelor's degree program	If S-STEM recipient was enrolled at two-year institution at first receipt of scholarship, did recipient ever enroll in a bachelor's degree program
Reason for leaving S-STEM institution	If degree not earned from S-STEM institution reason for leaving former institution
Current employment status	For former recipients not enrolled in an IHE, employed or not
Type of job	Occupation (STEM or not STEM related)
Hours worked	Average number of hours worked per week
Amount earned	Annual salary or hourly wage earned
Relationship of job to degree from S-STEM institution	Is current job related to major field of degree earned (or pursued but not earned)
Number of jobs worked during enrollment at S-STEM institution	Total number of jobs worked while enrolled at S-STEM institution
Amount of undergraduate loans	Total amount borrowed for undergraduate degree (earned or not)

Advantages of Approach

The approach to estimating differences in outcomes for S-STEM recipients relative to those of a matched comparison group of BPS study participants that is outlined above has several advantages and some notable limitations. The advantages of this approach include:

- The BPS (04/09) data are derived from a nationally representative sample of students who were enrolled in post-secondary education for the first time in the 2003-04 academic year. Additionally, it is representative at the student level for 3 institution types (public, private for-profit and private non-profit). These data thus provide a national context against which to compare educational and employment outcomes for S-STEM recipients; in addition, the data could be used to produce representative estimates at the state level (in 12 states).
- The BPS data include substantial numbers of variables measuring the same outcomes of interest for the proposed evaluation of S-STEM: enrollment history, including transfer of credits from one institution of higher education to another and periods of non-enrollment; persistence, progress and degree attainment in post-secondary education; major field of degree (e.g., STEM or other); financial aid received for undergraduate education; enrollment in post-baccalaureate education; and employment history.
- The BPS data also contain abundant variables on which to match S-STEM recipients to BPS respondents including, for example: major field of study, SAT/ACT scores, college credit earned for high school coursework, receipt of Pell grant in the first year of

undergraduate education, income, marital status, and number of dependent children, parents' highest educational levels, race, ethnicity, and other demographic characteristics. Similarly, the data include institutional characteristics of the colleges or universities respondents attended, such as Carnegie classification, institutional level (two- or four-year; types of degrees granted) and sector (private vs. public), selectivity, status as an HBCU (Historically Black College/University) or Hispanic-Serving Institution (HSI), and cost of attendance.

- The relationship of specific features of the S-STEM projects to outcomes of interest could be investigated in exploratory analyses. If there are promising strategies identified NSF might decide to pursue a rigorous examination of these by putting in place the conditions for a demonstration project.

Since this proposed approach is a retrospective study, in which S-STEM recipients are asked to report on both current and prior educational characteristics, findings will be available in a shorter period of time than a prospective study in which current S-STEM recipients were followed over time for several years. The proposed design also incurs a lower cost than some alternative approaches, because it exploits extant data to construct a comparison group, rather than collecting new, primary data from non-S-STEM participants.

Limitations

There are also several limitations of the proposed approach, although the study is designed to limit or mitigate several of these limitations.

Location and response rates Surveys will involve S-STEM participants from cohorts that span 2006-2010 of the S-STEM program. Reaching participants going back to the early cohorts may be difficult. Thus, the sampling approach will over sample S-STEM recipients so that there will be the required number of recipients for analysis even if response rates are lower than ideal. Earlier studies have shown location rates of 75%[26] and response rates of 80%.[27] As a result, we will boost our sample size of S-STEM recipients to account for these previously cited rates. Further, the analysis plan involves a specific approach to address potential non-response bias.

Recall bias. Another potential limitation of surveying S-STEM recipients from cohorts that span 2006-2010 is the potential for recall bias – an inability of respondents to accurately remember the events being examined. Recall bias has been extensively studied in the medical literature. For example, Litwin and colleagues documented that patients could not accurately recall their health status as little as three years after undergoing a surgical procedure (patients tended to report pre-surgery quality of life as better than it actually was).[28] Thus, the study will explore whether there are marked differences in the responses of individuals from earlier cohorts when compared to later cohorts.

---

26 Abt Associates Inc., (2009) Needs assessment of the NIGMS Research Supplements to Promote Diversity in Health Related Research: Final Report, April, 2009.

27 Abt Associates Inc., CAREER, GK-12, and NSF-International studies.

28 Litwin, M.S. and McGuigan, K.A. (1999). Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer, *Journal of Clinical Oncology* 17(9): 2882-8.

Propensity score matching. Propensity score matching deals with selection bias by explicitly balancing the observable differences between program participants and non-participants and constructing matched treatment and comparison groups that are then used to estimate the effects of the program. Propensity score estimators are valid under the "conditional independence" assumption, which states that the assignment status of a participant or a non-participant (to the treatment or comparison condition) is "ignorable" conditional on his/her propensity score.[29] In other words, propensity score matching relies on the statistical equivalence of matched treatment and comparison groups conditional on their observable characteristics. The major threat to the validity of propensity score estimators, therefore, comes from the existence of unobservable characteristics that affect both outcomes of interest and an individual's assignment status. For example motivation, an unobservable characteristic, is often an important factor that effects an individual's participation in a program as well as his/her outcomes. Program participants may be more motivated than non-participants and thus have better outcomes. In this case, using propensity score matching may not fully remove the inherent difference between the treatment and the comparison groups.

The best way to deal with the threat of unobservable characteristics is using as many "relevant" observable characteristics as possible in the propensity score matching process, so that the effect of these factors is reduced. This study will employ an extensive set of matching variables to minimize the selection bias. Assuming that the S-STEM award decisions were also based on the same information, this approach should account for some of the inherent differences between recipients and BPS comparison and minimize the selection bias.

Beginning Postsecondary Student (BPS) longitudinal data. There are three relevant limitations related to the use of the BPS data. First, the most recent administration of BPS for which it is feasible to obtain data within the study's timeline is out of phase with the study data collection, which is planned for 2013. As a result, this comparison group limits the ability to control for changes over time that coincide with (or overlap with) the onset of the S-STEM program or receipt of an award that could affect outcomes of interest independent of the program's effect. That is, the sample of BPS students to be used as the comparison group is a cohort of first time post-secondary students in 2003-04 who were surveyed longitudinally in 2004, 2006, and 2009. In contrast, the S-STEM recipient treatment group are students who were enrolled as undergraduates in 2006 or later. Because they were attending college at different historical times, systematic differences between the BPS respondents and S-STEM scholarship recipients could be unobservable –and thus might bias estimates of the differences in outcomes between the two groups.

Available institutional and student level data will be used for both S-STEM and BPS students to control for factors that could explain observed differences, thus reducing the number of plausible alternative explanations for observed effects to address these limitations. Furthermore, it is also necessary to control for characteristics of recipients that could affect receipt of S-STEM support, and independently could affect outcomes of interest to the evaluation. To control for these

[29] Blundell, R. & Costa Dias, M. (2000) Evaluation methods for non-experimental data, *Fiscal Studies* 21(4) (2000): 427-68; A. Abadiea and G. W. Imbens, (2009) *Matching on the estimated propensity score.* NBER Working Paper.

characteristics before they could be altered by exposure to S-STEM support, baseline data (i.e., pre-grant year data) measured in the years prior to receipt of S-STEM support will be used in the PSM process.

The second limitation of the BPS data is that the survey was not designed to directly measure many of the outcomes that are pertinent to our study. However, the necessary outcomes can be calculated using the data collected across the years. For example, the percent of students in two-year colleges who subsequently enrolled in a 4 year college needs to be calculated using school enrollment history collected over the years rather than using a direct question that asks whether the recipient had ever transferred. The use of the BPS data as a comparison will be limited to the subset of items surveyed and we have to calculate the outcomes from the data available using reliable and consistent assumptions.

Finally, the BPS sample has the potential to include some S-STEM recipients. Since personal identifying information for BPS respondents is difficult to obtain, these overlapping individuals cannot be removed from the BPS sample leading to the possibility of some level of contamination in the control group. However, as stated above, the potential overlap between these two samples might be minimal due to different enrollment years, and hence these samples can be treated as independent.

B.4. Methods for Maximizing the Response Rate and Addressing Issues of Nonresponse

The study team will draw upon the expertise of Abt Associates in collecting and analyzing similar data for other large, federally-funded, institutionally-based programs. Achieving strong response rates begins with a well-designed, user-friendly instrument, and continues with providing a clear and convincing rationale for the survey and the importance of respondents' participation. The web approach will allow us to easily identify non-respondents for follow-up emails and phone contact to encourage participation, thus increasing our response rates. (See Appendix F for copies of email and phone messages to non-respondents.)

Prior to data collection, respondents will be provided with an early notification, including an electronic letter from NSF. (See Appendix F for email text.) At the designated opening date, evaluators will send an opening e-mail message to respondents that will include their user name and password, detailed instructions, the closing date, and project staff contact information. Included will be detailed on-screen instructions and extensive help functionality for survey items, including a Frequently Asked Questions section, glossary, and navigation instructions. Throughout the data collection cycle, a toll-free number and e-mail address will be used to ensure that potential respondents can easily and quickly obtain answers to questions or concerns.

As part of our analysis, Abt will explore potential non-response bias. Such bias could occur if individuals who cannot be located or refuse to participate in our study would have given systematically different responses on the survey (had they been located or responded) than the individuals that were located and completed surveys. Such bias is not automatically connected to poor find and/or response rates, as the reason for not being located or the decision to not participate could be completely unrelated to the questions on the survey. Furthermore, high find and response rates can make the effects of bias negligible even if bias exists. Nonetheless, the evaluation team will make every effort to achieve high find and response rates.

Using recipient information, evaluators will compare characteristics (e.g. gender, race/ethnicity, academic degree, and characteristics of recipients' institution, etc.) for those recipients who were located and completed the survey and those who were not located/did not complete the survey. If the two groups are not statistically significantly different based on these variables, it will be assumed that the data are missing completely at random and that the respondents are representative of the census of applicants and no adjustments will be made. If the two groups are different on some characteristics, sampling weights will be created to account for non-accessibility and non-response based on the aforementioned characteristics. (Appendix G contains the details of our non-response bias analyses.)

B.5. Tests of Procedures or Methods

The study team is working from a vetted logic model (presented in Appendix H), which has been informed by guidance from the study's external evaluation group of subject matter experts as well as S-STEM program staff. The approach to the evaluation is guided by investigation of the data contained in the S-STEM monitoring database, and a review of information contained in PI annual reports, final reports and local evaluation reports. Researchers pilot-tested the instruments utilized in this study. Mean response times for the Recipient Survey, which is comprised of items from the BPS:04/09 survey and the NPSAS:04 survey, are based on response times reported in the corresponding BPS:04/09 Methodology Report (Wine, Janson & Wheeless, 2011) and the NPSAS04 Methodology Report (Cominole, Seigel, Dudley, Roe & Gilligan, 2004). The survey questions were pilot tested with 4 S-STEM recipients and 9 PIs, and interviews were conducted with these respondents to identify any questions that were confusing. Respondents were asked to comment on the clarity and content of the survey questions, as well as the proposed duration of the data collection to help with an accurate estimate of time burden. As a result of the pilot testing, some clarifications and roll-over definitions will be added to the final programmed surveys. The interview protocols were each pilot tested with 5 to 9 individuals from each respondent group. Interviews and focus groups lasted from 30 to 60 minutes, with an average of 40 minutes.

B.6. Individuals Consulted

The NSF point of contact for this study is **Connie Kubo Della-Piana,** Division of Undergraduate Education.

The contractors for collaboration and analysis of data in this study are Abt Associates, Inc., its data collection subsidiary Abt-SRBI, and frequent collaborator Sage Fox Consulting Group. Staff from these organizations have deep knowledge on statistical methods, experience in evaluation of large scale programs, expertise in scientific research, and content knowledge of STEM higher education programs.

Senior technical staff overseeing the study includes:

- Alina Martinez, Principal Associate, Abt Associates, Inc. (Principal Investigator)
- Kelly Daley, Abt SRBI
- Carter Epstein, Scientist, Abt Associates, Inc.
- Lorelle L. Espinosa, Senior Analyst, Abt Associates, Inc.
- Luba Katz, Senior Scientist, Abt Associates, Inc.

- Amanda Parsad, Senior Scientist, Abt Associates, Inc.
- Alan Peterfreund, SageFox Consulting Group
- Hiren Nisar, Senior Analyst, Abt Associates, Inc.

Members of the study's external evaluation group of subject matter experts include:

- Dr. Deborah Allen, Director, Center for Educational Effectiveness and Associate Professor of Biological Sciences, University of Delaware
- Dr. Eun-Woo Chang, Instructional Dean of Science, Engineering and Mathematics, Montgomery College
- Dr. Bert E. Holmes, Philip G. Carlson Distinguished Chair, University of North Carolina, Ashville
- Dr. Patricia Mead, Professor of Engineering, Norfolk State
- Dr. Lizanne DeStefano, Professor of Quantitative and Evaluative Research Methodologies, Educational Psychology, University of Illinois, Champaign-Urbana

References

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment. Washington, DC: U.S. Department of Education. Retrieved September 20, 2012 from http://www2.ed.gov/pubs/Toolbox/Title.html

Abadie, A., & Imbens, G. W. (2009). Matching on the Estimated Propensity Score. NBER Working Paper.

Angrist, J.D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants, *Econometrica, 66*: 249-288.

Blundell, R. & Costa Dias. M. (2000) Evaluation methods for non-experimental data, *Fiscal Studies 21*(4): 427-68

Cabrera, A.F., Stampen, J.O., and Hansen, W.L. (1990). Exploring the Effects of Ability to Pay on Persistence in College. *The Review of Higher Education, 13*(3), 303-36.

Caliendo, M. & Kopeinig, K. (2008). Some Practical Guidance for Implementation of Propensity Score Matching, *Journal of Economic Surveys*, *22*, 1, 31-72.

Cominole, M., Siegel, P., Dudley, K., Roe, D., and Gilligan, T. *2004 National Postsecondary Student Aid Study (NPSAS:04) Full Scale Methodology Report* (NCES 2006–180). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved October 19, 2011 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006180

Crisp, G., Nora, A., and Taggart, A. (2009). Student Characteristics, Pre-College, College, and Environmental Factors as Predictors of Majoring in and Earning a STEM Degree: An Analysis of Students Attending a Hispanic Serving Institution. *American Educational Research Journal, 46*(4) 924-942.

Dowd, A. C. (2008). Dynamic interactions and intersubjectivity: Challenges to causal modeling in studies of college student debt. *Review of Educational Research, 78*(2), 232-259.

Dowd, A. C., & Coury, T. (2006). The effect of loans on the persistence and attainment of community college students. *Research in Higher Education, 47*(1), 33-62.

Dynarksi, S. (2008) Building the Stock of College-Educated Labor. *Journal of Human Resources 43*(3): 676-610.

Eagan, K., Hurtado, S., & Chang, M. (November, 2010) What Matters in STEM: Contexts That Influence STEM Bachelor's Degree Completion Rates. Presented at the Annual Meeting of the Association for the Study of Higher Education, Indianapolis, IN. Retrieved September 24, 2012, from http://heri.ucla.edu/publications-main.php

Heckman, J. Ichimura, H. Smith, J. & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66,* 1017-1098.

Hedges, L. and Olkin, I. (1985). St*atistical methods for meta-analysis*. New York: Academic Press.

Hirano, K, Guido W.I. & Ridder. G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica, 71*(4): 1161-89;

Ho, D.E., Imai, K., King, G., & Stuart ,E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*. 2007; 15: 199–236.

Horn, L., and Nevill, S. (2006). *Profile of Undergraduates in U.S. Postsecondary Education Institutions: 2003–04: With a Special Analysis of Community College Students* (NCES 2006-184). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ishitana, Terry T (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *Journal of Higher Education.* Ohio State University Press. HighBeam Research. 24 Sep. 2012. http://www.highbeam.com

Knapp, L.G., Kelly-Reid, J.E., Whitmore, R.W., and Miller, E. (2007). *Enrollment in Postsecondary Institutions, Fall 2005; Graduation Rates, 1999 and 2002 Cohorts; and Financial Statistics, Fiscal Year 2005 (NCES 2007-154).* U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved April, 30, 2008 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007154

Litwin, M.S. & McGuigan, K.A., (1999). Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer, *Journal of Clinical Oncology 17*(9): 2882-8.

Melguizo, T., & Dowd, A. C. (2009). Baccalaureate success of transfers and rising four year college juniors. *Teachers College Record, 111*(1), 55-89.

Morgan S.L. and Harding D. J. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research, 35*(1), 3–60.

National Academies (2010). *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads*. National Academies Press: Washington, D.C.

National Science Foundation (2011). National Science Foundation: FY 2012 Budget Request to Congress. Retrieved 5/1/11 from http://www.nsf.gov/about/budget/fy2012/pdf/fy2012_rollup.pdf

Office of Management and Budget Standards and Guidelines for Statistical Surveys, September 2006. http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Rosenbaum. P. & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association 79*, 516-524.

Rubin, D.B. & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. Biometrics, 52, 249-264.

Seawright, J. & Gerring, J. (2008). Case selection techniques in case study research. Political Research Quarterly, 61(2), 294-308.

Singell, L. (2004). Come and Stay a While: Does Financial Aid Effect Enrollment and Retention at a Large Public University? *Economics of Education Review,* 23:459-472.

Singer, J.D., & Willett, J.B. (2003) Applied Longitudinal Data Analysis. Oxford, U.K.: Oxford University Press.

Stake, R. E. (1995). *The Art of Case Study Research.* Thousand Oaks, CA: Sage Publications

Temple University Institute for Survey Research and Caliber Associates, Inc. (no date) Evaluation of the National Science Foundation's Computer Science, Engineering, and Mathematics Scholarship (CSEMS) Program: Phase Two – Survey Finding 2003-2004 Summary Report, Temple University, Institute for Survey Research and Caliber Associates, Inc.

U.S. Department of Education, National Center for Education Statistics. (2010). The Condition of Education 2010 (NCES 2010-028).

Wang, X. (2010). Factors Contributing to the Upward Transfer of Baccalaureate Aspirants Beginning at Community Colleges. Wiscape working paper.

Wine, J., Janson, N., and Wheeless, S. (2011). *2004/09 Beginning Postsecondary Students Longitudinal Study (BPS:04/09) Full-scale Methodology Report* (NCES 2012-246). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved June 20, 2012 from http://nces.ed.gov/pubs2012/2012246_1.pdf

Yin, R. K. (2009*). Case Study Research: Design and Methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Appendices

A: Recipient Survey
B: Survey Items Mapped to Sources and Research Questions
C: PI Survey
D: Interview Protocols
E: Sampling Plan and Estimates of Differences
F: Recruitment and Reminder Materials
G: Non-response bias analysis
H: S-STEM Program Logic Model
I: 60-day Federal Register Notice