
OMB SUPPORTING STATEMENT

PART B: COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

In this document, the Department of Labor (DOL) requests clearance from the Office of Management and Budget (OMB) under the Paperwork Reduction Act (PRA) for **a new collection** associated with the Apprenticeship Evidence-Building Portfolio. The Chief Evaluation Office of the U.S. Department of Labor (DOL) commissioned the Apprenticeship Evidence-Building Portfolio evaluation contract to build the evidence on apprenticeship, including apprenticeship models, practices, and partnership strategies in high-growth occupations and industries.

We discuss here seven different instruments in two separate sections. The first section (B.I) addresses the two instruments that are part of an impact evaluation to examine the effectiveness of Scaling Apprenticeship and Closing the Skill Gaps grants on participants' outcomes, such as employment earnings and career advancement.

1. A baseline survey for program participants
2. A consent form for program participants

The remaining five instruments are discussed in section two (B.II) and are part of two implementation studies: implementation study of the Scaling Apprenticeship, Closing the Skill Gaps, and other similar DOL initiatives to develop typologies of apprenticeship models and practices, identify promising strategies across the portfolio, and to better understand the implementation of models to help interpret impact evaluation findings; and an implementation study on the VETS Apprenticeship pilot to understand service delivery design and implementation, challenges, and promising practices.

3. An interview guide for program staff
4. An interview guide for program partners
5. A focus group guide for program participants
6. An interview guide for military apprenticeship placement counselors
7. An interview guide for military participants

Section B.I Impact Evaluation Data Collection

A major component of this evaluation is to develop an impact design option to rigorously evaluate the effectiveness of apprenticeship models and partnerships, based on the Scaling Apprenticeship Through Sector-Based Strategies grants (Scaling Apprenticeship grants), Closing the Skills Gap grants, and other relevant apprenticeship investments. DOL contracted with the Urban Institute in partnership with Mathematica and Capital Research Corporation to conduct the evaluation. This stage of the impact evaluation will include two data collection instruments, a baseline survey and consent form. Study participants will be potential apprenticeship candidates seeking apprenticeship program services from the grantee organizations and their subgrantees and partners.

B.1.1. Respondent universe and sampling methods

The respondent universe for the baseline survey and consent form are all grantees (and associated sub-grantees) of the Scaling Apprenticeship grants, and all grantees (and associated sub-grantees) for the Closing the Skills Gap grants (yet to be awarded). A Randomized Controlled Trial (RCT) design to estimate program impacts with control groups or enhanced treatment groups is planned. If the design changes, a change request package will be submitted to OMB for review. The study will include a purposeful sample of up to 10 sites from this universe of potential grantees and 5,000 program participants split evenly between the research groups. This design is intended to rigorously address the following research questions to be honed during the ongoing knowledge development phase of the study:

1. In the study sites, what are the impacts of apprenticeship models, components, and/or strategies on apprentices' employment, earnings, career advancement?
2. In the study sites, what are the differential impacts for underserved populations such as veterans, those with disabilities, women, and people of color?
3. In the study sites, what are the proximate impacts of intervening strategies, and job features (e.g., supervisory/mentoring strategies, method of training, length of training) that may be related to employment outcomes (e.g., retention and completion in the program, receipt of industry-recognized credentials, continuous employment, workplace benefits)?

a. Selection of sites

The objective of site selection is to identify up to 10 sites that are deemed to be suitable candidates for participating in a random assignment evaluation to address the study research questions. The study will purposively select the sites based on factors related to what can be learned from them and the feasibility of implementing random assignment. The study will consider five factors in determining a site's suitability for participating in the random assignment study:

1. **Sufficient sample size for estimating impacts.** To ensure a study sample size to yield sufficient statistical power to detect impacts, each participating site must be able to recruit and enroll a sufficient number of participants. Clarifying discussions with sites (discussed further below) will record the planned enrollment and the expected intake period. The study will also assess whether sites have the ability to recruit additional participants to fill a control group.
2. **Implementation status and readiness for evaluation.** A second factor is that the site's program services are of sufficient high quality and maturity, include all features required by the original Solicitations for Grant Application, and are consistent with the literature on effective apprenticeship services. To help maximize what can be learned from the study, the study will prioritize grantees that are planning to use innovative apprenticeship and work-based learning models.

-
3. **Service differential between the contrasted groups.** A third factor is the differential between the apprenticeship-related services provided to the research groups. Control group members denied apprenticeship services will have access to other services available in the community (and perhaps even other services from the same grantee). Thus, it is critical a sufficient differential exists between the tested program services and those available elsewhere in the community. Similarly, for designs testing enhanced apprenticeship services, it will be critical to ensure that these enhanced services are sufficiently different from business-as-usual services to have a reasonable chance of yielding detectable program impacts.
 4. **Similarity of services and point of random assignment across study sites.** A fourth factor is the ability of participating sites to implement a relatively consistent point of random assignment (for example, at the community college admissions office) and deliver a relatively similar set of intervention services, thereby ensuring that the impact analysis (which may need to pool across grant programs because of sample size considerations) tests a consistent model with focused research questions. Variation across sites in services or points of random assignment can pose problems for the analysis because a pooled impact analysis effectively would treat them as the same program even if they actually vary substantially in the nature of their services or in how or when a worker is defined as a study participant.
 5. **Appropriateness of implementing random assignment.** A final factor is consideration of the feasibility of implementing random assignment. In some cases, sites may operate with established referral networks from employers or elsewhere that could prohibit the use of random assignment. In these cases, programs could find themselves in conflict with partners who simply refuse to participate in such a study. This could occur, for example, if the study focuses on incumbent workers or special populations with control groups. During site selection, the study will focus on the sites in which random assignment is more feasible and does not threaten the program's continued operations and recruitment sources.

The study will rate each site using these five criteria during a multi-step site selection process. First, the study will begin with a systematic examination of extant materials on all grantees including the grantee applications and progress reports. It will use a template for recording this information and the results will be recorded directly into a database.

This information gathering effort will continue with clarifying phone calls to all grantees, excluding sites who would clearly not be suitable for the study (for example, because of very few expected program participants). These calls will focus on clarifying services provided by the programs and the process by which participants are recruited and enrolled into the program. Following each call, a standardized write-up on the grantee will be completed, as well as updates to the study's grantee database.

The study will identify approximately 20 sites with the highest ratings in the five considered categories. Formal recruiting of these sites will then occur by phone, and the study will start the

process of tailoring random assignment procedures to fit the sites' contexts, and also develop Memoranda of Understanding between the sites and DOL and the evaluation team. As needed, the evaluation team will conduct in-person visits with up to 5 sites to further assess the feasibility of implementing random assignment and to learn more about their services and processes. It is expected that this process will yield up to 10 suitable grantees for the study.

b. Selection of participants

All participants who meet the program eligibility requirements and consent to be part of the study will be subject to random assignment. According to the Scaling Apprenticeship grant applications, almost half of grantees were funded at a level associated with serving at least 5,000 apprentices (11 out of 23). The remaining 12 grantees were evenly split between a minimum requirement of 800, 1,600, and 3,200 apprentices. The working assumption is that the Closing the Skills Gap grants will plan to serve similar numbers of participants.

Not all grantee participants will be part of the study population to address the specific research questions. For example, the study population may focus on incumbent workers or special populations. Further, based on actual counts from previous DOL-funded apprenticeship grant programs, the sample size targets of the grantees are ambitious. Thus, the study conservatively assumes an average of 500 eligible program applicants per site who will be subject to random assignment, yielding a respondent universe of 5,000 participants split evenly between the research groups. Universe and sample size estimates are provided in Table B.1.

Table B.1. Summary of universe and sample counts

Evaluation component	Universe Description	Estimated Size of Universe	Expected Sample Size	Sampling method
Scaling Apprenticeship grants				
Grantees	All grantees and associated sub-grantees of the Scaling Apprenticeship grants	23	5	Purposive
Participants	All eligible applicants within selected grantees	2,500	2,500	Universe of all eligible applicants
Closing the Skills Gap grants				
Grantees	All grantees and associated sub-grantees of the Closing the Skills Gap grants	28	5	Purposive
Participants	All eligible applicants within selected grantees	2,500	2,500	Universe of all eligible applicants

c. Response rates

Applicants eligible for study participation will only be enrolled in the study and randomly assigned if they complete the baseline survey and provide their identifying information as part of

the intake process. Therefore, the project team anticipates that 100 percent of study participants will provide these data.

B.1.2. Procedures for the collection of information

The evaluation team anticipates starting participant intake, randomization, and baseline data collection in April 2020. Grantee staff will use RAPTER® to conduct participant intake. RAPTER® is a secure, web-based system that program staff will use to administer consent to participants, collect their identifying and contact information, and conduct random assignment of study participants. Participants completing the 15-minute baseline survey via the web or program staff entering baseline survey information on behalf of participants will use the RAPTER® interface to complete baseline information, which will also be completed online. The evaluation team will program RAPTER® to conduct random assignment within strata to ensure key population subgroups (such as special populations and those targeted for specific occupations) are balanced across the research conditions to improve precision of the impact estimates.

a. Estimation procedures

With an experimental design, unbiased impact estimates can be obtained by comparing differences between the mean outcomes of the contrasted research groups. By using regression procedures that control for highly predictive covariates, however, the study will improve the precision of estimates and adjust for small baseline differences between groups that may arise by chance or from survey nonresponse or missing administrative records data. The study will estimate impacts not only for the full sample, but also for important subgroups defined by participant and program characteristics from the grant application and baseline survey. The analysis will be conducted using the RCT-YES software program (www.rct-yes.com) that uses state-of-the-art design-based impact estimators derived from the building blocks of experiments with minimal assumptions, and can estimate impacts for continuous, binary, and discrete outcomes.

Assessing baseline equivalence. Using data from the program application and baseline surveys, the study will conduct t-tests on each baseline measure in isolation to examine differences between the research groups due to random sampling. We will also conduct a joint F-test to assess the joint significance of the baseline differences. The analysis will control for baseline characteristics, correlated with the outcomes, to improve the precision of the estimates.

Estimating impacts for the full sample. Assuming 10 sites, the benchmark model will be a regression in which an impact is calculated for each site, adjusted for students' baseline demographic characteristics from the program application and baseline information forms:

$$(1) y_i = \sum_{k=1}^{10} \beta_k * Block_{i,k} + \sum_{k=1}^{10} \delta_k * Treat_i * Block_{i,k} + \gamma * X_i + \varepsilon_i,$$

where y_i is the outcome of worker i ; $Block_{i,k} = 1$ for a worker in site k and 0 otherwise; $Treat_i = 1$ for treatment students offered program services and 0 for controls; X_i are baseline characteristics; ε_i is the error term; and β_k , δ_k , and γ are parameters to be estimated. We will

select the baseline covariates that are correlated with the outcomes using Least Absolute Shrinkage and Selection Operator (lasso) procedures (Tibshirani, 1996; Hastie et al., 2009) that avoid model overfitting.

The average impact of the tested intervention across grant programs is $\sum_{k=1}^{10} \delta_k / 10$. The study will assess differences in impacts across sites using a joint F-test of the site-level impacts (δ_k) and by comparing them to each other. The study will also explore the extent to which the results are sensitive to different weighting schemes, where for example, each sample member is weighted equally, or each site is weighted according to the size of the program eligible population. All weighting schemes are valid approaches but will provide slightly different estimates if grantees are of different sizes and have heterogeneous impacts. The study will account for missing data on baseline covariates using multiple imputation procedures with chained equations and predictive mean matching.

The study will interpret the impact estimates by conducting both classical significance testing and a Bayesian approach, where the study will report the probability that the intervention had positive effects given our findings (a Bayesian posterior probability). The Bayesian approach reduces the chance of misinterpreting p -values and statistical significance findings while providing credible, understandable assessments of program effectiveness.

Estimating impacts for subgroups. These same analytic methods for the full sample can be used to obtain impact estimates for two types of subgroups to address the question of whether access to grantee services is more effective for some subgroups than others. First, the study will estimate impacts for subgroups defined by worker characteristics (for example, age, prior employment experiences, special populations) defined from the program application and baseline surveys. Second, the study will estimate impacts for subgroups defined by key program features obtained from the implementation analysis (separate ICR).

Impacts for subgroups will be estimated using a straightforward modification to Equation (1), where the model includes terms formed by interacting subgroup indicators with the treatment status indicator variable and using F -tests to assess whether differences in impacts across subgroup levels are statistically significant. In addition, for the second set of program-related subgroups, Equation (1) can be reformulated to be the first level of a multi-level model or hierarchical linear model (HLM). In the second level, the study can regress the estimated impact of each site (δ_k) on key program characteristics obtained from the implementation analysis, including key components, practices, and policies, to help identify promising innovations for future development of apprenticeship programs.

Accounting for survey nonresponse. A future OMB package will request clearance for the 12- and 30-month follow-up surveys for the evaluation. However, baseline survey and program application data will be used to assess and correct for potential follow-up survey nonresponse, which could bias the impact estimates if outcomes of survey respondents and nonrespondents differ. To assess whether survey nonresponse may be a problem for the follow-up survey, two general methods will be used:

-
- *Comparing the baseline characteristics of survey respondents and nonrespondents for each research group.* Baseline data (which will be available for the full research sample) will be used to conduct statistical tests (chi-squared and *t*-tests) to gauge whether those in a particular research group who respond to the interviews are fully representative of all those in that research group. Noticeable differences between respondents and nonrespondents could indicate potential nonresponse bias.
 - *Comparing the baseline characteristics of respondents across research groups.* Tests for whether the baseline characteristics of respondents across the research groups differ from each other will be conducted. Noticeable differences between respondents in different research groups could indicate potential nonresponse bias and limit the internal validity of the study if not taken into account.

Two approaches for correcting for potential nonresponse using the baseline data will be used in the estimation of program impacts based on follow-up survey data. First, adjustments for any observed baseline differences between respondents across the various research groups will be performed using regression models. Second, because this regression procedure will not correct for differences between respondents and nonrespondents, sample weights will be constructed so that weighted observable baseline characteristics are similar for respondents and the full sample that includes both respondents and nonrespondents. For each survey instrument, weights will be constructed for each research group separately, using the following two steps:

- *Calculate a propensity score for each sample member that measures the probability of survey response based on baseline variables.* The study will use machine learning methods to construct the propensity scores using data from the baseline survey and program application forms. Specifically, the study will use the Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) software to estimate propensity scores (Griffin et al. 2014). The TWANG algorithm produces propensity scores based on the nonparametric splits of the baseline variables according to the created trees. Individuals with large propensity scores are likely to be survey respondents, whereas those with small propensity scores are likely to be survey nonrespondents.
- *Construct nonresponse weights using the propensity scores.* The weight for a sample member will be inversely proportional to the person's estimated propensity score. These weights will then be used in the impact analysis. Thus, the weighted observed characteristics of respondents should be similar, on average, to the observed characteristics of the entire research sample.

Assessing and correcting for grantee nonparticipation. As part of the recruitment process, the study will collect data on key grantee characteristics and compare the characteristics of the selected sites that agree to participate to those that do not. This information will be used to help interpret the analysis findings. However, because the study will not randomly select sites, but rather, purposively select them based on their suitability for the study, there is not a well-defined universe of sites to which the study sites will generalize. Thus, our benchmark approach will not adjust the impact estimates for site nonparticipation, because external validity is not a well-

defined concept for this evaluation. However, the study will re-weight the data for sensitivity analyses.

Adjusting for no-shows and crossovers. In any experiment in the real world, some members of the treatment group may not receive intervention services (no-shows), and some controls may be exposed to the interventions (crossovers). To correct for these sample members, the study will use an instrumental variable approach by replacing the $Treat_i$ indicator in the models above with the indicator variable, $Part_i$, that equals 1 for those who received intervention services and 0 for those who did not, and the study will use $Treat_i$ as an instrument for $Part_i$.

b. Statistical Power

To adequately address the evaluation’s research questions, the design must have sufficient statistical power to detect impacts that are policy relevant and of practical significance. The sample sizes needed for the study were determined by focusing on minimum detectable impacts (MDIs) for the primary outcome of quarterly earnings but the study also present MDIs for completion of an apprenticeship program, a key proximal (mediating) outcomes. Enrolling 5,000 (split evenly between the two research groups) in an RCT would enable us to detect MDIs of \$246 on quarterly earnings and 4.0 percentage points on apprenticeship program completion (Table B.2). This is smaller than the gains from participation in apprenticeship programs found in other studies. For example, Reed et al. (2012) found that participating in Registered Apprenticeship was associated with a gain of \$6,595 in annual earnings (\$1,649 in quarterly earnings) compared to the earnings of nonparticipants. This \$1,649 earnings gain is also larger than our calculated \$779 MDI for a 10 percent subgroup analysis based on 500 participants. For a design comparing an enhanced-service treatment group to a business-as-usual treatment group, it is expected that the study will have sufficient power to detect likely program effects if the enhanced services are intensive (for example, providing intensive case management and supportive services).

Table B.2. Minimum detectable impacts on key outcomes for an RCT

Sample size (treatment and control)	Quarterly earnings (impact, dollars)	Apprenticeship program completion (impact, percentage points)
100	1,755	28.3
500	779	12.6
1,000	550	8.9
5,000	246	4.0

Notes: Calculations above were made using Microsoft Excel tables. Assumptions made include: individuals are randomly assigned; equal assignment probabilities to treatment and control; 50% control group mean for completion; \$3,102 standard deviation of earnings; covariates explain 20% of the variation in outcomes; attrition of 20% in survey and NDNH data; alpha level 0.05, two-sided test, 80% power. The MDIs are calculated using the

following formula: $MDI = Factor(\alpha, \beta, df) * standard\ deviation\ of\ outcome * \sqrt{(1 - R^2) * (\frac{1}{N_T} + \frac{1}{N_C})}$,

where $Factor(\alpha, \beta, df) = t^{-1}(\alpha, df) + t^{-1}(2 * (1 - \beta), df)$ where α is the significance level (0.05), β is the power (80%), $t^{-1}(A, df)$ is the inverse of the student's t distribution function evaluated at A with df degrees of freedom, and df is equal to the sample size (after accounting for attrition) minus 2. R^2 is the proportion of variation in the outcome explained by covariates (20%), and N_T and N_C refer to the sample size in the treatment and control group, respectively, after accounting for attrition.

c. Statistical methodology for sample selection

All participants who meet the program eligibility requirements and consent to be part of the study will be subject to random assignment. Stratified random assignment will be conducted online using RAPTER® with pre-specified random assignment strings, developed separately for each sample intake location. Strata will be formed using information from the baseline survey and program application forms to ensure the research groups are balanced along key dimensions such as age, special populations, and those targeted for specific occupations.

B.1.3. Methods to maximize response rates and minimize nonresponse

The study is expected to collect baseline information from 5,000 participants in up to 10 grantee sites. In each site selected for the evaluation, grantee staff will administer the consent form to participants. If the participant consents to participate in the evaluation, staff will enter the person's identifying and contact information into RAPTER®, a secure, web-based system, to make sure the participant has not been enrolled in the study. New study enrollees will then be asked to respond to the baseline survey. The study team will make use of survey methods and best practices to encourage high response rates while minimizing burden and non-response. These methods include:

Web administration. It is anticipated that most respondents will prefer to complete the survey online. This choice allows the respondent to complete on their own schedule and pace, as well as complete the survey over multiple sessions. The web survey system used by the data collection team also supports mobile browsers, such as tablets or cellular phones.

Multiple modes of administration. Program staff will either administer the baseline survey or the participant will self-administer the survey online. To comply with Section 508 of the Rehabilitation Act, participants who may have difficulty completing a web survey will be offered the option of completing the survey by telephone.

Technology to reduce burden. To reduce burden, the baseline survey will employ drop-down response categories so respondents can quickly select from a list, dynamic questions and automated skip patterns so respondents only see those questions that apply to them (including those based on answers provided previously in the survey), and logical rules for responses so respondents' answers are restricted to those intended by the question. These features should minimize data entry burden by participants and facilitate high quality responses.

Use a tested questionnaire. The collection of baseline data has been tailored to the specific circumstances of this evaluation, yet is based closely on prior baseline surveys. These include the America's Promise Job Driven Grant Program Evaluation (OMB control number pending), the

Parents and Children Together Evaluation (OMB control number 0970-0403), the Evaluation of the Supplemental Nutrition Assistance Program (SNAP) Employment and Training Pilots baseline survey (OMB control number 0584-0604), and the Trade Adjustment Assistance Evaluation (TAA) (OMB control number 1205-0460). These prior instruments were extensively tested using interviews, cognitive interviews, or debrief sessions under each of these evaluations with populations that are similar to this study. These populations include active participants of SNAP employment and training services, individuals receiving economic stability services funded by Responsible Fatherhood or Healthy Marriage grants, individuals who received a Trade Readjustment Allowance payment to cover employment/retraining services, and America's Promise grantees. In addition, the Parents and Children Together Evaluation, the Evaluation of the Supplemental Nutrition Assistance Program Employment and Training, and the Trade Adjustment Assistance Evaluation (TAA) have successfully fielded their instruments.

Methods to ensure data reliability. The study will use several well-proven strategies to ensure the reliability of the survey data. The survey will be extensively reviewed by project staff, staff at DOL, and a Technical Working Group (TWG). These steps have been taken to make the questions as simple and straightforward as possible while targeting a consistent interpretation across participants and grantees.

The online baseline will include checks to prevent outlier entry. Additionally, to ensure that respondents answer questions, all respondents are informed of the privacy of their responses and that reports will never identify a respondent's specific data and that respondent names will remain private.

The evaluation team will train and provide written materials to all program staff on administering consent, collecting and identifying contact information from participants, and administering the baseline survey.

Data from completed baseline surveys will be reviewed throughout the fielding period for accuracy and consistency. To prepare survey data for analysis, the study will run data checks, examine frequencies and means, and assess the extent of missing data. Participants that do not complete the baseline survey at all will be excluded from the analysis.

B.1.4. Tests of procedures or methods to be undertaken

All data collection procedures and instruments included in this request to be used in the evaluation have been reviewed by content and methodological experts to ensure clarity and optimal ordering of the questions.

Just as the instruments for the baseline survey are based closely on prior surveys that have been extensively tested to evaluate the clarity of the questions to be asked, to identify possible modifications to either question wording or question order that could improve the quality of the data, and to estimate respondent burden (B.I.3), the procedures used to collect the data will be based closely on the procedures used successfully for similar surveys, which ensures that they can be used effectively to conduct the data collection for this study..

B.I.5. Individuals consulted on statistical aspects of design and on collecting and/or analyzing data

Consultations on the methods used in this evaluation are part of the impact design options phase of the project to ensure technical soundness. Consultations include stakeholders within the U.S. Department of Labor and a TWG. Members of the study team and the TWG are listed in Table B.3.

Table B.3. Individuals who were consulted for the Apprenticeship Evidence-Building Portfolio

The Urban Institute	Barbara Butrica Pamela Loprest Project Directors Demetra Nightingale Co-Principal Investigator Daniel Kuehn Deputy Project Director William Congdon Senior Advisor, Methods Robert Lerman Senior Advisor, Apprenticeship
Mathematica	Peter Schochet Co-Principal Investigator Samina Sattar Project Director Annalisa Mastri Quality Control Advisor
Capital Research Corporation	John Trutko Project Director
Technical Work Group Members	Carolyn Heinrich Patricia and Rodes Hart Professor of Public Policy, Education, and Economics, Vanderbilt University Susan Helper Frank Tracy Carlton Professor of Economics at the Weatherhead School of Management, Case Western Reserve University Chris Magyar Chief Apprenticeship Officer, Tectonic Inc. Mary Alice McCarthy Director of the Center on Education & Skills, New America Jeffrey Smith Paul T. Heyne Distinguished Chair in Economics and Richard Meese Chair in Applied Econometrics, University of Wisconsin-Madison

Staff responsible for overseeing the collection and analysis of data are listed in Table B.4.

Table B.4. Individuals who will oversee the collection and analysis of data for the Apprenticeship Evidence-Building Portfolio

The Urban Institute	Barbara Butrica Pamela Loprest Project Directors Daniel Kuehn Deputy Project Director Lauren Eyster Shayne Spaulding Task Directors
Mathematica	Samina Sattar Project Director Ryan Callahan Survey Director
Capital Research Corporation	John Trutko Project Director

Section B.II Implementation Evaluation Data Collection

The Apprenticeship Evidence-Building Portfolio includes two implementation studies: an implementation study of the Scaling Apprenticeship, Closing the Skill Gaps, and other similar DOL initiatives to develop typologies of apprenticeship models and practices, identify promising strategies across the portfolio, and to better understand the implementation of models to help interpret impact evaluation findings (labeled below as the Program Implementation Study); and an implementation study on the VETS Apprenticeship pilot to understand service delivery design and implementation, challenges, and promising practices. In this section, we describe the data collection for each of these in turn.

Section B.II.1 Respondent Universe and Sampling

Program Implementation Study. The potential universe for site visits as part of this implementation study is 23 Scaling Apprenticeship Grants and 28 Closing the Skill Gap grants. From this universe, 21 sites will be visited for three- to four-day site visits (Table B.5) The site visits are designed to provide in-depth information about a group of grantees with a range of characteristics and apprenticeship program designs, including grantees whose programs are part of the impact evaluation. The site visit information will be used to inform development of typologies of types of apprenticeship programs and identify promising strategies. All reports including these findings will make clear that the results show not be generalized to all programs receiving these apprenticeship grants or all apprenticeship programs. In addition to including sites that are part of the impact evaluation, suitable sites for these visits will be identified using the following primary criteria:

- Grant type, with at least 9 site visits allocated to each grant program (Scaling Apprenticeship and Closing the Skill Gap);
- Models, strategies and practices being implemented;
- Industry focus of the grant (to include a mixture of sectors, such as IT, health care, advanced manufacturing);
- Organization type (for Closing the Skill Gap);
- Target populations to be served;
- Geographic area served by the grant.

Data sources for site selection information include grantee applications, workplans and quarterly reports, and notes from clarification calls made to a select number of grantees as part of the impact study design process. A list of alternative sites, having similar characteristics, will be identified in the event that the grantee is unable to participate in the data collection.

The potential respondents at each selected site include staff from the grantee organization and partner organizations, including employers. For each of the 21 sites selected, the member of the two-person site visit team primarily responsible for logistics will make initial contact by phone with the individual listed as the primary contact in the Office of Apprenticeship records. The site visit team will then send an e-mail to inform the grantee organization of the study and request its cooperation. The initial telephone contact will provide background about the project and seek additional information on organizations and partners in order to identify key respondents. Based on this information, the site visit team will contact respondents and determine the best timing for the visit in order to accommodate the schedule of local respondents.

At the outset of the interviews, we will inquire about environmental context for the grant, key features of the program, including apprenticeship and pre-apprenticeship programs, program organization and structure, role of key partners, overlap with other apprenticeship initiatives, target population and recruitment practices, employer engagement, sustainability of grant activities, and key lessons learned.

Table B.5. Respondent Universe for Site Visits Implementation Study

Data Collection Activity	Universe/Sampling Frame	Respondent Description
Site visits: Semi-structured interviews	Universe of 51 grant recipients/purposive sampling of 21 grantees that represent a range of grantee characteristics and program design features, including grantees in impact study	Grantee program coordinators, other program staff, employer partners
Site visits: Focus groups	Universe for each site ranging from 800 to 5,000 participants over course of program; contact for focus group participation made in 21 grant sites. Convenience sample of 210 students willing and able to attend the focus groups.	Apprentices participating in Scaling Apprenticeship or Closing the Skill Gaps programs

The study will also conduct one focus group of program participants at each site visited, where feasible. We will first work with grantees to select several programs or one program, depending on proximity and numbers of participants, where it will be possible to hold a single focus group. We will then obtain a list of program participants with their contact information from the grantee, and then send apprentices recruitment emails and if necessary, conduct follow-up phone calls. If the grantee prefers to have participants contacted by program staff, we will provide the recruitment materials to facilitate outreach efforts. Outreach for each grantee will be conducted until 10 participants accept the focus group invitation in each site.

For the focus groups, we will ask about participants' reasons for enrollment, how they were recruited, their educational and employment aspirations, perspectives on key features of the apprenticeship, credential attainment and key outcomes, and perspectives on program.

VETS Apprenticeship Pilot. Interviews will be conducted with military apprenticeship placement counselors at each of the 8 sites implementing the pilot. In some sites we will include two placement counselors and in some sites one, reflecting the number of placement counselors at each base.

The interviews will inquire about the vision for program, organizational structure of the pilot, apprenticeship placement counselors' role and services provided, key partners in the pilot and their expected and actual role, implementation challenges, participant flow through the program, their view of participants' expectations and motivations for participation, perspective on program outcomes and promising strategies.

The study will also interview participants in the pilot program. Participants will be interviewed from three bases. These bases will be selected from the 8 participating sites to reflect diversity in terms of geography, the scale of programs, and branches of the military represented. The number of participants at each site ranges from 24 to 72, with 383 expected participants in total. We aim to interview five service members from each of the three selected bases. We will select a set of 20 participants from each of the three bases randomly using data from the case management system, but then select among these to make sure we have a mix of participants based on gender, rank, and apprenticeship sector of interest/placement. We expect that some participants will not respond to our calls/emails, so we will plan to reach out to as many as needed to get to our targeted number of interviews. Once participants are identified, we may seek assistance from apprenticeship placement counselors in getting responses for interviews. We will make clear in any public report that findings based on these interviews are suggestive of participant experiences and perspectives but may not be representative of all pilot participants.

These interviews will gather information on how participants became interested in the program, the types of information and assistance they received, how participation has helped (or not helped) them learn about apprenticeship opportunities, gain placement in an apprenticeship, upgrade skills, improve earnings, and move along his/her career pathway, and if they have been placed in apprenticeship their views on the experience. Interviewees will receive a \$25 gift card as a thank you gift for participation.

B.II.2. Procedures for the collection of information

The data for both implementation studies will be collected through semi-structured interviews of program staff or partners held at selected grantees and military bases, focus groups of apprentices participating in grant programs, and phone interviews with selected participants. The interview topic guides are provided in Attachments D and E for implementation program study staff and partners, F for focus group guide for program participants, Attachment G for military apprenticeship counselors, and Attachment H for semi-structured phone interviews with participants.

Since these are qualitative studies of program implementation no statistical methods will be used to sample respondent populations. No statistical methods will be used to select the grantee organizations for the program implementation site visits as the sample is intended to be neither random nor representative. All VETS Apprenticeship Pilot sites will be included. The information collection for the implementation studies is designed to provide in-depth qualitative information about grantees and information suggestive of participants experiences and perspectives. No estimation procedures will be used. The data analysis will be descriptive. The site visits, focus groups, and participant interviews are one-time data collection efforts.

B.II.3. Methods to maximize response rates and minimize nonresponse

For both implementation studies, we expect that all program staff identified for interviews will agree to participate. For the Program Implementation Study, we will work with the primary contact person for the grant to identify appropriate staff and partners and to schedule interviews. We will use program staff to provide introduction to partner employers and discuss importance of the study to increase partner participation in interviews. In the VETS Apprenticeship Pilot we will interview all apprenticeship placement counselors. Should a potential respondent not be available during a visit, the research team will follow up with a time to interview the person by phone. For focus groups, we will recruit participants ahead of time, providing them with information on the importance of the study and confidential nature of their responses. We will continue recruitment until we have acceptances from 10 participants. For participants, we will work with apprenticeship placement counselors to introduce the study and invite selected participants to take part in interviews. We will send prospective participants information on the study prior to the interview to encourage participation.

B.II.4. Individuals consulted on statistical aspects of design and on collecting and/or analyzing data

Staff responsible for overseeing the collection and analysis of data are listed in Table B.4.