CONTENTS

TABLES

The U.S. Department of Labor (DOL), Employment and Training Administration (ETA) is currently undertaking the Self-Employment Training (SET) Demonstration. This demonstration is a reemployment program targeted towards dislocated workers, as defined by the Workforce Investment Act (WIA), who are interested in starting or growing a business in their fields of expertise.[1] The demonstration relies on self-employment advisors who deliver intensive business development counseling with the goal of connecting such workers to self-employment training, technical assistance, and other services (including seed capital microgrants) to help them become more successful in self-employment. Enrollment in the demonstration began in July 2013 and will be open for a period of up to 31 months.[2] Each program participant will have access to SET services for up to 12 months, for a total implementation period of up to 43 months.

The main objective of the evaluation of the SET Demonstration (the SET Evaluation) is to understand whether providing dislocated workers with access to intensive business development services and self-employment training increases their likelihood of reemployment, their earnings, and their propensity to start a business. The evaluation will use a rigorous experimental design in which approximately 3,000 applicants to the program in four metropolitan areas are randomly assigned to a program group or a control group with equal probability. An impact analysis will assess the SET Demonstration's effects on outcomes measured approximately 18 months after randomization. An implementation study will also be conducted to provide information about the training experiences of applicants that will help ETA further refine the self-employment services made available to dislocated workers and other customers of the workforce system. Additional information about the program model and the research questions that will be examined in these two components of the study is included in Part A.

[1] To receive training services under Title I of WIA, a dislocated worker is an individual who (1) (A) has been terminated or laid off or has received a notice of termination or layoff from employment, and (B) (a) is eligible for or has exhausted unemployment insurance or (b) has demonstrated an appropriate attachment to the workforce, but is not eligible for unemployment insurance, and (C) is unlikely to return to a previous industry or occupation; (2) has been terminated or laid off or received notification of termination or layoff from employment as a result of a permanent closure or substantial layoff, or is employed at a facility where the employer has made the general announcement that the facility will close within 180 days; (3) was self-employed but is unemployed as a result of general economic conditions in the community or because of a natural disaster; or (4) is a displaced homemaker who is no longer supported by another family member. Individuals are considered eligible for the SET Demonstration if they meet any of these four qualifications, irrespective of whether they register for staff-assisted services with a WIA American Job Center.

[2] Intake into the demonstration will proceed until the demonstration reaches its participation target (3,000 eligible applicants) across participating study sites, but no later than January 31, 2016.

The purpose of this is request is to ask that OMB grant clearance for an extension to the period of data-collection for the 18-month follow-up survey the SET Evaluation, one component of the data collection effort previously approved by OMB on January 30, 2013 (ICR 201209-1205-001; OMB Control Number 1205-0505). We specifically request to extend data collection for an additional 20 months, from the currently approved expiration date of January 31, 2016 through September 30, 2017. The follow-up survey is critical for gathering information needed for the impact analysis about the self-employment experiences and other labor-market outcomes of members of both the program and control groups. Given that study enrollment has proceeded more slowly than originally planned, an 18-month follow-up survey could be administered to only approximately 25 percent of the demonstration applicants by the current expiration date of January 31, 2016. Assuming a 70 to 80 percent response rate, this would result in approximately 525 to 600 respondents (= 3,000 respondents × [0.70 to 0.80 response rate] × 0.25 of study participants).[3] Extending the expiration date to September 30, 2017 will allow sufficient time to field the survey to all study applicants.

This request covers only the follow-up survey that was approved under ICR 201209-1205-001. No changes are proposed to the other four data-collection efforts previously approved and being conducted (consent and application forms, the program participation records, or the evaluation team's site visit and case study protocols) through the original information clearance request. Those other four data-collection efforts will be completed within the originally planned timeframe.

1. Respondent Universe and Sampling

The SET Demonstration is being implemented in purposively selected study sites, in which recruitment will target dislocated workers likely to meet the study's eligibility criteria (described below).[4] The follow-up survey will be conducted with up to 3,000 individuals applying to the program and meeting the eligibility criteria—the application process will be closed once this target is reached.[5] Sample members are selected based on the factors described in

[3] Respondent burden is discussed in Part A of this information clearance request. Section B.3 below provides and extensive discussion of plans for maximizing response rates and addressing both individual-level and item nonresponse.

[4] The program is active in four metropolitan areas: (1) Chicago, Illinois (City of Chicago and Cook County); (2) Cleveland, Ohio (Cuyahoga and Lorain counties); (3) Los Angeles, California (Los Angeles city and county); and (4) Portland, Oregon (Multnomah and Washington counties). Within each catchment area, the evaluation team has selected strong partner microenterprise development organizations that have the capacity to deliver the services specified for the SET Demonstration over the duration of the program period.

[5] The purposive selection factors described in this section, in conjunction with self-selection of applicants to the demonstration based on an unknown mechanism, mean that the study population cannot be construed as being sampled from a larger target population with well-

the two subsection that follows. A second subsection describes the expected sample sizes for the evaluation's analyses.

Selecting the Study Population within Study Sites. Successful applicants to the SET Demonstration are dislocated workers who, at baseline, already have established behaviors suggesting that they will be responsive to and benefit from self-employment training.[6] To identify dislocated workers who are likely to benefit from the program, applications to the SET Demonstration are screened based on prior work experience related to the applicants' proposed business idea. Study recruitment occurs after potential applicants attend a mandatory online orientation session. The orientations explicitly state the demonstration's eligibility criteria and inform potential applicants that (1) applications not meeting the eligibility criteria will be screened out, and (2) meeting the eligibility criteria qualifies them only a 50 percent chance to enter the SET program, based on the outcome of the random assignment lottery.

The 3,000 individuals meeting the eligibility screens are randomly assigned at each to the program and control groups; this is done with equal probability. The program group (N ≈ 1,500) is eligible to receive services through the SET demonstration. The control group (N ≈ 1,500) is not eligible for such services. Both groups have access to other existing services available through American Job Centers (AJCs) and community providers of standard self-employment assistance and training. As noted in Part A, the evaluation team selected partner microenterprise development organizations (MDOs) that help support the integrity of the evaluation's control-group design by providing SET services only to the members of the program group for the duration of the program period.

Expected Sample Size for the Follow-Up Survey. The impact analysis will be based on outcomes data collected from follow-up surveys, currently being initiated with all study members who went through random assignment. Based on the experience of the contractor in fielding surveys for similar study populations, as well as in fielding the survey to date, it is expected that the response rate for the follow-up survey will be 70 to 80 percent. This will result in a sample size of 2,100 to 2,400 respondents.[7] This group of individuals will be referred to as the analysis sample. Section B.3 describes the statistical methods that the study team will use to analyze and

defined probabilities. As discussed in Section B.1.b, this implies that it will not be possible to draw statistical inference about any larger population than the respondents included in the demonstration.

[6] Part A of this clearance request provides additional information on the practical and research-based motivations for selecting potential participants based on related work experience, as well as discussion of options for implementing the screening criteria.

[7] A justification for this expected response rate is presented in Part A of this package (Section A.10).

potentially account for nonresponse bias by applying sampling weights to the analysis sample.[8]

2. Analysis Methods and Degree of Accuracy

The methods used for the implementation study and impact analysis are presented separately in two following subsections. The main research questions of each component of the SET Evaluation and the data used to answer them described more fully in Part A of this Office of Management and Budget (OMB) package.

a. Implementation Study

Information from the follow-up survey will be used, in combination with other data collected under OMB Control Number 1205-0505, to provide a the contextual understanding of the SET program in each study site. The main way in which data from the follow-up survey will be used in the implementation study is to provide information about the experiences of the program group with the SET Demonstration and the training and services they received—from both SET providers and other organizations providing self-employment services.

The analysis team will use this information conduct a cross-site analysis to describe common elements and differences across sites in the implementation of the SET program. The team will examine variations in services across study sites and characterize the degree to which there is fidelity to the model (high, medium, or low) in each site, using a predeveloped rating scheme. In addition, the service-use data will be used to identify factors or considerations that might help understand why the impacts of the SET program vary from one site to the next.

This quantitative component of the analysis will be done using simple descriptive statistics (such as means and percentages) and cross-tabulations. The statistical significance of differences in continuous measures will be calculated using *t*-tests that allow for unequal variances across the main contrast dimension (e.g. assignment status and site). Hypothesis testing for binary and categorical measures will be based on chi-squared tests and Fisher’s exact tests depending on the sample size and prevalence of each outcome category.

[8] Because applicants to the SET Demonstration are not recruited from a sampling frame with known probabilities (that is, applicants are self-selected from an unknown population), American Association for Public Opinion Research (AAPOR) guidelines would suggest using the rate of participation, rather than the rate of response, when describing the fraction of the original random assignment sample completing the follow-up survey; the latter term is typically associated with probability sampling (AAPOR 2011). However, the text of this OMB package submission will continue to use response and nonresponse to avoid confusion with participation in the SET Demonstration program by individuals who were randomly assigned to the program group.

b. Impact Analysis

The objective of the impact analysis is to provide statistically valid and reliable estimates of the effects of the SET Demonstration on the economic outcomes of the dislocated workers served by the program. As discussed further in Part A, the main research questions of the impact analysis are:

1. What is the net impact of the SET Demonstration program on participants' overall employment status and total earnings?
2. Does the SET Demonstration increase self-employment?
3. Does the SET program improve intermediate business development outcomes?
4. How does participation in the SET Demonstration affect job satisfaction and participation in other workforce programs?
5. Do program impacts differ for subgroups of participants defined by baseline characteristics?
6. Through what programmatic mechanisms might the SET Demonstration's program influence participant outcomes?

These questions will be answered using information about outcomes and program participation experiences only available from the follow-up survey to which the requested extension applies. The analysis will also make use of the other baseline and contextual data collected for the evaluation under OMB Control Number 1205-0505. A classical experimental design, in which applicants are assigned randomly to program and control groups, will enable the evaluation team to calculate estimates of the causal impact of the SET program. The measured impacts will be internally valid for the four study sites. However, because the study sites will be chosen purposively and the pool of applicants to the demonstration will be self-selected and then purposively selected as a quota sample, the evaluation's results cannot be generalized to a wider population with a known degree of statistical precision.

A description of the study's outcome measures and discussion of the methods that will be used to estimate the program's impacts and compute variances for the point estimates follows, after which is a description of the expected precision of the estimates by characterizing the minimum detectable impacts (MDIs) of the program that are likely to be obtained using data from the follow-up survey.

Study outcome measures. The primary study outcomes to be examined in the impact analysis include the following:

1. Self-employment at the time of the follow-up survey
2. Employment in any job at the time of the survey

3. Total earnings during the one-year period between random assignment and the date of the survey

These outcomes will be used to summarize the effectiveness of the program. Measuring the program's impact on self-employment is an important goal of the demonstration because of the nature of services being delivered. Additionally, self-employment is of particular interest because of the autonomy that self-employed workers are expected to achieve. The other two primary outcomes—employment in any type of job and total earnings—capture the demonstration's *overall* success at helping participants become reemployed, which is the major objective of ETA for the SET Demonstration.

In order to better understand whether and how the SET program works, the evaluation will also consider how effectively it encourages participants to take steps associated with self-employment success. The study will specifically consider intermediate milestones such as whether participants were able to gain access to startup capital, register their businesses, and develop and complete a business plan. Additional, secondary outcomes that will be considered include: receipt of self-employment services; achievement of important intermediate business development milestones; earnings from self-employment and from wage/salary employment; receipt of unemployment insurance (UI) payments; and participation in other government programs; (See Part A for further details.) Exploratory analyses of these outcomes will seek to shed light on the mechanisms by which the SET program operates and the diverse set of effects the program might have. Further, as described in following sections, the exploratory analyses will seek to examine how program impacts vary across subgroups. Results from the exploratory analysis will be treated cautiously because of the large number of comparisons being made.

Calculating estimates of program impacts. Random assignment will enable estimation of the net impact of the SET Demonstration by comparing average outcomes across the program and control groups. These estimates will assess the impact of the offer of SET program services, rather than the impact of services received, as some individuals in the program group could chose not to use the business development counseling provided by the demonstration's self-employment advisors. In addition to capturing the direct effects of SET services, the impact estimates also implicitly measure the effects of differences in the quantity and quality of other self-employment services received, such as classroom training and one-on-one technical assistance, as a result of the SET program.[9]

[9] As discussed in Part A, one of the major functions of the self-employment advisor is to help SET participants identify and marshal the most appropriate and effective training resources that are already available in the community.

The core statistical approach for estimating net impacts predicts the outcome of interest as a function of program group membership, site, and a set of background characteristics. The general form of this model for a continuous outcome variable is

(1) $y_{is} = \beta_s p_i + \boldsymbol{\gamma}'\mathbf{x}_i + \boldsymbol{\alpha}'\mathbf{q}_{is} + \varepsilon_{is}$,

where y_{is} is the outcome of interest for individual *i* in site *s*, p_i is a binary variable indicating membership in the program group, and $\mathbf{x}_i$ is a vector of baseline characteristics of individual *i* measured before random assignment. The (4 × 1) vector $\mathbf{q}_{is}$ denotes a set of dummy variables for each study site—for individual *i* at site *s*, the *s*th element of $\mathbf{q}_{is}$ is equal to one and all other elements are equal to zero—and so $\boldsymbol{\alpha}$ represents a set of four site-specific intercept terms.[10] Finally, ε_{is} is an individual-level random error term that denotes the effects of unobserved factors that influence the outcomes. Because of the randomized design, the error term is expected to have a mean of zero within each site, conditional on the program assignment status of individual *i* (p_i). The main coefficient of interest in equation (1) is β_s, which measures the average effect of the SET Demonstration program on participants' outcomes at site *s*. Estimates of program effects using equation (1) are based on the *offer* of demonstration services and are estimated using all sample members in the program and control groups, irrespective of their actual utilization of SET services, in a classical intention-to-treat (ITT) framework.

For ease of exposition, the outcome variable is assumed to be continuous throughout this section. When considering binary outcomes, Equation (1) could be re-specified as a nonlinear probit or logit model. However, a regression coefficient from a linear probability model often provides a reasonable approximation to the marginal effect of a variable that would be obtained from a nonlinear binary response model (Wooldridge 2002). Because of its advantages for interpreting point estimates, the linear model will be used if the regression estimates are similar to the marginal effects obtained from the nonlinear model.

Point estimates. Equation (1) can be estimated using ordinary least squares (OLS) to obtain the estimated impact of the program at each sites *s*, $\hat{\beta}_s$ within the analysis sample (that is, the set of individuals that completed surveys). However, the goal of the evaluation is to draw inferences about the

[10] Sites will be purposively selected based on the criteria described previously, thus statistical inference will be valid for the set of study sites only and cannot be generalized to any broader population. Consequently, site-level intercepts will be specified as fixed effects, rather than random error components.

effects of the SET Demonstration on the full study population of individuals who were randomized at baseline. As discussed further in Section B.3, an analysis will be conducted to assess the extent to which there is the potential for nonresponse bias in the estimates obtained from the analysis sample. In the event that nonresponse adjustments are required, Equation (1) will be estimated using weighted least squares (WLS), with individual-level nonresponse factors used as the elements of a diagonal matrix of regression weights. Equations (4.10) and (4.31) in Cameron and Trivedi (2005) provide the formulas that will be used to calculate the OLS and WLS point estimates, respectively. Irrespective of the estimation technique, estimates of $\hat{\beta}_s$ will be reported separately for each site.

Combining estimates across sites. It is also reasonable to estimate a pooled effect of the program across all sites because each site will be asked to implement the same program model. In addition, one of the key criteria in selecting sites is that the AJC and MDO infrastructure is sufficient to effectively deliver the program. The estimated pooled effect ($\hat{\tau}$) is computed as a weighted average of the estimated effects in each site, where the weights are set equal to the proportion of the sample located in each site. That is,

(2) $\hat{\tau} = \sum_s f_s \hat{\beta}_s$.

Without nonresponse adjustments, f_s is equal to the fraction of the analysis sample from site *s*; when applying nonresponse adjustments, f_s is equal to the fraction of the baseline sample from site *s*. Because program assignment within each site will be independent of baseline characteristics, $\hat{\tau}$ will be approximately equal to what would be obtained by estimating a regression in which the impact of the program is constrained to be the same in every site. Thus, the pooled estimate $\hat{\tau}$ can be interpreted as the average effect of the SET Demonstration program across all sample members. Sensitivity analyses will consider whether results differ when sites are weighted equally or are weighted by the inverse of the site-specific variances when calculating the pooled estimate.

Covariates included in the regression. If random assignment has been properly implemented and there are no concerns about nonresponse, it is not strictly necessary to control for baseline characteristics ($\mathbf{x}_i$) in the regression. However, including these variables in the regression is advantageous because doing so will improve the precision of the estimated program effects. This occurs because baseline measures that are predictive of the sample members' outcomes will absorb some of the variability in the outcome measures, resulting in a greater signal-to-noise ratio when estimating the impact of the program.

In addition the model described by Equation (1), an alternative approach that will be considered is to allow the relationships between the baseline characteristics and the outcome (that is, the parameters in $\boldsymbol{\gamma}$) to vary across sites. This set-up could potentially improve the precision of the impact estimates, $\hat{\beta}_s$, because the baseline characteristics will be allowed to explain more of the site-specific variation in the outcome. However, this approach implies estimating a substantially larger numbers of parameters, leading to a smaller number of degrees of freedom, which could, all else being equal, reduce the precision of the impact estimates. Thus the net effect on precision of allowing the coefficients on the baseline characteristics to vary across sites is ambiguous. The study team will consider both approaches; the decision about which approach is preferred will be guided, in part, by information on the relationship between survey response rates and baselines characteristics. If that relationship differs substantially across sites, it could be advantageous to allow the coefficients on the baseline characteristics also to vary across sites.

Potential baseline characteristics that could be controlled for include measures of demographics (age, sex, race/ethnicity); family structure (marital status and number of dependents); education level; receipt of UI benefits at the time of random assignment; and baseline measures of employment status and earnings from both self-employment and wage/salary jobs. The specific characteristics included will be selected based on the substantive knowledge of the evaluation team or, alternatively, through a stepwise variable-selection procedure (Neter et al. 1996).

Subgroup analyses. Additional analyses will consider the extent to which the effects of the program differ across different groups of sample members defined by baseline characteristics and whether the effects differ according to specific contextual or programmatic factors measured at the site level. Subgroup impacts will be measured using a straightforward modification to Equation (1).

For ease of exposition, consider the case in which two subgroups of interest are defined by different levels of a binary variable d_i. For example, d_i could be set to one for individuals receiving UI benefits at baseline and to zero for individuals not receiving UI benefits. In this case, subgroup impacts would be estimated using the model

(3) $y_{is} = \beta p_i + \eta(p_i \times d_i) + \theta d_i + \boldsymbol{\gamma}'\mathbf{x}_i + \boldsymbol{\alpha}'\mathbf{q}_{is} + \varepsilon_{is}$.

Equation (3) differs from Equation (1) in two ways. First, an interaction term between assignment status and the subgroup indicator is included, $p_i \times d_i$. (For clarity, the uninteracted measure of subgroup membership, d_i, has been denoted separately from the other baseline characteristics, $\mathbf{x}_i$.). Second, the coefficients β and η are not allowed to vary across sites.[11] With this set-up, the average effect of the program on the subgroup for which $d_i = 0$ (for example, individuals not receiving UI benefits at baseline) across all sites is measured by β. The average effect of the program on the subgroup for which $d_i = 1$ (for example, recipients of UI benefits at baseline) across all sites is measured by $\beta + \eta$.

Potential subgroups of interest include those defined by the baseline characteristics controlled for in the regression, as discussed previously. In addition, subgroups can be formed based on different levels of contextual or programmatic factors particular to each site, in which case d_i would be replaced in Equation (3) with a site-level measure of those factors. Another potentially beneficial approach is to focus on UI recipients and form subgroups according to factors associated with their likelihood of exhausting their available benefits. As discussed in Part A, such an analysis might provide useful information to states interested in examining how Worker Profiling and Reemployment Services (WPRS) systems are used to identify candidates for new or existing Self-Employment Assistance programs. The specific subgroups analyzed will be determined by the contractor in conjunction with ETA based on findings from the implementation study, evidence from prior self-employment assistance demonstration projects, and results from other research on the correlates of success in self-employment (for example, Evans and Leighton 1998; Fairlie and Robb 2008).

Variance estimation. Because the SET Demonstration sites were chosen purposively and the study population will not be sampled probabilistically from a known population, inference will be limited to the baseline sample of individuals who went through random assignment in the four study sites. Therefore, variances can be straightforwardly estimated using fairly standard linear regression formulas. A Huber-White “sandwich” estimator will be used to account for potential heteroskedasticity of the error term (Huber 1967; White 1980). Asymptotic formulas for heteroskedastic-consistent estimates of the variance–covariance matrix for coefficients

[11] This simplifying decision was made because, based on sample sizes, it is not expected that site-specific subgroup differences can be measured with a reasonable degree of precision. Allowing β and η to vary across sites would also imply allowing the basic coefficient on the subgroup indicator, θ, to also vary across sites. This site-interacted specification would further reduce the precision of the subgroup impact estimates through a reduction in the number of degrees of freedom.

calculated using OLS and WLS are given by Equations (4.21) and (4.32) and the surrounding discussion in Cameron and Trivedi (2005). Estimated variances will be computed based on these formulas using a standard statistical package, such as Stata, that incorporates the scalar "HC1" degrees-of-freedom correction, described by McKinnon and White (1985), as a finite sample adjustment.

When conducting inference on the multisite pooled estimates, which is calculated as a sample-weighted average, the estimated variance of the pooled estimate will take into account the potential correlations among the site-specific estimates. Those correlations are non-zero when the coefficients on the baseline characteristics are constrained to be the same across sites. The variance formula for the pooled estimate given by Equation (2) is

(4) $$\hat{V}(\hat{\tau}) = \sum_{s}\left[(f_s)^2 \times \hat{V}(\hat{\beta}_s) + \sum_{s' \neq s}\left[f_s f_{s'} \times \hat{C}(\hat{\beta}_s, \hat{\beta}_{s'}) \right] \right],$$

where $\hat{V}(\cdot)$ and $\hat{C}(\cdot)$ represent estimated variances and covariances and f_s is as defined above.

Minimum detectable impacts. Table B.1 presents MDIs calculated for the three primary study outcomes measured: (1) self-employment at the time of the follow-up survey, (2) employment in any job at the time of the survey, and (3) average quarterly total earnings (from all sources) during the six quarters between random assignment and the survey. The MDIs have been calculated using the following assumptions:

- The level of statistical power is 80 percent and inference will be conducted using a two-tailed test with the significance level set to 5 percent.
- The overall prevalence of self-employment will be 40 percent, the prevalence of employment in any job will be 75 percent, and the standard deviation of quarterly total earnings will be $10,200.[12]
- At baseline, the sample members in each site are assigned with equal probability to the program or control groups.

[12] Because the pool of applicants to be included in this evaluation is expected to be more focused and more experienced than those in the Project GATE evaluation, the rate of self-employment is assumed to be slightly higher than what was seen in the 18-month follow-up for Project GATE for individuals who were unemployed at baseline. The rate of employment in any job is assumed to be approximately equal to the average of the 6- and 18-month rates for initially unemployed members of the Project GATE sample. Likewise, the standard deviation of total quarterly earnings is based on the average of standard deviation of total earnings since random assignment for the baseline-unemployed sample at the Project GATE 6- and 18-month follow-up surveys; this number is expressed in 2014 dollars. All of these estimated sample statistics from Project GATE are reported in Benus and Michaelides (2010, Tables 3 and 45).

- The response rate for the follow-up survey will be approximately the same in both groups.
- Baseline measures included in the regression explain 20 percent of the variance in the outcome.
- Point estimates are based on an unweighted regression.
- Variance estimates do not account for heteroskedasticity.

The final two assumptions were made so that an analytic expression for the MDI could be derived. Specifically, using formulas (1) and (5) from Schochet (2008), MDIs are calculated using the approximation:

$$MDI = \left[T^{-1}(1 - \{\alpha/2\}, df) + T^{-1}(\beta, df) \right] \times \sqrt{\frac{1}{Np(1-p)}} \times SD \qquad (5)$$

In this expression: $T^{-1}(\cdot)$ represents the inverse of the student's *t* distribution function; α is the significance level for the test, β is the level of statistical power; *df* is the number of degrees of freedom, which is equal to the number of respondents minus the number of groups minus the number of sites; *N* is the number of respondents; *p* is the fraction of respondents assigned to the treatment group; and *SD* is the standard deviation of the outcome.

Table B.1 shows MDIs for two response rates: (1) 80 percent, which would ideally be achieved; and (2) a slightly lower rate of 70 percent. In addition to presenting MDIs for the full sample, Table B.1 also displays MDIs for a 50 percent subsample and a 33 percent subsample—which could shed light on the impacts that could be detected in subgroup analyses—as well as MDIs for a single site. Using the full sample obtained from all study sites, the expected MDIs with a response rate of 80 percent are 5.0 percentage points for self-employment, 4.4 percentage points for employment in any job, and $1,044 for quarterly total earnings. With a response rate of 70 percent, MDIs would be 5.4 percentage points for self-employment, 4.7 percentage points for any employment, and $1,116 for quarterly total earnings. As might be expected, the subgroup and single-site MDIs are higher than the MDIs calculated for the full sample and pooled across all sites.

Table B.1. Minimum Detectable Impacts for Key Outcomes

			Outcome Variable (Units)		
Sample	**Survey Respondents in Program Group**	**Survey Respondents in Control Group**	**Self-Employment (Percentage Points)**	**Employment in Any Job (Percentage Points)**	**Quarterly Total Earnings ($)**
Response Rate = 80 percent					
All Sites					

Sample	Survey Respondents in Program Group	Survey Respondents in Control Group	Outcome Variable (Units)		
			Self-Employment (Percentage Points)	Employment in Any Job (Percentage Points)	Quarterly Total Earnings ($)
Full Sample	1,200	1,200	5.0	4.4	1,044
One-Half Subsample	600	600	7.1	6.3	1,477
One-Third Subsample	400	400	8.7	7.7	1,810
Single Site					
Full Sample	375	375	10.0	8.9	2,090
Response Rate = 70 percent					
All Sites					
Full Sample	1,050	1,050	5.4	4.7	1,116
One-Half Subsample	525	525	7.6	6.7	1,579
One-Third Subsample	350	350	9.3	8.2	1,935
Single Site					
Full Sample	263	263	10.7	9.5	2,235

Note: MDI calculations were calculated using equation (5) based on the following assumptions: (1) the level of power is 80 percent and a two-tailed test will be applied at a 5 percent significance level; (2) at the follow-up survey, the overall prevalence of self-employment will be 40 percent, the prevalence of employment in any job will be 75 percent, and the standard deviation of quarterly total earnings will be $10,200; (3) individuals at each site are assigned to the program and control groups with equal probability; (6) 20 percent of the variance in the outcome is explained by baseline covariates included in the regression; (7) point estimates are based on an unweighted regression; and (8) variance estimates do not account for heteroskedasticity.

To put the MDIs in Table B.1 in perspective, they can be compared with actual impacts found in a randomized evaluation of the Enterprise Project, a demonstration program that provided self-employment assistance to UI recipients in Massachusetts during the early 1990s (Benus et al. 1995).[13]

- The Enterprise Project increased self-employment by 11 percentage points during the 21 months after random assignment. Over the same period, Enterprise Project program group members were 13 percentage points more likely to be employed in any job. The MDIs in Table B.1 indicate that the SET Evaluation could detect such effects under response rates of 70 to 80 percent even when analyzing a 33 percent subgroup and when estimating the impact of the SET Demonstration at a single site.
- The Enterprise Project also increased total earnings, although this was largely due to increases in wage/salary earnings rather than self-employment earnings. The findings reported in Benus et al. (1995) suggest that the Enterprise Project increased total quarterly earnings of the program group by approximately $1,964 (in 2014 dollars). Based on Table B.1, the SET Evaluation could detect such an impact for the full sample, as well as the 50 and 33 percent subsamples when calculating pooled estimates across sites. This capacity to detect earnings impacts of the size observed in the Enterprise Project holds true for both response rates considered.

Thus, the SET Evaluation has the potential to statistically detect program effects of a realistic size, given the Demonstration's design and findings from other research about a similar intervention with a similar target population.[14] Moreover, the capacity to detect impacts of the size found in the Enterprise Project does not differ substantially across response rates of 70 percent and 80 percent.

[13] Benus et al. (1995) also evaluated a second demonstration program in Washington State, the Self-Employment and Enterprise Development (SEED) Project, which also provided self-employment assistance to UI recipients. However, the results from SEED evaluation were not considered to benchmark the MDIs calculated for the SET Demonstration for two reasons. First, although the SEED project specified that sample members could "cash out" their remaining UI entitlement, receiving a lump-sum payment after achieving certain benchmarks and business milestones, as described in Part A, the SEED program's lump-sum payments were: (a) substantially larger than the microgrants offered in the SET demonstration; and (b) only offered to that participants that had already secured adequate financing, which will not be required for participants to access the SET micrograNts. Second, the SEED Project was open to all UI recipients. Based on the WIA eligibility criteria noted previously, it is expected that the dislocated workers enrolled in the SET Demonstration will more closely resemble the likely UI exhaustees enrolled in the Enterprise Project.

[14] Applying nonresponse weights would reduce the precision of the SET Evaluation's impact estimates, due to a design effect from unequal weighting. However, as described in Section B.3, the contractor conducting the evaluation's follow-up survey will use a variety of proven techniques to maximize response rates for important subgroups.

3. Methods to Maximize Response Rates and Data Reliability

The contractor is using well-established methods to maximize response rates and data reliability for the follow-up survey. These methods have been used by the contractor in other data collection efforts, such as the Trade Adjustment Assistance Study Follow-Up Survey (OMB Control Number 1205-0460) and the Individual Training Account 2 (ITA2) Follow-up Questionnaire (OMB Control Number 1205-0441). Following a discussion of approaches for maximizing response rates and ensuring data reliability is a description of (1) the methods that are being used for addressing item non-response on the survey and (2) a detailed description of plans for analyzing and addressing individual-level survey nonresponse.

a. Maximizing Response Rates for the Follow-Up Survey

The strategy for maximizing response to the SET follow-up survey is based on the approaches described in following sections. The methods employed address all types of individual nonresponse, including failure to locate the sample member or his or her refusal to participate in the survey.

Web Administration of the Survey. Based on the pervasive use of the web by a cross-section of the general population, it is anticipated that the majority of sample members are likely to be most comfortable with a self-paced, self-administered web survey.

Contact with sample members. The contractor sends an advance letter on DOL letterhead to sample members shortly before the fielding of the survey begins to provide information about the content of the follow-up survey and average administration time, and explain how to access the web-based instrument. This letter (1) explains the voluntary and private nature of participation, (2) extends the incentive offer, (3) provides web survey log-in information, and (4) gives a toll-free number for telephone calls. The contractor is working with partner organizations in the study sites to encourage participation in the survey by sample members. The envelope for hardcopy advance letters is printed with the DOL logo to capture the sample members’ attention and to communicate the legitimacy of the study. Electronic copies of the advance letter are also be mailed to study members who provide an email address at baseline. The contractor’s return address is used to facilitate the processing of returned mail and locating procedures. The advance letter is followed up with timed reminders offering the option to complete the survey via the telephone or the web.

Before the mailing of these materials, interviewing staff, such as interviewers, project supervisors, monitors, and locators at Mathematica’s Survey Operations Center (SOC) were thoroughly trained on how to address respondents’ questions about the study and questionnaire. A list of frequently asked questions and answers (FAQs) was developed for the self-administered web survey, and web survey respondents have access to them throughout the survey.

Locating sample members. A key component to obtaining a high response rate is locating sample members. The process of locating members of the SET study population begins before sending out the first mailing. This locating process involves the use of an independent vendor that checks the full sample against current address databases. This first step is critical given that some sample members could have moved since the date at which they submitted their applications. Extensive tracking and locating procedures that have proven successful in other Mathematica studies are used for sample members whose mail is returned as undeliverable. These include using other independent databases, checking with neighbors and family members, and searching social networking sites. When talking with contacts, the specific purpose of the call is not disclosed, but it is stated that the effort to reach the sample member is for an important study being sponsored by the government.

Gaining and maintaining cooperation. A key component to achieving high response rates is gaining cooperation after locating respondents. Mathematica's interviewers are highly trained in establishing rapport with gatekeepers, gaining cooperation, and avoiding refusals. Sample members who are difficult to contact and who have not yet completed the survey on the web are sent a reminder email one week after the advance letter and a follow-up postcard one week later. Trained interviewers begin making reminder calls three weeks after the advance letter, and additional reminder emails are sent one week later, and postcards are sent two weeks after that to remaining nonrespondents. To those sample members who refuse to participate, a targeted refusal-conversion letter that will address their specific concerns is mailed first. Next, expert refusal-conversion interviewers make follow-up calls to try to gain the sample members' cooperation.

Incentives for survey participants. Offering an incentive for the SET follow-up survey could be important for obtaining the desired response rates and reducing overall survey costs. According to Singer et al. (2000), incentives can help to achieve high response rates by increasing the sample members' propensity to respond. By doing so, incentive payments were been found to contain evaluation costs by significantly reducing the number of calls required to resolve a case. Incentives also may increase the likelihood of participation from subgroups with a lower propensity to cooperate with the survey request. This can be an important component of ensuring the representativeness of the survey respondents and the quality of the data being collected. For example, Jäckle and Lynn (2007) found that incentives increased the participation of sample members more likely to be unemployed. There is also evidence that incentives bolster participation among those with lower interest in the survey topic (Schwartz et al. 2006;

Jäckle and Lynn 2007; Kay 2001), resulting in data that are more complete. Furthermore, paying incentives did not impair the quality of the data obtained (such as item nonresponse or the distribution of responses) from groups that would otherwise be underrepresented in the survey (Singer et al. 2000).

Part A of this clearance package provides additional discussion about the potential benefits of incentive payments for response rates and data quality. As discussed there, as part of the current OMB-approved effort to field the follow-up survey to the earlier groups of SET applicants the evaluation team conducted an incentive experiment to determine whether to offer sample members survey completion incentives, which are received in the form of a check. The results from the experiment were presented to OMB in a memo by DOL, which determined that the best incentive scheme to use was to offer sample members $50 for completing within the first four weeks or $25 for completing thereafter. This respondent payment scheme will continue to be used during the proposed extension period—see Section A.9 for details. Based on the findings from this incentive experiment, it is expected that this graduate $50/$25 will help ensure that survey response rates remain in the 70 to 80 percent range. To fully assess and leverage the benefits of offering incentives in the SET evaluation's follow-up survey, the advance letter to study participants explicitly mentions the payment. Sample members who elect to complete the survey via the telephone are also reminded of this incentive by the interviewers when contact is first established.

Survey length. The SET follow-up questionnaire is designed to be easy to complete. The questions are written in clear and straightforward language. The average time required for the respondent to complete the survey is estimated at 20 minutes.

Targeted response rate. Employing these procedures, a response rate of 70 to 80 percent for the SET follow-up survey is anticipated based the contractor's experience fielding the survey to date and the results of the incentive payment experiment already noted. As discussed above, this range of response rates is expected sufficient statistical precision for the evaluation. The evaluation team is taking active steps to reduce the potential risks of (1) nonresponse bias and (2) the extent to which precision might be reduced through nonresponse differentials that require adjustments through weighting. In particular, the study is monitoring response rates to assess whether there are systematic differences between the program and control groups or across demographic in the likelihood of nonresponse. This analysis uses baseline information data from the study application package that is available for all sample members, and its results are being used to target additional reminders and encouragement. As discussed in the following subsection, additional analyses will be conducted when the survey is completed to assess the extent of any remaining nonresponse differentials. If it appears that the survey respondent sample is not representative of the study sample, weights to adjust for nonresponse will be developed using propensity scoring methods.

b. Data Reliability for the Follow-Up Survey

The follow-up survey is unique to the current evaluation and is being used across all SET study sites, ensuring consistency in the collected data. The survey has been extensively reviewed by project staff and staff at ETA, was thoroughly pretested, and has been fielded for 10 months by the contractor. Potential respondents are referred to the survey web site by the advance letter and by AJC staff. If a respondent starts the web survey but encounters problems or must complete it at a later time, the survey can be resumed later. Every aspect of the web program was thoroughly tested before being put into production. Additionally, to ensure that respondents answer questions, all interview respondents are ensured of the privacy of their responses to questions.

Addressing item nonresponse. The follow-up survey primarily collects data on outcome measures to be used in the impact analysis. Although the past experience of the contractor conducting surveys for similar evaluations suggests that rates of item nonresponse on the follow-up survey will be very low, some item nonresponse is inevitable. Imputation of outcome data could lead to biased estimates due to imperfect matches on observables when using a hot-deck procedure (Bollinger and Hirsch 2006). Thus, sample members with missing data on a given outcome will be omitted from the sample when analyzing that outcome.

Addressing individual-level nonresponse. As with almost any survey, some nonresponse in the follow-up survey is inevitable. Some sample members will not be located and others will not be able or willing to respond to the survey. Even after the efforts of the contractor noted above, there is a potential for differential patterns of response that could indicate bias. A nonresponse analysis will use various baseline data items, including demographic characteristics, employment status, and earnings. The nonresponse bias analysis will consist of the following steps:

1. Compute response rates for key subgroups. A key subgroup comparison considers the difference between members of the program group and members of the control group and additional subgroups will be formed based on characteristics, as discussed in Section B.2.
2. Compare the distributions of respondents’ and nonrespondents’ characteristics.
3. Identify the characteristics that best predict nonresponse and use this information to generate nonresponse weights.

4. Compare the distribution of characteristics of respondents using response-adjusted analysis weights with the distribution of characteristics of the baseline sample.

These analyses will be conducted within and across sites to assess whether the potential for nonresponse bias differs among sites. Each of these steps is discussed in greater detail in the following subsections.

Compute response rates for subgroups. The response rate for the subgroups will be computed using the American Association for Public Opinion Research (AAPOR) definition of the participation rate for a nonprobability sample: the number of respondents who have provided a usable response divided by the total number of individuals from whom participation in the survey is requested (AAPOR 2011).[15] Overall response rates will be computed for the full sample and by site. Response rates will then be computed for subgroups defined by characteristics available from the baseline information form (collected under OMB Control Number 1205-0505) to examine if these rates differ systematically from the overall response rate.

Compare the characteristics of respondents and nonrespondents. Next, the characteristics of respondents and nonrespondents will be calculated according to characteristics available from the baseline information form. The statistical significance of the difference between the respondent and nonrespondent subgroups will be assessed using *t*-tests. This type of analysis can be useful in identifying patterns of differences in observable characteristics that might suggest nonresponse bias. However, this approach has low power to detect substantive differences when sample sizes are small, and the large number of statistical tests conducted can also result in high rates of Type I error. Consequently, the results of this item-by-item analysis will be interpreted cautiously.

Identify the best explanatory factors of nonresponse and generate nonresponse weights. Logistic regression modeling is commonly used to develop adjustment weights for nonresponse. This approach is also known as response propensity modeling and can be viewed as an extension of the classical weighting-class nonresponse adjustment procedure that makes it possible to include more factors (that is, binary, categorical, and continuous factors) in nonresponse adjustments.

[15] This OMB package submission uses the terms response and nonresponse, rather than participation and nonparticipation, to avoid confusion with “participation in the SET Demonstration program” by individuals who were randomly assigned to the program group. This terminology is not intended in any way to imply that the baseline sample for the SET Evaluation is sampled with known probabilities from a known population. Applicants will be self-selected from an unknown population and the evaluation will seek to draw inference about only the baseline sample of individuals that were randomly assigned.

The logistic nonresponse model will be fitted by first identifying a pool of covariates to work from using stepwise regression and then assessing candidate models using various measures of goodness of fit and predictive ability. The covariates will include factors or attributes that can be obtained from the baseline information form and which (1) are likely to be associated with differences in the likelihood that a sample member is located and interviewed and (2) have been shown by previous research (Benus and Michaelides 2010; Fairlie and Robb 2008) to be related to the outcomes of interest for this study among individuals seeking self-employment. Specific examples include demographics (age, sex, race/ethnicity); family structure (marital status or number of dependents); education level; receipt of UI benefits at the time of random assignment, and baseline measures of employment status and earnings from both self-employment and wage/salary jobs. Another important variable to be included in this analysis is the assignment (program or control) status of the individual.

A chi-squared automatic interaction detector (CHAID) will be used to refine the list of candidate independent variables and identify interactions among them.[16] The CHAID procedure iteratively segments a data set into mutually exclusive subgroups that share similar characteristics based on their effect on nominal or ordinal dependent variables. It automatically checks all variables in the data set and creates a hierarchy that shows all statistically significant subgroups. The algorithm finds splits in the population, which are as different as possible based on a chi-square statistic. It is a forward stepwise procedure, and it finds the most diverse subgrouping, and then each of these subgroups is further split into more diverse sub-subgroups. Sample size limitations are set to avoid generating cells with small counts. The algorithm stops when splits no longer are significant; that is, the group is homogeneous with respect to variables not yet used or when the cells contain too few cases. The CHAID procedure results in a tree that identifies the set of variables and interactions among the variables that have an association with the propensity of a baseline sample member to complete a follow-up survey.

The variables and interactions identified using CHAID then will be processed using forward and backward stepwise regression to further refine the candidate variables and interaction terms. After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, a set of models will be evaluated to determine the final model.

[16] CHAID is normally attributed to Kass (1980) and Biggs et al. (1991), and its application in SPSS is described in Magidson (1993). Decisions about variables and interactions will be based on statistical tests with the significance level (alpha level) set to 0.30. The test size of 0.30 is used instead of the standard 0.05 because the purpose of the model is to improve the estimation of the propensity score and not to identify statistically significant factors related to response.

Computing nonresponse adjustment factors through this process will contribute substantially to the nonresponse bias analysis by identifying the main effects and interaction among main effects that are statistically associated with nonresponse. This information will be used in the bias analysis to form levels of categorical variables for computing response rates and point estimates of program impacts using nonresponse adjustment weights.

Compare the nonresponse-weighted distribution of respondent characteristics with the distribution for the full random assignment sample. In this last step, the weighted distribution of respondent baseline characteristics will be compared with the unweighted distribution of the original study population that went through random assignment. Comparisons will be made for the full study population and for key subgroups, as described earlier in this subsection. This analysis can highlight measures in which the potential for nonresponse bias is greatest and in which greater caution should be exercised in the interpretation of the observed findings.

4. Tests of Procedures or Methods

All data collection procedures, instruments, and protocols to be used in the conduct of the SET Evaluation were tested to ensure that the procedures can be feasibly and efficiently carried out, 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. The forms contained in the follow-up survey instrument were thoroughly tested with up to nine individuals from nonparticipating sites with backgrounds similar to SET Demonstration participants. After each pilot test participant completed the forms, project staff debriefed each participant using a standard debriefing protocol to determine if any words or questions were difficult to understand and answer. Like actual study participants, participants in the pilot test of the follow-up survey were given an incentive for their time.

5. Individuals Consulted on Statistical Methods

Consultations on the statistical methods used in this study have been used to ensure the technical soundness of the study. Specifically, ETA has contracted with Mathematica to conduct the SET Evaluation. Table B.2 displays the technical staff who were consulted in planning for the implementation and evaluation of the SET Demonstration.

Table B.2. Contractor Technical Staff

Affiliation and Name	Role on Project	Telephone Number
Mathematica Policy Research		

Affiliation and Name	Role on Project	Telephone Number
Dr. Irma Perez-Johnson	Project director	(609) 275-2339
Dr. Heinrich Hock	Task leader, impact analysis	(202) 250-3557
Ms. Samia Amin	Task leader, implementation study	(609) 275-2375
Mr. Ryan Callahan	Survey director	(312) 994-1015
Mr. Shawn Marsh	Senior survey advisor	(312) 585-3319
Ms. Annalee N. Kelly	Survey researcher	(609) 275-2885
Ms. Stephanie A. Boraas	Survey researcher	(202) 484-3292
University of California, Santa Cruz		
Dr. Robert Fairlie	Consultant	(831) 459-3332

This page has been left blank for double-sided copying.

REFERENCES

American Association for Public Opinion Research (AAPOR). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Seventh edition. Lenexa, KS: AAPOR, 2011.

Benus, Jacob, and Marios Michaelides. “Are Self-Employment Training Programs Effective? Evidence from Project GATE.” Unpublished Manuscript. Munich Personal RePEc Archive Paper No. 20883. Available at http://mpra.ub.uni-muenchen.de/20883. Accessed May 27, 2011.

Benus, Jacob, Theodore Shen, Sisi Zhang, Marc Chan, and Benjamin Hansen. “Growing America Through Entrepreneurship: Final Evaluation of Project GATE.” Final report submitted to the U.S. Department of Labor, Employment and Training Administration. Columbia, MD: IMPAQ International, LLC, December 2009.

Benus, Jacob. M., Terry R. Johnson, Michelle Wood, Neelima Grover, and Theodore Shen. “Self-Employment Programs: A New Reemployment Strategy: Final Report on the UI Self-Employment Demonstration.” Unemployment Insurance Occasional Paper 95-4. Washington, DC: U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service, 1995.

Biggs, David, Barry de Ville, and Ed Suen. “A Method of Choosing Multiway Partitions for Classification and Decision Trees.” *Journal of Applied Statistics,* vol. 18, no. 1, 1991, pp. 49–62.

Bollinger, Christopher R., and Barry T. Hirsch. “Match Bias in the Earnings Imputations in Current Population Survey: The Case of Imperfect Matching.” *Journal of Labor Economics,* vol. 24, no. 3, July 2006, pp. 483–520.

Cameron, A. Colin, and Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press, 2005

Evans, David S., and Linda S. Leighton. “Some Empirical Aspects of Entrepreneurship.” *American Economic Review,* vol. 79, no. 3, June 1989, pp. 519–535.

Fairlie, Robert W., and Alicia Robb. *Race and Entrepreneurial Success: Black-, Asian-, and White-Owned Businesses in the United States*. Cambridge, MA: MIT Press, 2008.

Huber, Peter J. “The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions.” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, edited by L.M. LeCam and J. Neyman. Berkeley, CA: University of California Press, 1967.

Jäckle, Annette, and Peter Lynn. “Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias.” Working paper presented to the Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom, 2007.

Kass, G. V. “An Exploratory Technique for Investigating Large Quantities of Categorical Data.” *Applied Statistics,* vol. 29, no. 2, 1980, pp. 119–127.

Kay, Ward R. “The Use of Targeted Incentives to Reluctant Respondents on Response Rates and Data Quality.” *Proceedings of the American Association for Public Research*. Montreal, Canada: American Association for Public Opinion Research, 2001.

Magidson, Jay. *SPSS for Windows CHAID Release 6.0*. Belmont MA: Statistical Innovations, Inc., 1993.

McKinnon, James, and Halbert White. “Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties.” *Journal of Econometrics,* vol. 29, no. 3, September 1985, pp. 305–325.

Neter, John, Michael Kutner, Christopher Nachtsheim, and William Wasserman. *Applied Linear Statistical Models.* New York: McGraw-Hill, 1996.

Schwartz, Lisa K., Lisbeth Goble, and Edward M. English. “Counterbalancing Topic Interest with Cell Quotas and Incentives: Examining Leverage-Salience Theory in the Context of the Poverty in America Survey.” *Proceedings of the American Association for Public Research*. Montreal, Canada: American Association for Public Opinion Research, 2006.

Shochet, Peter Z. “Statistical Power for Random Assignment Evaluations of Education Programs.” *Journal of Educational and Behavioral Statistics*, vol. 33, no. 1, March 2008, pp. 62-87.

Singer, Eleanor, John Van Hoewyk, and Mary P. Maher. “Experiments with Incentives in Telephone Surveys.” *Public Opinion Quarterly,* vol. 64, no. 2, summer 2000, pp. 171–188.

White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica,* vol. 48, 1980, pp. 817–830.

Wooldridge, Jeffrey. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press, 2002.