Washington State's Integrated Basic Education and Skills Training (I-BEST) Program

Appendices for Three-Year Impact Report

PACE
Pathways for Advancing
Careers and Education

Washington State's Integrated Basic Education and Skills Training (I-BEST) Program: Appendices for Three-Year Impact Report

A Pathways for Advancing Careers and Education (PACE) / Career Pathways Intermediate Outcomes Study Publication

This report and other reports sponsored by the Office of Planning, Research, and Evaluation are available at www.acf.hhs.gov/opre.

✉ **Sign-up for the OPRE Newsletter**

Follow OPRE on Twitter @OPRE_ACF   **Like OPRE's page on Facebook OPRE.ACF**   **Follow OPRE on Instagram @opre_acf**   **Connect on LinkedIn company/opreacf**

OPRE   Abt ASSOCIATES

Contents

List of Exhibits

Appendix A: Baseline Characteristics and Adjustments

This appendix starts with a description of the specification for baseline characteristics, including our approach to handling missing values (Section A.1). The next section compares distributions for treatment and control group members on these and other baseline measures (Section A.2), and the last section explains how the analyses control for these covariates in estimating impacts (Section A.3). It should be noted that Sections A.1 and A.2 are nearly unchanged from parallel appendices in the first, short-term report on this program (Glosser et al. 2018). In contrast, the approach to covariate control in Section A.3 describes some important procedural changes from those used in the prior report.

A.1   Details on Baseline Covariates

Exhibit A-1 shows the specifications and data sources for baseline covariates. Item nonresponse rates on these covariates were generally low. Across all nine Pathways for Advancing Careers and Education (PACE) sites, item nonresponse rates were less than 4 percent except for parental college attendance (6.0 percent), typical high school grades (7.2 percent), family income (9.5 percent), and expected near-term future work hours (6.0 percent).

We imputed values for missing covariates using SUDAAN®/IMPUTE, a weighted hotdeck imputation procedure (Research Triangle Institute 2012). This imputation step entailed a single computer run on the combined sample from all nine PACE sites.[1] With this process, we replaced each missing value with an observed response from a similar case. Within specified strata, we random-matched cases with missing values to cases with reported values; we then copied over the reported value to the case where the value was missing. The strata represented a cross-classification of treatment-control status, site, National Student Clearinghouse (NSC)-reported enrollment status (*some* or *none*),[2] NSC-reported credential award (*some* or *none*), and number of months of NSC-reported enrollment.[3]

---

[1]   Using the combined data set better controlled for school enrollment status as measured in NSC in the smaller sites.

[2]   NSC has information on monthly enrollment and many credentials for 96 percent of college students. https://nscresearchcenter.org/workingwithourdata/

[3]   In instances where this level of matching was too restrictive because we found no matched case with a reported value, we re-ran the procedure matching only on treatment status and NSC-reported enrollment status. In this second pass imputation, matches were allowed across sites.

Exhibit A-1:   Operationalization of Baseline Measures Used as Covariates in Regression-Adjusted Impact Estimates

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Demographic Background		
Age	Categorical measure: Under 21 21-24 25-34 35+ [a]	BIF: B2_dob RABIT: R_RA_Date_Assigned
Female	Binary variable: 1 if female 0 if male	BIF: B7
Race/ethnicity	Categorical measure: Hispanic, any race Black, non-Hispanic White, non-Hispanic[a] Another racer, non-Hispanic	BIF: B9
Family structure	Categorical measure: Spouse/partner, with children Spouse/partner, without children Single, with children[a] Single, without children (Only biological and adopted children of randomized participant considered here. Stepchildren, grandchildren, younger siblings, and other children not considered.)	BIF: B13
Living with own parents	Binary variable: 1 if living with own parent(s) 0 otherwise (Presence of parents of spouse not considered)	BIF: B13
Educational Background		
Parent attended college	Binary variable: 1 if either parent attended college 0 otherwise	BIF: B21
Usual high school grades	Categorical measure: Mostly A's Mostly B's Mostly C's or below[a]	BIF: B23
Highest level of education completed	Categorical measure: No college[a] Less than one year of college credit One or more years of college credit Associate degree or above	BIF: B17
Career Knowledge		
Career Knowledge Index (average of items)	Proportion of responses to seven questions about career orientation and knowledge to which respondent answered, "strongly agree." Missing if four or more of seven responses blank.	SAQ: S13

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Psycho-Social Indices		
Academic discipline[b]	Average of 10 items (scale ranging 1= strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if seven or more of 10 responses blank.	SAQ: S11a
Training commitment[c]	Average of 10 items (scale ranging 1= strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if seven or more of 10 responses blank.	SAQ: S11b
Academic confidence[d]	Average of 12 items (scale ranging 1= strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if nine or more of 12 responses blank.	SAQ: S11d
Emotional stability[e]	Average of 12 items (scale ranging 1= strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if nine or more of 12 responses blank.	SAQ: S11e
Social support[f]	Average of 10 items (scale ranging 1= strongly disagree to 4= strongly agree). Missing if seven or more of 10 responses blank.	SAQ: S12
Resource Constraints (Financial)		
Family income in past 12 months	Categorical measure: Less than $15,000 $15,000-$29,999 $30,000+ [a]	BIF: B27
Received food assistance (WIC/SNAP) in past 12 months	Binary variable: 1 if yes 0 if no	BIF: B26b
Received public assistance or welfare in past 12 months	Binary variable: 1 if yes 0 if no	BIF: B26c
Financial hardship in past 12 months	Binary variable: 1 if yes to ever missed rent/mortgage payment in prior 12 months or reported generally not having enough money left at the end of the month to make ends meet over the last 12 months. 0 if otherwise	SAQ: S8, S9
Resource Constraints (Time)		
Current work hours	Categorical measure: 0-19[a] 20-34 35+	BIF: B24
Expected work hours in next few months	Categorical measure for covariate: 0-19[a] 20-34 35+	SAQ: S2
Expecting to attend school part-time if accepted	Binary variable: 1 if yes 0 if no	SAQ: S1

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Life Challenges		
Frequency of situations interfering with school, work, job search, or family responsibilities	Average of six items of frequency of problems in past 12 months (childcare, transportation, alcohol or drug use, health, family arguments, physical threats). Scale ranges from 1=never to 5=very often. Missing if four or more of responses blank.	SAQ: S15
Stress[f]	Average of four items about feeling in control of important things and able to handle personal problems (scale 1=never to 5=very often over the past month) after reversing responses to negatively phrased items. Missing if three or more of four responses blank.	SAQ: S14

Key: BIF = Basic Information Form. RABIT = Random Assignment and Baseline Information Tool. SAQ = Self-Administered Questionnaire. SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] Category omitted in creating binary (dummy) variables for regression-adjustment models.

[b] Modified version of the Academic Discipline scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson et al. (2006).

[c] Modified version of Commitment to College scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson et al. (2006).

[d] Modified version of the Academic Self-Confidence scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson et al. (2006).

[e] Modified version of the Emotional Control scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson et al. (2006).

[f] Modified version of the Social Provisions Scale; Cutrona and Russell (1987). Original scale has 24 items. This short version developed by Hoven (2012).

[g] Cohen et al. (1983).

A.2   Comparing Treatment and Control Groups at Baseline

Exhibit A-2 shows tests for similarity in characteristics of treatment and control group members at baseline. If the means in the two columns are congruent, then "baseline balance" was achieved. Assessment of congruence involved testing for equality of the two means separately for each characteristic.

The last column contains *p*-values for tests of hypotheses of no systematic differences between the treatment and control groups. If we were to repeat the randomization process a large number of times, out of 28 tests, on average, three will fall outside a 90 percent confidence interval due to chance. In this case, the *p*-values in Exhibit A-2 show there were three statistically significant differences, which are highlighted in red. The research team carefully reviewed data processing and other operations but could find no causes for these differences. It is likely that these are simply chance results. Furthermore, as described in the next section, regression adjustment helps to control for any effects that such chance differences might have on the impact estimates.

Exhibit A-2:   Baseline Balance

Characteristic	All Participants	Treatment Group	Control Group	*p*-Value
Age (%)				.067
20 or under	22.2	23.2	21.1	
21-24	14.9	11.1	18.6	
25-34	29.8	31.4	28.1	
35+	33.2	34.3	32.2	
Female (%)	42.5	44.9	40.1	.231
Race/Ethnicity (%)				.346
Hispanic, any race	26.0	28.9	23.1	
Black, non-Hispanic	7.6	6.2	9.1	
White, non-Hispanic	54.9	53.1	56.7	
Another racer, non-Hispanic	14.1	13.4	14.8	
Family Structure (%)				.591
Not living with spouse/partner and not living with children	47.2	48.7	45.8	
Not living with spouse/partner but living with children	16.6	14.6	18.6	
Living with spouse/partner and not living with children	17.3	18.2	16.3	
Living with spouse/partner and children	18.9	18.5	19.3	
Living with parents (%)	28.6	27.2	30.1	.412
One parent has at least some college (%)	45.3	45.5	45.2	.955
Usual High School Grades (%)				.170
Mostly A's	6.9	7.7	6.2	
Mostly B's	33.2	36.8	29.5	
Mostly C's or below	59.9	55.6	64.3	
Highest Level of Education (%)				.497
Less than a high school diploma	30.7	28.2	33.1	
High school diploma or equivalent	40.0	42.0	38.0	
Less than one year of college	11.1	12.1	10.2	
One or more years of college	9.5	10.1	8.9	
Associate degree or higher	8.8	7.7	9.8	
Received vocational or technical certificate or diploma (%)	19.3	19.7	19.0	.853
Career Knowledge Index (mean)	0.41	0.41	0.41	.934
Psycho-Social Indices (means)				
Academic Discipline Index	5.07	5.05	5.08	.528
Training Commitment Index	5.42	5.42	5.43	.828
Academic Self-Confidence Index	4.47	4.47	4.48	.960
Emotional Stability Index	4.95	4.95	4.94	.943
Social Support Index	3.21	3.21	3.21	.898
Stress Index	2.31	2.30	2.31	.887
Depression Index	1.60	1.61	1.59	.729
Family Income in Past 12 Months (%)				.551
Less than $15,000	47.3	46.5	48.1	
$15,000-$29,999	23.9	26.0	21.9	
$30,000+	28.8	27.6	30.0	
Family income (mean $)	22,110	23,002	21,240	.378

Characteristic	All Participants	Treatment Group	Control Group	*p*-Value
Public Assistance/Hardship Past 12 Months				
Received WIC or SNAP (%)	58.6	55.0	62.1	.092
Received public assistance or welfare (%)	21.3	18.1	24.3	.094
Reported financial hardship (%)	48.5	49.8	47.1	.499
Current Work Hours (%)				.993
0	66.6	66.9	66.3	
1-19	8.5	8.5	8.5	
20-34	11.7	11.7	11.6	
35+	13.2	12.8	13.6	
Expected Work Hours in Next Few Months (%)				.228
0	41.1	41.4	40.8	
1-19	9.9	8.6	11.2	
20-34	32.0	35.2	28.9	
35+	17.0	14.8	19.1	
Life Challenges Index (mean)	1.56	1.56	1.57	.906
Owns a car (%)	62.7	62.1	63.4	.733
Has both computer and internet at home (%)	72.0	70.2	73.7	.338
Ever arrested (%)	29.0	28.4	29.6	.740
Sample sizes	631	315	316	

Source: PACE Basic Information Form; PACE Self-Administered Questionnaire.

Note: Tests for statistically significant imbalance were based on SAS/FREQ procedure for categorical outcomes and on the SAS/TTEST procedure for other outcomes. Significant imbalances are highlighted in red, using a threshold for statistical significance of 10 percent. All values are based on baseline balance prior to imputation of missing values.

A.3   Regression Adjustment

This section describes the regression adjustment approach used to improve precision and minimize effects of sampling error on impact point estimates. In a rigorous evaluation, random assignment ensures that if the sample size is large enough, differences in average potential outcomes between the treatment and control groups will become vanishingly small so that any observed differences in average outcomes across the two groups must almost certainly be the result of treatment.[4] Even when sample sizes are modest, random assignment ensures that that differences in average potential outcomes between the treatment and control groups arise from chance rather than biases of program operators or program evaluators. This means that unbiased estimates of the effects of treatment can be obtained by simply comparing average outcomes across the treatment and control groups. Moreover, it is easy to run formal tests of the hypothesis that the program has no effect (and that therefore the observed difference in

---

[4]   Potential outcomes are a central concept in the Neyman-Rubin causal model (Holland 1986). In this model, each person has an innate pair of possible outcomes: one if treated and the other if not treated. Only one of the two potential outcomes is ever observed for each person. The average difference in potential outcomes across a specific population is said to be the local average treatment effect (LATE) or more simply, just the effect of treatment, with the context making clear the population to which it applies and supplemental analyses exploring whether the effect is homogenous within that population.

mean outcomes is the result of those accidental imbalances in potential outcomes across the two groups).

Despite these favorable properties of analysis based on simple comparisons of observed means, use of regression adjustment can reduce the impact of accidental imbalances in potential outcomes across the groups, thereby increasing power to detect small program impacts (Lin 2013). To achieve this benefit, the variables used in the regression adjustment must be predictive of potential outcomes. Including other variables will increase the variance on the estimated program impact rather than decreasing it.

Opinions and practice differ on how strong the evidence for correlation between a baseline variable and the outcome must be before it makes sense to include the baseline variable in the regression adjustment.[5] Some favor a lean approach, including just those baseline variables that have a demonstrated strong relationship to the outcome, while others favor a more comprehensive approach including all baseline variables that have a plausible theoretical relationship to outcomes of interest, believing that doing so generally bolsters confidence in study findings (Tukey 1991).

Given demands to minimize burden on participants, all measured PACE baseline variables have at least plausible relationships to PACE outcomes, but some baseline variables have been discovered to have only weak empirical relationships with PACE outcomes. Moreover, one could combine the directly measured characteristics into a limited number of interactions. So some judgment must be exercised about which covariates to include in regression adjustments and which to exclude.

Opinions and practice also differ on how much to customize decisions about covariate inclusion across outcomes in evaluations (like this evaluation of I-BEST) with multiple outcomes. A single uniform set of decisions promotes transparency, making it easier for readers to understand the procedure, while a more customized approach is likely to improve variances for at least some outcomes given that the correlation between a covariate and an outcome will vary by outcome.

In preliminary analyses for the first round of PACE reports, the team planned to use a fairly comprehensive approach with a uniform set of decisions but discovered that this approach was causing the variances on adjusted impacts to be larger than the variances on unadjusted impacts. The discovery prompted a switch to a different approach for the first round of reports, which ultimately proved not to work as well as hoped (Judkins 2019). In response, the team developed a new approach for the current round of PACE reports. This new approach emphasizes transparency and control on imbalanced covariates, while still trying to maximize precision as far as possible given those priorities. Details follow.

Equation (A.1) below shows the conventional regression-adjustment model:

$$Y_i = X_i\beta + \delta T_i + e_i \qquad\qquad\qquad (A.1)$$

---

[5]   For a current review of practice, see Ciolino, et al. (2019).

where Y_i is the outcome; X_i is a row vector of baseline characteristics (hereafter referred to as covariates); β is the vector of parameters indicating the influence of each covariate on the outcome; δ is the effect of treatment; T_i is a 0/1 dummy variable indicating treatment group membership; and e_i is an error term. We fit models of this sort using SAS/SurveyReg, a procedure that uses a robust estimator of the variance of $\hat{\delta}$ and can accommodate the required nonresponse-adjustment weights for survey-measured outcomes. (See Appendix Section C.3 for a discussion of nonresponse-adjustment weights.)

This method is known as ordinary least squares (OLS) and has excellent properties when the sample size is many times larger than the number of baseline characteristics used as covariates (Lin 2013), even when the outcomes are not normally distributed (Judkins and Porter 2016). Estimates of the treatment effect are "asymptotically unbiased," and under most conditions, the variance of the estimated treatment effect declines from the simple difference-in-mean-outcomes estimator of impact in proportion to the amount of outcome variation explained by the covariates.

Specifically, the relationship between the variance of the estimated treatment effect from the OLS estimation of Equation (A.1) and the explanatory power of the covariates is $var\left(\hat{\delta}\right) \approx (1 - R^2) var(\bar{y}_t - \bar{y}_c)$, where R^2 is proportion of the variance in Y_i explained by the baseline characteristics (X_i) in OLS estimation of Equation (A.2) below:

$$Y_i = X_i\beta + e_i \qquad\qquad (A.2)$$

However, as mentioned above, when there are a large number of potential covariates, not all of which are useful in predicting every outcome of interest, the effect of adjustment can be the opposite of the intended effect: variances are increased rather than decreased.[6] To avoid unnecessary variance inflation, the analyst needs to drop or otherwise reduce the influence of extraneous covariates that do not have a strong influence on the outcome of interest.

Simulation research (Judkins 2019) showed that dropping (with "backward selection") or downweighting covariates[7] based on simple analyses of the same data used in the evaluation yields slightly biased estimates of the variance of the estimated treatment effects (but still

---

[6]   Mathematically, the presence of extraneous variables causes the coefficients of the true determinants of the outcome to be less accurately estimated. For example, if the best prediction model is $Y = 2X$ but the model is fit with many extraneous covariates, the fit prediction formula could easily end up having coefficients of 1.9 or 2.1 for X instead of the best value of 2. If the wrong slope is used to correct for a treatment-control imbalance in X, the adjusted estimate of impact can be worse than an unadjusted estimate of impact.

[7]   An example of a method that downweights covariates is the "modified Koch method" developed for and used in the first round of PACE reports (Judkins et al. 2018; Koch et al. 1998).

unbiased estimates of the treatment effect itself).[8] This bias is negative, meaning that the variance estimates are slightly too small, making confidence intervals for impact estimates misleadingly narrow and hypothesis tests too likely to conclude that a nonzero impact has occurred when the true impact is zero or negative.

To select covariates in a manner that does not compromise variance estimation, we use the relatively recently developed technique "least absolute shrinkage and selection operator" (LASSO) with "10-fold cross-validation."[9] With the LASSO, the sum of absolute values of the estimated regression coefficients in Equation (A.2) is constrained to be less than a preselected value (the "constraint"). If the value for this constraint is small enough, many coefficients in Equation (A.2) will be forced to zero to fit within the cap on the sum of absolute coefficient values and thus can be removed from the list of baseline covariates. The 10-fold cross-validation is used to optimize the value of the constraint, rather than just relying on an arbitrary choice for it.

Details of the procedure are as follows.

(1)   With 10-fold cross-validation, the sample (both treatment and control group members) is divided into 10 equal and mutually exclusive random subsamples.

(2)   For each of a range of candidate values of the constraint, the LASSO procedure is run to select covariates on a sample in which one of the 10 subsamples has been dropped.

(3)   The model in Equation (A.2) is fit on the same sample using just the variables selected in the second step for each of the candidate values of the constraint.

(4)   The model is used to create out-of-sample predictions of the outcome for everyone in the dropped piece of the sample, and the prediction error $\hat{Y}_i - Y_i$ is measured for each of the candidate values of the constraint.

(5)   Steps 2 through 4 are repeated 10 times for each candidate value of the constraint. On each iteration, a different one of the 10 subsamples is dropped. In this manner, out-of-sample prediction errors are obtained for the entire sample.

(6)   Mean squared prediction errors across all 10 replicates are then calculated for each of the candidate values of the constraint.

(7)   The value of the constraint that minimizes this cross-validated mean squared prediction error and thus captures most of the variation reduction possible with the available

---

[8]   If the sample size is very large, the estimated variance of the estimated effect of treatment will be nearly unbiased even if the evaluation data are used to cull or downweight extraneous covariates. However, simulations clearly show that PACE sample sizes are not large enough to avoid biased variance estimates if "backward selection" on local data is used to prune covariates or if the modified Koch method is used to downweight extraneous covariates. Accordingly, impact analyses at three years for I-BEST and all other PACE programs are not using the modified Koch method used in the first, short-term round of reports covering the first 18 months of follow-up.

[9]   See Bühlmann and van de Geer (2011) for a full explanation of these techniques.

covariates is selected as the optimal constraint.[10] Whichever variables have nonzero coefficients in the model for that optimal constraint are used as covariates in the impact regressions. All other baseline characteristics are discarded. All of this is done automatically in SAS/GLMSELECT. Simulations under PACE-like conditions (in terms of sample sizes and the numbers of covariates) when developing the analysis plan for the entire suite of PACE three-year reports (Judkins et al. 2018) demonstrate that this technique reduces the true variances without biasing variance estimates.[11]

In principle, we could repeat the LASSO with 10-fold cross-validation independently for every outcome for each of the nine PACE programs. But such an approach would produce a different final covariate list for each outcome and program, leading to some loss in transparency and making it harder for outside researchers to replicate the PACE results. At the other extreme, we could run the LASSO just once for each program for the most important confirmatory outcome and then use the resulting set of selected covariates for all impact estimates for the program. But we believe that this would result in more precision loss than can be justified for the sake of transparency.

As a compromise between these extremes, we selected one set of covariates for each of three domains and customized them for each of the nine PACE programs. The three domains are (1) analyses of *employment and earnings* outcomes that are conducted on the dataset of merged data from the three-year follow-up survey and the National Directory of New Hires (NDNH); (2) analyses of *education* outcomes (whether based on the survey, NSC, or local or state college records); and (3) analyses of all *other* outcomes (most of which concern personal and family well-being and economic independence). The pool of potential covariates was the same for all three domains—with one important exception: indicators of pre-baseline earnings based on NDNH data are only allowed in analyses of NDNH-based outcomes.[12]

To identify covariates for this report, we ran the LASSO procedure for the most salient outcome within each of the three domains (*earnings and employment, educational progress, other*) at each of the nine PACE programs.[13] For NDNH analyses, the confirmatory outcome is average quarterly earnings for the 12th and 13th quarters after randomization (Q12, Q13), so that is a natural choice for the outcome around which to optimize covariate selection. In the *educational progress* domain, the most important outcome varies by PACE program. As discussed in the

---

[10] One could simply use the LASSO to select covariates with a pre-specified value of the constraint, but the 10-fold cross-validation provides a principled method for selecting the constraint.

[11] See Judkins (2019) for additional detail.

[12] This is because we analyzed survey outcomes on Abt's secure server rather than on the ACF secure server. Though both systems have very high security procedures, agreements with the Office of Child Support Enforcement permit the NDNH data to reside only on the ACF secure server. It would have been possible to analyze all survey outcomes on the ACF secure server, but doing so would have significantly burdened the study's analytic operations without any commensurate benefit. It would also prevent us from analyzing survey data for people whose names and Social Security numbers do not match properly according to OCSE.

[13] Selection started with the set of baseline covariates used in the analyses of follow-up data at 18 months after random assignment for the short-term impact report (shown in Exhibit A-3).

body of this report, for the Integrated Basic Education and Skills Training (I-BEST) program, the most salient education outcome is receipt of a credential requiring a year or more of college study. As the most salient outcome for the third domain, we selected whether anyone in the household draws means-tested public benefits. We made this last decision because of the centrality of the concept of self-sufficiency in the rationale for creating the PACE project.[14] We made these choices prior to reviewing any impact estimates.

In addition to covariates based on the above procedures, regression models included covariates for which baseline distributions differ for treatment and control group members at the 5 percent level.[15]

Exhibit A-3 below shows the covariates that we selected with the LASSO procedure. The procedure would have included covariates that were out of balance, but there were not any out of balance covariates for I-BEST. For categorical variables, the LASSO procedure worked on dummy variables for the individual levels; so for a variable with four levels, it was possible for just one of three dummy variables to be selected. For example, only one level of future work hours was selected as a covariate for employment outcomes. The table shows all possible levels of categorical variables and indicates which specific categories we selected as covariates.

---

[14]   The original name for PACE was "Innovative Strategies for Increasing Self-Sufficiency." The promotion of self-sufficiency is also central to the goals of the career pathways framework, as articulated by Fein (2012).

[15]   Baseline balance was assessed prior to imputation of missing data. See Exhibit A-2.

Exhibit A-3:   Covariates Selected, by Outcome Domain

Baseline Covariate	NDNH-Based Employment and Earnings Domain	Educational Progress Domain	Other Domain
Age			
20 or under			
21-24			
25-34			
35+			
Gender			
Female	LASSO	LASSO	
Male			
Race/Ethnicity			
Hispanic, any race			
Black, non-Hispanic			
White, non-Hispanic			
Another race, non-Hispanic			
Family Structure			
Not living with spouse/partner and not living with children		LASSO	
Not living with spouse/partner but living with children			
Living with spouse/partner and not living with children			
Living with spouse/partner and children			
Living with parents			LASSO
One parent has at least some college		LASSO	
High School Grades			
Mostly A's			
Mostly B's			
Mostly C's or below			
Current Education			
High school diploma or less			
Less than one year of college			
One or more years of college			
Associate degree or higher			
Career Knowledge Index			
Family Income in Past 12 Months			
Less than $15,000			
$15,000-$29,999			
$30,000+			
Pre-Randomization Quarterly Earnings (NDNH)		Not available	Not available
4 quarters prior to randomization			
3 quarters prior to randomization	LASSO		
2 quarters prior to randomization	LASSO		
1 quarter prior to randomization			

Baseline Covariate	NDNH-Based Employment and Earnings Domain	Educational Progress Domain	Other Domain
Pre-Randomization Quarterly Employment (NDNH)		Not available	Not available
4 quarters prior to randomization			
3 quarters prior to randomization			
2 quarters prior to randomization			
1 quarter prior to randomization	LASSO		
Psycho-Social Indices			
Academic Discipline Index	LASSO		
Training Commitment Index			
Academic Self-Confidence Index			
Emotional Stability Index			
Stress Index			
Life Challenges Index	LASSO		
Public Assistance/Hardship Past 12 Months			
Received WIC or SNAP	LASSO		LASSO
Received public assistance or welfare			
Reported financial hardship		LASSO	
Current Work Hours			
0-19			
20-34			
35+			
Expected Work Hours in Next Few Months			
0-19			
20-34			
35+	LASSO		LASSO
Plan to attend school only part-time if admitted to I-BEST	LASSO		

Key: SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.
Source: PACE Basic Information Form; PACE Self-Administered Questionnaire.
Note: "LASSO" flags that the covariate was selected by the LASSO for variance reduction.

Exhibit A-4 below shows impacts on selected confirmatory and secondary outcomes before and after regression adjustment without weights.[16] The two sets of estimates lead to similar conclusions. Regression adjustment did reduce the standard errors for all three of the targeted outcomes (earnings, receipt of a credential requiring a year or more college study, and receipt of means-tested benefits). To get variance reduction on every estimate, it would probably be necessary to run a separate LASSO for each outcome.

---

[16] We did not use the weights in the preparation of this table because they are not required for the first panel (Full Sample), and because in this section we want the focus to be on the role of covariates. See Appendix Exhibit C-8 for the impact of nonresponse-adjustment weights on these estimates.

Exhibit A-4: **Comparison of Confirmatory and Secondary Impact Estimates Unadjusted and Adjusted for Baseline Imbalances**

Domain (Data Source), Outcome	Impact (Unadjusted Estimate)	Standard Error	Impact (Adjusted Estimate)	Standard Error
Confirmatory Outcome: Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q12-Q13 after randomization ($)	388	371	404	344
Secondary Outcome: Employment (Survey)	**Survey Respondents without Weights**			
Employed at survey follow-up (%)	6.3*	4.8	7.0*	4.8
Employed at $14 per hour or above (%)	6.9*	4.3	5.6*	4.3
Employed in a job requiring at least mid-level skills (%)	−3.8	3.1	−4.0	3.2
Confirmatory Outcome: Education (SBCTC Records)	**Full Sample**			
Received credential taking 1+ year of college study	2.9	2.4	2.4	2.3
Secondary Outcomes: Education (SBCTC Records)	**Full Sample**			
Number of workforce and academic credits	11.3***	2.7	10.9***	2.6
Full-time-equivalent months enrolled in college	2.5***	0.5	2.4***	0.5
Receipt of any college credential (%)	30.5***	3.5	31.0***	3.5
Secondary Outcomes: Education (Survey)				
Receipt of an exam-based certification or license (%)[a]	9.8**	4.6	10.2**	4.7
Secondary Outcomes: Other (Survey)	**Survey Respondents without Weights**			
Indicators of Independence and Well-Being				
Has health insurance coverage (%)	−0.8	3.6	−1.2	3.7
Receives means-tested public benefits (%)	1.4	4.9	2.2	4.5
Personal student debt ($)	345	636	380	586
Any signs of financial distress (%)	−6.1	4.9	−5.4	4.8
Indices of Self-Assessed Career Progress (average)				
Confidence in career knowledge[b]	0.05	0.06	0.05	0.06
Access to career supports[c]	−0.02	0.03	−0.01	0.03
Sample sizes (across treatment and control groups):				
SBCTC 631				
NDNH 610				
Survey 419				

Source: SBCTC records; PACE 18-month and three-year follow-up surveys; National Directory of New Hires.

[a] Blended 18-month and three-year survey results.

[b] Seven-item scale tapping self-assessed career knowledge; response categories range from 1=strongly disagree to 4=strongly agree.

[c] Six-item scale tapping self-assessed access to career supports; response categories range from 1=no to 2=yes.

Statistical significance levels based on one-tailed *t*-tests of differences between research groups: *** 1 percent level; ** 5 percent level; * 10 percent level.

Appendix B: College Records Data

This appendix explains the data sources and strategy for measuring the confirmatory outcome (*receipt of a credential requiring a year or more of college study*) as well as other important college outcomes. Section B.1 discusses our decision to base most such measures on records from Washington's State Board for Community and Technical Colleges (SBCTC), rather than on potential alternative sources such as the three-year follow-up survey and the National Student Clearinghouse (NSC). This decision required the imputation of college outcomes for students who studied at colleges outside the SBCTC system. Then Section B.2 documents the methodology for this imputation.

In addition to preparing impact estimates with the college records data, we also prepared alternative impact estimates using NSC data and three-year follow-up survey data. See Appendices D and E, respectively, for these sensitivity analyses. It is important to note that the survey is used to measure educational progress not just at colleges but also at schools that are not accredited to grant degrees. Details of these measurements are discussed in Appendix C.

B.1   Rationale for Use of SBCTC Records

The SBCTC records offer several advantages over other potential sources for defining college education outcomes. For study participants who enrolled in one of 34 SBCTC member colleges,[17] the records provide complete data on enrollment and credentials. This is a key advantage over the three-year follow-up survey, which covers respondents only and may be subject to response bias or recall error. Moreover, credential duration is classified by the college, allowing us to accurately distinguish credentials requiring a year or more of study from credentials requiring less time. This is an important advantage over the NSC credential data, which do not have sufficient information to categorize credentials as taking less than or more than a year to complete.

The SBCTC data, however, have several limitations. First, due to data use limitations, we only had access to academic records from SBCTC through March 2017. That end date implies that we have three years of post-randomization follow-up for 83 percent of study participants, but only 30 months of follow-up for 17 percent of participants (i.e., those randomized near the end of the study enrollment period). However, for the 17 percent with incomplete SBCTC data, analysis of NSC data found that the vast majority had already completed their training by 30 months after randomization; only four participants had an NSC-reported spell at an SBCTC college after March 2017.[18] So the length of available SBCTC data is not a major limitation.

Another, and more serious, limitation of the SBCTC records is that they do not cover credentials earned at other colleges (four-year colleges in Washington, for-profit two-year colleges in

---

[17]   Information about SBCTC member colleges is available here: https://www.sbctc.edu/our-colleges/.

[18]   NSC data on college enrollment are very complete, as discussed in Appendix D, so this finding suggests that almost everyone missing SBCTC data on months 31-36 had already completed their college training by month 30.

Washington, or all out-of-state colleges). If a large share of study participants attended other colleges, then outcomes based on SBCTC records would underestimate credential attainment. Importantly, if attendance at other colleges differed between the treatment and control group, then our impact estimates would likely be biased. Because NSC has excellent coverage on colleges across the country, it can be used to measure the seriousness of this problem and to adjust the SBCTC data to compensate for this weakness.

Analysis of NSC records showed that the majority of study participants who enrolled in college attended an SBCTC college (Exhibit B-1). More than 94 percent of those with any college attendance attended an SBCTC college: more than 90 percent attended *only* an SBCTC college; less than 5 percent attended both an SBCTC college and another college. Less than 6 percent attended only another college. However, almost all this attendance at non-SBCTC colleges is among members of the control group, so there is a potential for serious bias in impact estimates if we were to rely only on SBCTC records.

Exhibit B-1:   NSC-Reported Enrollment at SBCTC and Other Colleges by Research Group

	Enrolled at SBCTC Colleges			Enrolled Only at Colleges Other Than SBCTC Colleges (%)	Total Ever Enrolled (%)
	At SBCTC Colleges (%)	Only at SBCTC Colleges (%)	At SBCTC Colleges and Other Colleges (%)		
Through 30 Months After Random Assignment					
Treatment	99.2	98.1	1.2	0.8	100.0
Control	94.6	90.5	4.2	5.4	100.0
T-C difference	4.6	7.6	−3.0	−4.6	

Source: National Student Clearinghouse.

These results suggest that SBCTC records provide sufficient coverage to use as the primary source for defining college education outcomes, but that an imputation procedure is necessary to ensure that attendance at other colleges is also captured in the outcomes. Failure to do so (i.e., only using SBCTC records) would likely lead to bias. We developed a procedure to impute the outcomes of enrollment spells at colleges outside of the SBCTC system as well as for the outcomes of the spells of the four participants that occurred after the end of the available SBCTC data, as discussed above.

B.2   Imputation of Enrollment, Credits Earned, Credentials Earned at Colleges Other Than SBCTC Colleges

The prior section motivated the decision to base most education outcomes on SBCTC records, supplemented with imputed experiences at other colleges. This section documents the imputation procedure used for this purpose. The procedure is based largely on information from the NSC, but it also uses information from baseline forms and both rounds of follow-up surveys. In a nutshell, our approach entails matching NSC-reported spells at colleges other than those in the SBCTC system to spells at SBCTC colleges, and then copying *SBCTC-reported variables*

associated with that SBCTC spell to the person who attended the other college, dropping all NSC data about the spell.[19]

For example, if the NSC reports that "Bill" attended Washington State University (which is not an SBCTC college) for a certain number of months and was (or was not) awarded a credential, the imputation procedure identifies a person ("Susan") who attended an SBCTC college for a similar number of months with similar outcomes according to both NSC and the survey. The procedure then duplicates Susan's spell at an SBCTC college as a spell for Bill. Continuing the example, if SBCTC records show that Susan received an associate degree from an SBCTC college, then the procedure imputes that Bill did the same. Summary statistics about Bill's training history then sum across this imputed spell as well as any other spells that Bill may himself have had at SBCTC colleges.

This imputation yields unbiased impact estimates if, conditional on the information used in the matching, a spell at an SBCTC college has the same expected outcome as a spell at another college. For example, the conditional probability of the matched spell leading to a credential from an SBCTC college should be the same as the conditional probability of the other-college spell resulting in a credential from that other college. Whether this assumption is reasonable depends on the quality of the matching, which in turn depends both on the relevance of information available to use in a matching process and on the details of how that information is used in the matching process. As implemented, the matching process used NSC-reported spell duration, NSC data on credential receipt, survey data on credential receipt, and baseline information to identify suitable SBCTC spells.

One potential shortcoming of the procedure is that it does not use information on receipt of support services. Because the Integrated Basic Education and Skills Training (I-BEST) program did boost some forms of supports, such as help arranging supports from third parties, to help students manage school, work, or family responsibilities (Glosser et al. 2018, Exhibit 4-5), it is possible that the imputation does not do full justice to the program. On the other hand, the procedure did control for survey-reported receipt of credentials. So for the lack of control on receipt of support to be a problem, the supports would need to have had important effects other than boosting receipt of credentials. An example might be a service-induced boost in training persistence that did not result in additional credentials. We did not study whether this was true, and we cannot rule it out, but we believe it is unlikely.

At a more detailed level, the imputation strategy involved six steps (briefly listed and then explained in detail in the next sections).[20]

(1) Find an SBCTC record for as many NSC-reported SBCTC spells as possible. This step filled in credits and credentials earned for most NSC-reported SBCTC spells. The research team

---

[19] Alternatively, we could have kept some of the NSC information such as whether the school was a four-year school and specific credentials awarded, but we believed it would make for a more coherent set of imputations not to mix actual NSC data about the person with imputed SBCTC for the person.

[20] We used a similar procedure to model SBCTC outcomes for the short-term impact report, but step 3 was not part of that earlier procedure, and step 4 at 18 months did not use survey data.

referred to this step as the "exact matching" process, because there is a single correct match in the SBCTC records for almost all the NSC-reported SBCTC spells.

(2)   Resolve NSC-reported SBCTC spells that did not match to an SBCTC record. For such NSC-reported spells, the team assumed that no credits were earned and no credentials were earned.

(3)   Impute educational attainment variables for nonrespondents to the three-year follow-up survey based on baseline data and data from the 18-month follow-up survey.

(4)   Summarize the available data for each NSC-reported SBCTC spell and the student to whom the spell belonged. The team summarized these data using statistical models that predict four critical SBCTC-reported outcomes for each spell: (a) receipt of a degree requiring at least one year of credits by 36 months after randomization; (b) receipt of any credential by 36 months after randomization; (c) full-time-equivalent months of academic or vocational coursework by 36 months after randomization; and (d) cumulative academic and vocational credits earned by 36 months after randomization.

(5)   Match each SBCTC-reported other-college spell with a similar SBCTC-reported SBCTC spell in terms of the four predicted critical outcomes from step 4. The team referred to this step as "statistical matching," because many possible NSC-reported SBCTC spells could be matched to every NSC-reported other-college spell. The team matched only other-college spells of students in the treatment group to SBCTC spells of other students in the treatment group. A parallel restriction was placed on the matching of other-college spells of students in the control group. We imposed this restriction to avoid "washing out" any effects of making control experiences artificially more similar to treatment experiences.

(6)   Lastly, for the NSC-reported spells at other colleges, copy the SBCTC information from the match identified in step 5. The information copied is the information not available in the SBCTC: total full-time-equivalent (FTE) months enrolled, enrollment by quarter, credits earned, and credentials earned.

The following sections give more information for each of these six steps.

Step 1: Exact matching

The team conducted the exact matching of each NSC-reported spell at SBCTC colleges with a SBCTC-reported spell by determining the amount of overlap between the spells, based on the start and end dates of each spell. If only one SBCTC-reported spell overlapped with an NSC-reported spell at an SBCTC college, then the team considered the two spells to be matched without regard to how well start and end dates aligned between the two systems. If multiple SBCTC-reported spells overlapped with one NSC-reported spell at an SBCTC college, then the team considered the SBCTC-reported spell with the most months of overlap to be matched to the NSC-reported spell. If one SBCTC-reported spell overlapped with multiple NSC-reported spells at SBCTC colleges, then the SBCTC-reported spell was broken into pieces that better matched the NSC-reported spells. The team then transcribed the outcomes associated with the SBCTC-reported spell in the SBCTC record system—including total FTE months enrolled, enrollment by quarter, credits earned, and credentials earned—over to the NSC-reported spell.

Step 2: Unmatched NSC-reported SBCTC records

In a small number of cases, NSC-reported spells at SBCTC colleges did not overlap with any SBCTC-reported spells. Data investigation with SBCTC staff determined that many of these cases were due to early course drops or withdrawals. Because the courses were dropped early in the term, they were not included in the file that SBCTC provided to the research team. However, the SBCTC colleges do appear to include these records in extracts they send to the NSC. For these spells, the team assumed that the student had experienced zero hours of instructional credits and earned no credits or credentials.

Step 3: Impute survey outcomes for survey nonrespondents

We reasoned that data on credential attainment from the three-year follow-up survey would help us find better SBCTC matches for NSC-reported spells at colleges other than SBCTC colleges. However, this strategy works only if these data are available on the full sample, including nonrespondents to the three-year survey. To overcome this difficulty, we imputed a small collection of critical survey-measured education outcomes for survey nonrespondents. On the pooled set of all nine Pathways for Advancing Careers and Education (PACE) programs, we imputed eight three-year follow-up survey variables for survey nonrespondents:

- count of credentials earned from taking regular college classes;
- count of credentials earned from taking vocational classes;
- count of licenses and certifications earned from authorities other than schools;
- earning of a short-term college credential;
- earning of a long-term college credential less than a degree;
- earning of an associate degree;
- earning of a bachelor's or graduate degree; and
- self-assessed career progress.

We had considerable information to guide this imputation because about half of the nonrespondents had previously responded to the short-term follow-up survey. This imputation involved modeling the probability of earning a credential requiring a year or more of college study (including degrees) in terms of baseline variables and variables from the short-term follow-up survey,[21] using the predictions to form propensity strata, and then hotdecking within cells defined by site, treatment group, and propensity stratum.[22]

Step 4: Data summarization

We needed to create a parsimonious set of key variables on which to statistically match SBCTC spells to NSC spells at other colleges. The available data about each spell included NSC-reported spell duration and timing, NSC-reported credentials awarded in connection with the

---

[21] This included 21 short-term follow-up survey outcomes that also had to be imputed for nonrespondents to that earlier survey. About 40 percent of nonrespondents to the first follow-up survey responded to the three-year survey, so there was also considerable information available to guide this imputation.

[22] We used quintiles of estimated propensity.

spell, self-reported baseline variables, and the actual or imputed three-year follow-up survey variables from step 3. To facilitate matching, the team created statistical models for four SBCTC-reported spell-level outcomes on the set of exactly matched records in terms of these variables. The SBCTC-reported outcomes were:

- receipt of a degree requiring at least one year of credits by 36 months after randomization;
- receipt of any credential by 36 months after randomization;
- FTE months of academic or vocational coursework by 36 months after randomization; and
- cumulative academic and vocational credits earned by 36 months after randomization.

The procedure involved first fitting models for these four outcomes on NSC-reported SBCTC spells and then using estimated coefficients to predict values for both SBCTC spells[23] and spells at other colleges. The models involved up to 36 characteristics from NSC, the baseline, and the three-year follow-up survey. Additionally the model considered the interaction of each characteristic with treatment.

The baseline variables fell into several categories:[24]

- age: age 21 to 24, age 25 to 34
- gender
- race/ethnicity: Hispanic (of any race); Black, non-Hispanic; another race (neither White nor Black), non-Hispanic[25]
- educational attainment: less than one year of college, one or more years of college, associate degree or higher
- income: less than $15,000, between $15,000 and $29,000
- receipt of public benefits: receipt of WIC or SNAP, receipt of any public assistance or welfare
- current work hours: working 20 to 34 hours per week, working 35 or more hours per week
- intent to attend school only part-time if admitted to I-BEST

The NSC variables in the model search included the number of months with any enrollment in the spell from randomization onward (and at 12, 18, and 35 months after randomization); number of FTE enrolled months in the spell from randomization onward (and at 12, 18, and 35 months); and number of completions in the spell from randomization onward (and at 12, 18, and 35 months).

---

23 This model building ignored SBCTC spells that were not reported in the NSC.

24 For each variable listed below, there was also an excluded category that is not listed.

25 White, non-Hispanic was the excluded category.

The survey variables consisted of the eight outcomes that were imputed in step 3 above: (1) count of credentials earned from taking regular college classes; (2) count of credentials earned from taking vocational classes; (3) count of licenses and certifications earned from authorities other than schools; (4) earning of a short-term credential (i.e. a college credential requiring less than one year of credits); (5) earning of a long-term academic credential less than a degree (those requiring a full year or more's worth of credit, but less than an associate degree); (6) earning of an associate degree; (7) earning of a bachelor's or graduate degree; and (8) self-assessed career progress.

LASSO selection procedures were used in the modeling. To quantify the improvement due to adding survey variables in addition to the baseline and NSC variables used in the prior report, this process was performed twice. In the first iteration, only baseline and NSC variables were included as candidates in the lists—the same as was the case for the short-term report. In the second iteration, the three-year follow-up survey outcomes listed in the prior paragraph were also included. Thus, for each outcome two models were created, one from each of these two iterations. If the two models differed, the better of the two, defined as the one with the higher R-squared, was chosen.

The chosen models from each iteration, along with their R-squared values, are shown in Exhibit B-2. Note that the use of the three-year follow-up survey variables led to a substantial improvement in the model for attainment of a credential (R-squared boosted from 38.2 to 48.6 percent). Using the survey data for the other three SBCTC outcomes provided little to no benefit.

Step 5: Statistical matching

Each college spell was assigned a point in four-dimensional Euclidean space based on the predicted values for the four different spell-level outcomes mentioned in step four.[26] Once each spell was assigned a point, the team could calculate the Euclidean distance from a given spell to any other using the following formula

$$D_{ij} = \sqrt{\sum_{l=1}^{4} \left(z_{li} - z_{lj}\right)^2}$$

Where z_{li} and z_{lj} are the normalized predicted values for outcome l for the potential match between donor i and unmatched spell j. This metric allowed the team to use k-means clustering to sort the spells into six different groups, with the goal of all spells in a given group being "close" to one another. Each non-SBCTC college spell was matched with a random cluster-mate (where the associated student had same treatment status) who had an SBCTC-reported spell with a similar start date.

---

[26]   Where each dimension corresponds to the normalized value of one of the predicted outcomes.

Exhibit B-2:   Predictive Power of Models for SBCTC-Reported Education Spell-Level Outcomes

SBCTC-Based Spell-Level Outcome	Without Three-year Follow-up Variables		Including Three-year Follow-up Variables	
	Independent Variables	*R*-squared	Independent Variables	*R*-squared
Receipt of a degree requiring at least one year of credits by 36 months after randomization (any_ge1year_36_cat)	Baseline: ● Gender NSC: ● FTE months enrolled at 35 months ● Completions at 12 months ● Completions at 35 months	0.496	Baseline: ● Gender NSC: ● FTE months enrolled at 35 months ● Completions at 12 months ● Completions at 35 months Three-year follow-up survey: ● Associate degree	0.498
Receipt of any credential by 36 months after randomization (anycredential_36_cat)	Baseline: ● *None selected* NSC: ● Completions at 35 months	0.382	Baseline: ● Gender ● Associate degree or higher ● Receipt of public benefits ● Treatment*Hispanic ● Treatment*Income between $15,000 and $29,000 ● Treatment*Intent to attend school only part-time if admitted to I-BEST NSC: ● FTE months enrolled at 35 months ● Completions at 35 months ● Completions following randomization Three-year follow-up survey: ● Receipt of a degree requiring less than a year of coursework ● Receipt of a degree requiring a year or more of coursework ● Treatment*Receipt of a credential requiring less than a year of college study ● Treatment*Progress toward career goals	0.486
Full-time-equivalent months of academic or vocational coursework by 36 months after randomization (AcadOrVocFTEMonths)	Baseline: ● *None selected* NSC: ● FTE months enrolled at 35 months	0.789	Baseline: ● *None selected* NSC: ● FTE months enrolled at 35 months Three-year follow-up survey: ● *None selected*	0.789

SBCTC-Based Spell-Level Outcome	Without Three-year Follow-up Variables		Including Three-year Follow-up Variables	
	Independent Variables	*R*-squared	Independent Variables	*R*-squared
Cumulative academic and vocational credits earned by 36 months after randomization (CumAcadVocCredits_Q12)	Baseline: ● Age less than 20 NSC: ● FTE months enrolled at 35 months ● FTE months enrolled following randomization ● Completions following randomization ● Treatment*Completions at 18 months	0.807	Baseline: ● Age less than 20 NSC: ● FTE months enrolled at 35 months ● FTE months enrolled following randomization ● Completions following randomization ● Treatment*Completions at 18 months Three-year follow-up survey: ● *None selected*	0.807

Exhibit B-3 provides statistics about the matching. Spells in cluster 1 were predicted to have low scores on all four variables, whereas spells in cluster 3 were predicted to have high scores on all four variables. Spells in cluster 2 were predicted to have above-average scores on three of the four variables. Spells in cluster 4 were nearly average.

Members of the treatment group had 19 spells at non-SBCTC colleges. Each of these was matched to a single one of the 382 SBCTC spells experienced by members of the treatment group. Similarly, members of the control group had 28 spells at colleges other than SBCTC colleges. Each of these was matched to a single one of the 247 SBCTC spells experienced by members of the control group. Spells in cluster 5 were predicted to earn credentials with few months of study and few credits. Spells in cluster 6 were predicted to be above average on all four outcomes, but not as high as spells in cluster 3.

Exhibit B-3:   Cluster Structure and Spell Counts for Matching Non-SBCTC Spells to SBCTC Spells

	Predicted Standardized Scores				Spell Counts	
	Degree	Any Credential	FTE Months	Credits	Treatment	Control
Cluster						
1	−0.7	−0.5	−0.6	−0.6	138	186
2	0.9	−0.2	1.5	1.5	26	33
3	3.0	2.0	2.7	2.7	14	23
4	−0.1	−0.4	0.1	0.1	48	67
5	0.3	1.9	−0.3	−0.3	8	44
6	1.5	2.0	0.9	1.0	10	18
NSC-Reported College						
At colleges other than SBCTC colleges					19	28
At SBCTC colleges					382	247

Source: National Student Clearinghouse; SBCTC records; PACE Basic Information Form; PACE Self-Administered Questionnaire; PACE three-year follow-up survey.

Step 6: Propagating SBCTC values

The final step entailed copying matched data on total FTE months enrolled, enrollment by quarter, credits earned, and credentials earned from SBCTC records to serve as values for spells at other colleges. The procedure involved both the exact matching and the statistical matching. The outcomes for an NSC-reported spell at another college were copied over from SBCTC-reported outcomes of the SBCTC-reported spell that had been exactly matched to the NSC-reported SBCTC spell that had been statistically matched to the other-college spell.[27] We did this separately for every outcome based on SBCTC records.

---

[27] Let ON1 be an NSC-reported spell at a college other than a SBCTC college. Let PN1 be the NSC-reported spell at an SBCTC college that was statistically matched to ON1 in step 5. Let PP1 be the SBCTC record that was exact matched to PN1 in step 1. Then the outcomes on PP1 were transcribed over to ON1.

Appendix C: Three-Year Survey Data

This appendix documents key technical detail underlying analyses of the three-year follow-up survey data.[28] Section C.1 documents coding for scales based on follow-up survey data. Section C.2 describes the imputation process for some missing survey data elements in the construction of outcomes. Section C.3 analyzes survey nonresponse and documents the process we used to build the nonresponse weights used in the impact analysis. Sections C.4 and C.5 present evidence about the quality and completeness of survey responses. Before getting into those details, we provide an overview of the measurement goals and structure of the survey instrument.

The survey sought to collect a complete history of jobs and periods of schooling since randomization (including the progression and interleaving of these spells), credits and credentials earned; earnings growth, and self-employment. In addition, the survey measured several psycho-social skills, family formation and growth, income and material well-being, and child outcomes.

The Integrated Training and Employment History module of the three-year survey aimed to collect a complete history of training and employment between randomization and the day of interview three years later. Given data collection plans, the approach needed to work over the phone. The instrument development team reviewed several past efforts to collect such histories, but only one of the past approaches seemed likely to be workable over the phone—an approach developed for a German survey instrument that studies the training and work histories of German youth.[29] This was the first time that the German approach had been attempted in the United States.

Conceptually, a history could be built either forward from randomization or backward from the day of interview. The German study worked forward with apparent success, so we adopted that approach. One modification we made was to take each respondent through his or her training and employment history twice instead of just once. First, the survey collects the spell history (dates, whether work or school, and place names). This is the "scaffolding." Once the scaffolding has been built, the interviewer takes the respondent back through the history a second time to systematically collect more information about each training spell.

---

[28] The full instrument is available at http://www.career-pathways.org/career-pathways-pace-three-year-instrument/.

[29] The 2011 BIBB Transitional Study was a retrospective longitudinal survey conducted by the *Bundesinstitut für Berufsbildung* (Federal Institute for Vocational Education and Training) on a representative basis that recorded in detail the whole of the educational and occupational biographies of persons born between 1987 and 1992 and resident in Germany. For full details, see Beicht and Friedrich (2008). For a brief English synopsis of one report using some of the survey data: http://www.bibb.de/en/64317.htm.

There are two motivations for this two-pass approach:

1. By asking the respondent to focus on one type of information at a time, collection of date data may be more consistent across spells.

2. This approach allows more-straightforward programming.

C.1 Measures Based on Follow-up Survey Data

Exhibits in this section detail the operationalization of survey-based outcomes used in impact analyses in the main report. These exhibits also reference the underlying survey questions. Exhibit C-1 provides details on outcomes in the education domain, as reported in Chapter 3. Exhibit C-2 provides similar details on outcomes in the employment/earnings domain as reported in Chapter 4. Finally, Exhibits C-3 and C-4 do the same for intermediate outcome domains and other life outcomes domains, respectively, as reported in Chapter 5.

Exhibit C-1: **Details on Specifications for Survey-Based Education Outcomes in Chapter 3**

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Secondary Outcomes		
Education		
Received an exam-based certification or license	Respondents were asked whether they had "received a professional, state, or industry certification, license, or credential from an authority other than a school." This measure uses the18-month survey for exam-based credentials reported through the time that survey was completed and uses the three-year survey for exam-based credentials that were reported to be earned after completion of the short-term survey.	3-year: I3d, I3di, I3h 18-month: A56, A56a
Exploratory Outcomes		
Full-time-equivalent months enrolled at any school through 35 months after randomization	Students were asked for the dates of attendance of each school attended and their status while enrolled. If their status was "part-time," then the number of months was multiplied by 0.25 to estimate full-time-equivalent months. Similarly, if their status was "equal mix," then number of months was multiplied by 0.50 to estimate full-time-equivalent months. We developed this rule based loosely on guidance in NSC documents about how schools should classify less-than-full-time enrollment. Because the survey response categories were different from those used in the NSC and because students might have different understandings than schools did, this decision was fairly arbitrary. Alternate rules might have worked just as well.	C2, C3, D2
Received any type of credential from any school	Respondents were asked whether they had received "a diploma, certificate, or academic degree for completing any regular college classes" and whether they had received "any diplomas or certificates from a school for completing any vocational training."	I2, I2c, I3, I3c

Key: NSC = National Student Clearinghouse.

Exhibit C-2:   Details on Specifications for Survey-Based Employment/Earnings Outcomes in Chapter 4

Outcome	Details on Derivation of Outcome	Follow-up Survey Question(s)
Secondary Outcomes		
Employed at survey follow-up	Determined based on reported dates of jobs and date of interview.	Most of modules B, C, and E
Career Progress		
Employed at $14 per hour or above	Analyzed response to survey question for control group. Selected $14 per hour as the threshold because it was close to the 60th percentile of hourly wages among employed control group members. This percentile was picked as being a reasonable goal for graduates of I-BEST.	F5
Employment in job requiring at least mid-level skills	Three open-ended questions about the kind of work done, the usual activities completed, and the job title were coded into an SOC code. We then looked up the Job Zone[a] for each SOC code in the O*NET system.[b] Job Zone 3—occupations that need medium preparation—seemed a reasonable goal for graduates of I-BEST.	G2a, G3, G4
Exploratory Outcomes		
Works at least 32 hours per week	Currently employed respondents were asked about their typical hours worked.	F6
Currently employed, working straight day, evening, or night shifts	Currently employed respondents were asked about their typical work schedule. Answer possibilities included straight shifts, rotating shifts, split shifts, irregular schedules, and other.	G6, G6a
Currently working in a job that offers health insurance	Currently employed respondents were asked whether health insurance was available through the employer as a fringe benefit.	G8a
Currently working in a job with a supportive working environment	Questions about job benefits and conditions were used to cluster jobs into three categories. Jobs in this category generally provided employees with flexibility to balance work and family, a supportive set of co-workers and supervisors, a rich set of benefits, and opportunities for advancement.	G7, G8a-G8e, G9, G10
Working in a healthcare occupation	Three open-ended questions about the kind of work done, usual activities completed, and the job title were coded into a SOC code. If the first two digits of the SOC were 29 (Healthcare Practitioners and Technical Occupations) or 31 (Healthcare Support Occupations), then the respondent was considered working in a healthcare occupation.[c]	G2a, G3, G4
Quarterly earnings (for Q1-Q12)	Respondents were asked to provide a complete history of jobs, as well as the starting/ending wage and hours for each job. We combined these to establish weekly earnings for the first and last weeks of a job and then interpolated to get wages for each intervening month. We then summed weekly wages across jobs for multiple-job holders to get weekly earnings for every week between randomization and interview. Using this data, we were able to calculate earnings for each quarter.	F1, F1a, F1b, F1c, F4, F5, F5a, F6, F7, F7a, F8

Key: SOC = U.S. Department of Labor Standard Occupational Classification.

[a] https://www.onetonline.org/help/online/zones [accessed September 12, 2016].

[b] https://www.onetonline.org/ [last accessed September 12, 2016]. There are five Job Zones. A Job Zone is a group of occupations that are similar in education needed to do the work, related experience needed to do the work, and amount of on-the-job training needed to do the work. Job Zone 3 is described in the O*NET system documentation as "Employees in these occupations usually need one or two years of training involving both on-the-job experience and informal training with experienced workers. A recognized apprenticeship program may be associated with these occupations."

[c] Being employed in a healthcare occupation is usually associated with employment in the healthcare industry, but this is not always true. School nurses are one example of a healthcare worker being employed in an industry other than healthcare. Conversely, many people employed in the healthcare industry are not healthcare workers. Hospital janitors are one example. The survey did not ask about industry of employer.

Exhibit C-3:   Details on Specifications for Survey-Based Intermediate Outcomes in Chapter 5

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Secondary Outcomes		
Access to career supports	This was a new scale created for PACE at the 18-month follow-up. It is a six-item scale counting number of types of career-supportive relationships in workforce and education settings. The motivation for creating this scale was the theory that richer social networks are one of the benefits of higher education (e.g., Goldrick-Rab and Sorensen 2010). Say you need advice of help in taking a next step on a career pathway of interest to you. Please tell me if there is anyone you'd be comfortable turning to: ● Who has a college degree? ● Who is currently going to college? ● Who works at a local college, either as a teacher or staff member providing help to applicants or students? ● Who works for a local community organization helping people find education and training, work, and related supports? ● Who works in an occupation of interest to you? ● Who has a management job in a work setting matching your career interests?	K4
Confidence in career knowledge	This seven-item scale was based on a review of six survey instruments as well as literature. The first two scale items (a, b) were adapted from the Career Decision Self-Efficacy–Short Form (Betz and Taylor 2001). Three items (d, e, f) were adapted from the Career Exploration Survey (Stumpf et al. 1983). Two items (c, g) were new and written specifically for the PACE Basic Information Form. Response categories ranged from 1=strongly disagree to 4=strongly agree. a. You know how to accurately assess your abilities and challenges? b. You know how to make a plan that will help achieve your goals for the next five years? c. You know how to get help from staff and teachers with any issues that might arise at school? d. You know the type of job that is best for you? e. You know the type of organization you want to work for? f. You know the occupation you want to enter? g. You know the kind of education and training program that is best for you?	K6
Exploratory Outcomes		
Perceived career progress	This was a new scale created for PACE at the 18-month follow-up. It is a three-item scale of self-assessed career progress. Response categories range from 1=strongly disagree to 4=strongly agree. It was designed specifically to measure a respondent's sense of progress in a career pathways program as described by Fein (2012). ● I am making progress towards my long-range educational goals ● I am making progress towards my long-range employment goals ● I see myself on a career path	I5, I6
Grit	Existing scale from Duckworth et al. (2007). The eight-item scale captures persistence and determination. Response categories ranged from 1=strongly disagree to 4=strongly agree.	K1

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Core self-evaluation	Existing scale from Judge (2009). The 12-item scale response categories ranged from 1=strongly disagree to 4=strongly agree. Core self-evaluations (CSEs) represent a stable personality trait that attempts to capture one's self-perception. A positive self-image will correspond to a higher CSE, whereas those who view themselves more negatively will score lower in this category. This trait involves four personality dimensions: locus of control, neuroticism, generalized self-efficacy, and self-esteem. Various studies have shown CSE scores to have predictive ability for work outcomes such as job satisfaction and job performance.[a]	K3
Life Challenges Index	A new scale adapted for PACE from a longer instrument by Kessler et al. (1998). Average of five items of frequency of situations that interfered with school, work, job search, or family responsibilities. The response categories ranged from 1=never to 5=very often. Missing if four or more responses are blank. (childcare, transportation, alcohol or drug use, health, family arguments, physical threats). ● Childcare arrangements ● Transportation ● Alcohol or drug use ● An illness or health condition ● Another situation	K7
Social Support Index	Existing scale from Hoven (2012). The 10-item scale response categories ranged from 1=strongly disagree to 4=strongly agree. It is a short-form version of the Social Provisions Scale of Cutrona and Russell (1987), a scale that has 24 items.	K5
Stress Index	Existing scale from Cohen et al. (1983). This scale was first used in the PACE Basic Information Form and has since then been included in both follow-up instruments. The four-item scale captured perceived stress. The response categories ranged from 1=never to 4=very often.	K8

[a] Judge, Locke, and Durham (1997, 1998); Judge and Bono (2001).

Exhibit C-4:   Details on Specifications for Survey-Based Other Life Outcomes in Chapter 5

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Secondary Outcomes		
Personal student debt	Students were asked about personal borrowing to go to school since randomization. For those who had difficulty answering the question about the exact amount, a categorical response option was offered. These were then imputed to continuous levels.	M6, M6a
Has health insurance coverage	Includes the offer of healthcare by employer or actual receipt if not offered by employer.	G8a, M12
Receives means-tested public benefits	Respondents were asked whether they or anyone else in their household received TANF, SNAP, WIC, Medicaid, subsidized childcare, Section 8 or Public Housing, LIHEAP, or FRPL.	M3a, M3b, M3c, M3e, M3f, M3g, M3h, M3i
Any signs of financial distress	For the three-year follow-up, this scale is an expanded version of the financial hardship measure used in the 18-month follow-up survey. It flagged any signs of financial distress in terms of troubles paying bills (rent/mortgage, gas/oil/electricity), utility disconnects (gas/electric/oil, telephone), delayed healthcare, delayed dental care, delayed prescription drug procurement, not having enough to eat (sometimes or often), or not having enough money to make ends meet at the end of the month.	M9a-g, M10, M11
Exploratory Outcomes		
Personal income	Respondents were first asked to provide an open-ended amount for the prior month, specifically excluding income tax refunds. If no answer was given, the respondent was asked to choose one of seven bracketed amounts. Item nonresponse was multiply imputed. Exact amounts were also multiply imputed for people who chose a bracket.	M2, M2a
Household income	Respondents were first asked to provide an open-ended amount for the prior month, specifically excluding income tax refunds, where the household was clarified to include anyone who lived in the household for at least half of the prior month. If no answer was given, the respondent was asked to choose one of seven bracketed amounts. Item nonresponse was multiply imputed. Exact amounts were also multiply imputed for people who chose a bracket. People who lived alone were not asked this question. Instead, their personal income was assumed to equal the household income.	M4, M4a
Unsecured debt of $5,000 or more	Respondents were asked about debt other than student debt and secured debt (such as mortgages or title loans). Debts in the name of their spouse or partner were included.	M8
Parental student debt	Respondents were asked about borrowing by parents on behalf of the student to go to school since randomization. For those who had difficulty answering the question about the exact amount, a categorical response option was offered. These were then imputed to continuous levels.	M7, M7a

Didn't experience food insecurity	Respondents were asked about adequacy of household food over prior six months. The possible responses were:
	    1=Enough of the kinds of food you want
	    2=Enough but not always the kinds of food you want
	    3=Sometimes not enough to eat
	    4=Often not enough to eat
	Response of 1 or 2 counts as did not experience food insecurity.
		M10

Key: FRPL = free or reduced-price lunch. LIHEAP = Low Income Home Energy Assistance Program. SNAP = Supplemental Nutrition Assistance Program. TANF = Temporary Assistance for Needy Families. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

C.2 Imputation in the Three-Year Survey

As in any survey, some respondents did not answer every question. We used a variety of approaches to allow us to use these cases despite their partial responses. Our approach varied across questions, depending on whether the question was embedded in a sequence of questions in which all questions needed to be answered to calculate the value of a scale, whether the question was embedded in a block of unanswered questions, and the frequency of nonresponse to the question across respondents.

The default rule was to drop persons from any analysis involving unanswered question but to include them for all other analyses. Where this rule would result in a sharp drop in sample size—either for the question by itself or for a scale involving the question—then we instead imputed responses for those people for those questions, rather than dropping them. Additionally, we imputed blocks of responses for two groups of people: those with large blocks of missing data and those who, based on administrative data, appeared to have failed to report one or more education spells.

The goals of imputation were variance and bias reduction.[30] Both goals are achievable with the rich set of parallel outcomes measured in the three-year survey. For example, indications of problems paying bills is valuable information for imputing missing income. Specifically, we imputed seven types of missing data:

1. number of college credits;

2. credential award dates;

3. income (personal and household);

4. early certifications and licenses (first 18 months after randomization);

5. skipouts (i.e., missing data on spells caused by trying to avoid respondents ending the survey);

6. spell start and end dates (job spells and school spells); and

7. survey data on school spells reported to the National Student Clearinghouse (NSC) but not by respondent.

This section briefly describes each of these imputations and their prevalence. We used a common methodology for the first four types of missing data. Section C.2.1 provides the detail on these imputations. Section C.2.2 gives details on the imputation methodology for the other three types of missing data.

---

[30] Systematic nonresponse (e.g. those without college credentials are less likely to answer questions about credential attainment) can cause biased estimates. Effective imputation can reduce this bias. Making use of more data also increases sample size, thereby reducing the variance of impact estimates.

Types and Rates of Imputation. Exhibit C-5 below lists the seven types of imputation and shows the imputation rates for the survey respondents in the evaluation sample for the Integrated Basic Education and Skills Training (I-BEST) program. The instrument asked about credits spell by spell. It was fairly common for respondents to be unable to recall the number of credits they had earned during one or more training spells. They also had trouble recalling the dates on which they received credentials. Income was also frequently missing. The instrument prompted respondents to give a categorical answer ("bracketing") if they could not give an exact figure.

Exhibit C-5:   Imputation Rates among Survey Respondents in I-BEST

Type of Imputation	Job Spells (%)	School Spells (%)	Credentials (%)	People (%)
1. Number of college credits	n/a	n/a	n/a	19.6
2. Credential award dates	n/a	n/a	6.9	n/a
3. Income				
Personal (categorical)	n/a	n/a	n/a	2.9
Personal (exact)	n/a	n/a	n/a	4.1
Household (categorical)	n/a	n/a	n/a	10.3
Household (exact)	n/a	n/a	n/a	22.9
4. Early certifications and licenses	n/a	n/a	n/a	12.4
5. Skipouts	7.6	5.9	5.2	5.5
6. Spell start and/or end dates (job, school)	6.7	9.8	n/a	n/a
7. Survey data on school spells reported to NSC but not by respondent	n/a	9.3	4.7	8.6

Source: PACE three-year follow-up survey.

Note: Exact income was missing more often than categorical income because respondents unable or unwilling to provide an exact amount were encouraged to report a bracketed amount. n/a indicates not applicable.

The "Early Certifications and Licenses" row refers to the rate of study participants who were not interviewed at 18 months after randomization but who were interviewed at three years. This imputation involved creating a composite scale using the 18-month interview to measure receipt in the first 18 months and the second interview to measure receipt in the second 18 months. Section C.4 provides information about the rationale for this composite scale.

The "Skipouts" row refers to block missingness in the survey's Integrated Training and Employment History module. The German survey upon which this module was modeled experienced a high level of breakoff (12 percent; Beicht and Friedrich 2008), meaning people discontinued the interview midstream and declined to restart it. To prevent similar problems for this three-year analysis, the PACE survey added a skipout feature in the module. If a person refused to answer any question in the module or gave a response of "don't know" to any of several critical flow-controlling questions in the module, the interview flow automatically skipped ahead to the next modules (e.g., on 21st century skills, family structure, income and material

well-being, and child outcomes).[31] With this approach, complete interview breakoffs were nearly eliminated, but a large block of missing data was created for about 7 percent of respondents (across the entire PACE three-year sample) and 5.5 percent of I-BEST treatment and control group respondents—much lower than the breakoff rate on the German study, but still high enough to require special attention.

Nonresponse was non-negligible for start and end dates of both job and school spells, particularly start dates. This is not surprising given that the reference period was up to three years long (and longer for people interviewed later in the survey period and for spells that started prior to randomization).

The final row of Exhibit C-5 refers to an adjustment for undercoverage of NSC-reported spells. This adjustment started with a match of survey reports with administrative data on college attendance from the NSC. We flagged respondents who had spells of college attendance according to the NSC but who did not themselves report any training (college or other type of school) since randomization. Although the NSC is not error-free, its enrollment coverage is generally high (see Appendix D). Accordingly, we imputed all the data from the matched NSC spells to survey respondents who did not report such spells.

C.2.1. College Credits, Credential Award Dates, Income, and Early Certifications and licenses (Imputations 1-4)

As mentioned above, four of the seven types of imputation utilized a common imputation procedure: college credits, credential award dates, income, and certifications and licenses in the first 18 months. This section discusses the basic procedures used and provides additional details for each of the four types of missing data.

Core Imputation Procedure. The core imputation methodology involved a number of steps. The first step entailed assembling a list of potential predictors and imputing any missing data in them.[32] The list of potential predictors included program, treatment status, the interaction of program with treatment status, baseline variables, parallel outcomes, and two-way and three-way interactions of both baseline variables and parallel outcomes with program and treatment status.

The second step entailed the use of a cross-validated LASSO procedure to fit a linear model for the target variable in terms of the assembled predictor list.[33] We did this on a pooled dataset that contained respondents from all nine PACE sites (*n*=6,773, of whom 5,910 responded to both follow-up surveys) and some respondents from Health Profession Opportunity Grants

---

[31] The original intent was not to skip past questions about credential attainment and current job conditions, but a mistake in the specifications caused these sections to also be skipped.

[32] The only purpose of the imputation of potential predictors was to facilitate automated variable selection in the next step. After we used these imputed values of the predictors to predict new exam-based certifications and licenses as of the time of the 18-month survey, we discarded them. We carried out this imputation with SAS/MI/FCS.

[33] See Appendix A.3 for details on the cross-validated LASSO.

(HPOG)-only programs, as well.[34] Note that though this procedure allowed program, treatment, their interaction with each other, and their interactions with many other variables to enter the model, it did not force any of them in. We discuss the implications of this decision after first finishing a description of the procedure.

The third step used predicted values from the final linear model to create a nested set of three partitions for each combination of site and treatment status.[35] The finest partition involved splitting the sample into 20 equal-sized groups based on the predicted probability of having reported an exam-based certification or license if respondents had been interviewed at 18 months. The middle partition corresponded to deciles of this same probability, and the coarsest partition corresponded to quintiles of this same probability.

The fourth and final step used the hotdeck imputation procedure in SUDAAN to randomly match each nonrespondent with a respondent within cells defined by PACE program, PACE treatment status, and the nested partitions. Most cases were matched within cells defined by the 20-level partition. If there were no matches within those cells, then the procedure sought matches within the coarser partitions, first with the 10-level version and then with the five-level version if necessary. If even that did not permit a match, then the procedure randomly matched any unmatched nonrespondents with any respondent in the same PACE program with the same treatment status.

We ran the final hotdeck procedure five times with different random seeds to produce multiple imputations. We used these multiple imputations in the formal analysis runs to add between-imputation onto the naïve variance estimates on the full sample, using Rubin's classic formula.[36]

We now return to the implications of our decision not to force the interactions of site and treatment group with every other variable in the model. First, it is critical to note that we constrained matches to be from the same site and treatment group. This provided strong protection against imputation-caused bias in the estimated treatment impact. We used the models from the pooled dataset only to guide the matching of respondents and nonrespondents *with the same treatment status in the same site*. One way to think of this is that we used the pooled dataset to define a distance metric that we then applied within site and treatment group. An alternative procedure would have been to just randomly match respondents and nonrespondents within cells defined by site and treatment group. The point of using a distance metric rather than randomly matching is to reduce variance and the possibility of nonresponse bias. For a site with a large sample size, forcing in all the interactions of site and treatment

---

[34] ACF's Health Profession Opportunity Grants (HPOG) Program, like PACE, provides training to low-income individuals, but only for healthcare occupations. The impact study of 32 first-round HPOG awardees (HPOG 1.0) included three awardees and one subgrantee (Carreras en Salud) also studied in PACE. For more: https://www.acf.hhs.gov/ofa/programs/hpog.

[35] A "partition" of a sample is an exhaustive and mutually exclusive collection of subsets of the sample.

[36] See for example, Rubin (1987).

group with other variables might not cause much deterioration in model quality, but in small sites forcing would almost certainly have made it more difficult to detect subtle main effects.[37]

Life Trajectory Clusters. The survey contained multiple measures of financial and social-emotional well-being. We theorized that these variables would be useful predictors of several types of missing data, particularly the missing data created by skipouts because none of these questions were involved in the bad skip pattern. However, interpretation of high-dimensional models is difficult. As a way of incorporating these rich data on well-being into imputation models while still keeping the models fairly easy to interpret, we condensed all these measures into a partition of the sample using cluster analysis. We were able to identify five clusters of respondents who vary clearly in terms of quality of life and core self-evaluation and family dependence. For shorthand, we refer to them as "life trajectory" clusters because one of the variables that they vary on most clearly is a sense of career progress:

- "Overextended"—above average income but also above average financial stress and low scores on psycho-social skills.

- "Family supported"—below average income but strong family supports that protect them from financial stress.

- "Strivers"—strong psycho-social skills and sense of career progress but low income (personal and household) and dependent on public support.

- "Down and out"—very low psycho-social skills, low sense of career progress, severe life challenges, low income (personal and household), and strong reliance on public support.

- "Winners"—strong psycho-social skills and sense of career progress, high income (personal and household), few bill problems, and little dependence on either family or public support.

Missing College Credits

For missing credits, we assembled a rich set of predictors from the PACE Basic Information Form (BIF) and the Self-Administered Questionnaire (SAQ) administered at baseline, the NSC, the 18-month follow-up survey, person-level scales in the three-year survey, and spell-level data from the School Experiences module of the three-year survey. This was a spell-level file pooling data across the nine PACE sites, but not HPOG-only sites as no NSC data were available for the HPOG-only sample. We also added a large number of two- and three-way interactions with site and treatment group. After creating dummy variables for categorical variables, the total number of potential predictors was 1,584. The LASSO procedure working on this predictor set selected just six variables, yielding a model with an *R*-squared of 27 percent. Four of the six variables were significant predictors with standardized regression coefficients of at least 0.01. They were:

- adjusted spell duration (adjusted for the longest break);

---

[37] Algorithmically, the way to force in all interactions is to run the LASSO on a dataset restricted to just the cases in a particular site and treatment group. Even for the largest PACE site, this would not have provided nearly as much power to detect subtle main effects.

- spell duration interacted with full/part-time student status;

- credits reported at 18 months; and

- NSC-reported full-time-equivalent months of enrollment through 35 months after randomization.

After controlling on the six factors, program and treatment were not important and nor were any of their interactions with each other or with other predictors. After imputing credits at the spell level, we summed to the person level for respondents with multiple school spells.

Missing Credential Award Dates

On the pooled PACE/HPOG credential sample, we modeled the lag between randomization and credential award date for those respondents with reported award dates (n=12,392, with 11,628 responses). The potential predictor list included site, treatment, the interaction of site with treatment, type of credential (10 categories), life trajectory cluster, 20 parallel outcomes at the person level, the lag between randomization and interview, 16 baseline variables, and a large set of two- and three-way interactions with site and treatment group. After creating dummy variables for categorical variables, the total number of potential predictors was 1,160. The LASSO procedure working on this predictor set selected 14 variables, yielding a model with an R-squared of 8.4 percent. The significant predictors with standardized regression coefficients of at least 0.01 were:

- HPOG versus PACE;

- credential was awarded for regular college classes and typically takes less than a year of study to earn;

- credential is an associate degree;

- credential is a bachelor's degree;

- self-assessed career progress;

- student debt;

- two interactions of HPOG with main effects;

- one interaction of treatment status with a main effect; and

- two 3-way interactions of HPOG status with treatment status with main effects.

After matching nonrespondents with respondents, we adjusted for the difference in randomization dates between the two people, by adding the lag from the respondent to the randomization date for the nonrespondent. If this was past the interview date for the nonrespondent, we truncated the award date to equal the interview date.

Missing Income

The instrument yielded four related measures of income in the past month: (1) exact personal income; (2) categorical personal income; (3) exact household income; and (4) categorical household income. As could be seen in Exhibit C-5, missing data rates were considerably higher for the continuous variables than the categorical variables. This is because categorical income is missing only if both exact (which can be put in the appropriate income category) and

categorical income are missing. For prediction purposes, we assembled a person-level file with program, treatment status, the interaction of program with treatment status, self-reported earnings by quarter, 10 variables about economic well-being, four variables about psycho-social skills, nine measures of educational progress, 12 baseline characteristics, and a large collection of two- and three-way interactions with site and treatment group. We used this list for modeling both personal and household income. We ran the LASSO on the pooled PACE/HPOG three-year dataset (n=14,467, with 12,782 exact personal income reports and 9,219 exact household income reports). After creating dummy variables for categorical variables, the total number of potential predictors was 1,414.

The LASSO procedure working on this predictor set selected 11 variables for personal income, yielding a model with an R-squared of 58 percent. The significant predictors with standardized regression coefficients of at least 0.01 were:

- dummy variables for three of the five life trajectory clusters;
- personal earnings for the 12th quarter after random assignment;
- a dummy variable for having earned an associate degree since randomization;
- a scale for being able to make ends meet at the end of the month; and
- an interaction of earnings with a dummy for receipt of any means-tested public benefits.

For household income, the LASSO procedure selected 26 variables, yielding a model with an R-squared of 52 percent. The significant predictors with standardized regression coefficients of at least 0.01 were:

- dummy variables for three of the five life trajectory clusters;
- personal earnings for the 12th quarter after random assignment;
- a dummy variable for being an Earned Income Tax Credit claimant;
- a dummy variable for living with a spouse;
- a dummy variable for living with one's parents;
- a dummy variable for living alone;
- annual baseline family income below $15,000;
- baseline SNAP (Supplemental Nutrition Assistance Program) or WIC (Special Supplemental Nutrition Program for Women, Infants, and Children) receipt;
- a dummy variable for having earned an associate degree since randomization;
- a scale for being able to make ends meet at the end of the month;
- an interaction of earnings with a dummy for receipt of any means-tested public benefits;
- an interaction of personal earnings with living arrangements; and
- three 2- and 3-way interactions involving program.

Note that neither the model for personal income nor the model for household income involves three-way interactions of program with treatment status that are both statistically significant and

substantively large. This does not mean that there are no program effects on income. Rather, it means that the measured parallel outcomes already capture whatever program effects might be present.

Certifications and licenses in the First 18 Months

As mentioned earlier and as is discussed in detail in Section C.4 below, measures of ever-receipt of certifications and licenses blended reports from the 18 and three-year surveys. This decision also required imputing what nonrespondents to the 18-month survey would have reported if they had responded at that time.[38] We used the core imputation described above for this imputation.

On the pooled PACE three-year survey respondent sample (*n*=6,773 people, of whom 5,906 responded to both the 18-month and three-year follow-up surveys and 867 responded to only the three-year survey), we modeled the receipt of such credentials among those who responded to the 18-month follow up. The potential predictor list included program, treatment status, the interaction of program with treatment status, and about 40 baseline and three-year follow-up variables. After creating dummy variables for levels of categorical variables, this led to 80 potential predictors in total.

The LASSO selected 10 of the 80 predictors, yielding a model with an *R*-squared of 12.0 percent, a high value for a binary outcome. The selected variables included treatment status, dummy variables for two programs, one treatment-by-program interaction, five measures of educational progress and well-being at three years, and a dummy variable for employment in healthcare at three years. Of these, the predictors with standardized coefficients of at least 0.01 were:

- treatment status;
- one dummy variable for site;
- one treatment by site interaction;
- number of licenses obtained at three years;
- report of a short-term college credential at three years;
- report of a long-term college credential at three years; and
- current employment in healthcare.

After imputing new exam-based certifications and licenses for 18-month survey nonrespondents, we separated exam-based certifications and licenses reported in the three-year survey using the donor's interview date into two categories—early (would have been reported by the nonrespondent in the 18-month survey if the interview had taken place) versus late (would have been earned after the 18-month survey if the interview had taken place). We then created a blended flag for having earned an exam-based certification or license as of the

---

[38] Nonrespondents here were people who could not be located, refused to be interviewed, or were otherwise unavailable for an interview. The concept does not include people who skipped questions about credentials when interviewed at 18 months. We assumed that these respondents did not earn any credentials by the time of the 18-month interview.

three-year survey. The flag was set to yes if the 18-month nonrespondent had an imputed early exam-based certification or license or had reported a late exam-based certification or license in the three-year survey.

C.2.2   Skipout, Start and End Dates, and Unreported School Spells

The remaining three types of missing data required more customized procedures. This section provides details on the approach to each type.

Skipout

We considered several approaches to this type of missing data. One option we considered and rejected was to treat respondents with skipouts as nonrespondents and give them nonresponse-adjusted weights of zero. This simple option would have significantly boosted the overall nonresponse rate and wasted information collected after the skipout. A second rejected approach would have been to treat respondents with skipouts as nonrespondents only for analyses involving educational progress and employment. This option would have required the creation of a second set of nonresponse-adjusted weights, and would have led to inconsistencies across analyses. A third rejected option was to impute each outcome and scale requiring any data from the Integrated Training and Employment History module. This option was more attractive but would not have supported estimation of career trajectories.

The approach we adopted was to use a block imputation approach that was initially used in medical expenditure surveys in the United States (Williams and Folsom 1981). The general method involves matching a nonrespondent to a respondent and then copying the entire block of missing data from the respondent to the nonrespondent. Our objective was to find a respondent whose training and employment history would align well with the nonrespondent's baseline characteristics and measures of well-being at three years. If the matched person had a missing response to a question within the Training and Employment History module, we copied this missing value over the skipout along with all the other variables.

We used sequential hotdecks as in the core imputation methodology, but we formed the partitions in a different manner. Rather than modeling a single variable and then forming a nested set of partitions based on model-based predictions of that single variable, we crossed the life trajectory clusters discussed above with other important measures. We used a sequence of four hotdecks, where the first had the most stringent criteria for matches, and each succeeding hotdeck had loosened criteria.

The first hotdeck matched nonrespondents to respondents within cells defined by program, treatment status, any schooling reported prior to skipout, any work reported prior to skipout, life trajectory cluster, and lag between randomization and interview in whole months. This was on the pooled PACE/HPOG sample (n=14,169, with 13,245 respondents who did not skip out).[39] This run found donors for 815 of the 924 skipouts on the pooled dataset.

---

[39]   This excludes 302 three-year survey respondents who reported no training or employment between randomization and the survey interview.

The second hotdeck replaced program with site. This run found donors for 86 of the remaining 109 skipouts on the pooled dataset. The third hotdeck replaced the exact number of months in the reference period with a dummy variable for whether the number was greater than 38 months. This run found donors for 22 of the remaining 23 skipouts on the pooled dataset. The fourth hotdeck used a collapsed version of self-assessed goal progress in place of life trajectory cluster and the binary recode of length of the reference period. This found a donor for the last remaining skipout.

Given the challenges in matching many of the nonrespondents to appropriate respondents, we did not carry out multiple imputation for skipouts. For the imputation of skipouts, our judgment was that the donor pools would be frequently small and that multiple random matches would, in fact, be the same match over and over. This lack of variation in the matched donors would have rendered variance estimates based on multiple imputations little better than variance estimates based on single imputation.

Because I-BEST respondents with skipouts were missing a long stretch of data that are important to most of the secondary outcomes in this report, we prepared impact estimates with and without these cases, as displayed in Exhibit C-6. The two sets of impact estimates are similar, although imputing skipouts did shift the significance of the estimated impact for employment at survey follow-up past the 10 percent level. Similarly, imputing skipouts pushed the significance of the estimated impact for employment at $14 per hour or above past the 5 percent level. The imputation allowed us to use as many as 23 more cases for I-BEST (about 5 percent of the respondent sample), with the exact count depending on item nonresponse.

Exhibit C-6:   Comparison of Selected Impact Estimates of I-BEST

Outcome and Sample	Impact Estimate	Standard Error	Sample Size	*p*-Value
Employed at Survey Follow-Up (%)				
Full sample	6.7*	5.1	419	.094
Omitting skipouts	5.4	5.2	396	.149
Employed at $14 Per Hour or Above (%)				
Full sample	7.5**	4.5	409	.047
Omitting skipouts	6.1*	4.5	386	.089
Employed in a Job Requiring at Least Mid-Level Skills (%)				
Full sample	−3.4	3.0	410	.871
Omitting skipouts	−3.4	3.0	387	.874
Receipt of an Exam-Based Credential (%) (blended three-year and 18-month surveys)				
Full sample	10.5**	5.0	419	.016
Omitting skipouts	9.3**	5.1	396	.034

Source: PACE three-year follow-up survey; PACE 18-month follow-up survey.

Note: "Full sample" rows include values imputed for skipouts. All estimates are regression-adjusted as discussed in Appendix Section A.3. Statistical significance levels based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt) are: *** 1 percent level; ** 5 percent level; * 10 percent level.

Spell Start and End Dates

As mentioned earlier, respondents were frequently unable to remember dates. We decided to impute them to make the most use of the partial information in each respondent's reported history. Our primary objective was to create a high-quality measure of the duration of study over the entire reference period. Secondary objectives included the ability to estimate quarterly earnings over the entire reference period and supporting a broader set of exploratory analyses of career trajectories (transitions between school, work, and other activities).

For this imputation, we used a different approach from any of those discussed above. This decision was motivated by the complexity of partial information in the Training and Employment History module. Across the pooled PACE/HPOG sample, respondents had as many as six school spells and as many as 11 job spells. Even when respondents could not remember dates, we had many bounding conditions (e.g., spell #4 started after spell #3 ended). We devised a method that would respect these bounding conditions to create a coherent history while also supporting high-quality estimates of the site-specific impact of treatment on duration of study and quarterly earnings.

Before explaining the method, it will be useful to understand bounding conditions.

- For every spell, we knew whether it ended before the three-year follow-up interview or was ongoing at that time.

- For all closed spells, we knew whether there was another spell that started after it but prior to the three-year interview.

- For most spells, we knew
    - whether it started before or after randomization;
    - whether it started in the middle of another spell or after some period during which the person was neither working for pay nor enrolled in school; and
    - whether a new spell started during it.

- For spells that followed other spells, we would most often know the end date of the prior spell.

- For spells that preceded other spells, we would most often know the start date of the succeeding spell.

- For spells that started during other spells, we would most often know the start and end dates of the "mother" spell.

- For spells that spanned the start of a new spell, we would most often know the start and end dates of the "daughter" spell.

Our general approach to imputing missing dates involved the following steps on the pooled PACE/HPOG sample.

1. Express the date as a lag to some benchmark date. Specifically, we expressed start dates of main spells (those that did not start in the middle of any other spell) as the lag between randomization and the start of the spell, start dates of daughter spells as the lag from the start of the mother spell to the start of the daughter spell, and end dates of all spells as the lag from spell start date to spell end date.

2. Construct a statistical model for lag, and extract the predicted lag for spells with both known and unknown dates. (More details on this modeling process follow below. We constructed nine separate models.)

3. Identify the nearest neighbor case in the pooled dataset in terms of the predicted lag. Copy the lag from the spell with the known relevant date (start or end) to the case with an unknown value for the relevant date.

4. Add the imputed lag onto the benchmark date for the spell with an unknown date to obtain a preliminary date.

5. If the preliminary imputation violates any of the constraints, truncate it to just barely satisfy the constraints. For example, if preliminary imputation of an end date placed the end date past the date of follow-up interview but the respondent had reported that the spell ended before the interview, then we truncated the lag so that the job ended the month before the interview.

Before providing details on the nine models constructed in step 2, we offer some general observations about this methodology. We considered to conducting this process separately for each site. We rejected that approach because of the complexity of the boundary constraints on dates and the rarity of patterns for respondents with multiple spells. Instead, we focused on constructing high-quality models and then finding the best match available.

The pooled sample size consisted of 27,939 job spells plus 13,093 school spells. After discarding spells reported by skipouts and spells that ended prior to randomization, the total number of spells was 40,672. Among these spells, either the start date or the end date was missing for 3,302, or 8 percent. Missing start dates was the more common problem, with 538 spells missing just the end date and 2,764 missing just the start date or both dates. Missing dates were slightly more common for school spells than for job spells (10 percent versus 7 percent). Missing dates for closed spells were much more common than for open spells (10 percent versus 4 percent). For I-BEST, the overall missing data rate for spell dates was 8 percent, the same as on the pooled sample.

Exhibit C-7 below lists the models we created for each type of lag and some features of each, including average imputed values for the various lags. Main spell #1 was always the ongoing spell at the time of randomization for those respondents working or going to school at the point of randomization, and so always has a negative lag. Main spell #2 was always the first spell after randomization for those not working or going to school at the point of randomization. Other

main spells always followed main spell #1 or #2. Given this structure, we prepared separate models for the start date of each group (lag types 1, 5, and 6 below) and we modeled other features associated with the first spell separately, as well (lag types 2, 3, and 4). For other lag types, we modeled on a pooled dataset combining main spells #2 and higher (lag types 7, 8, and 9) and their associated subspells.

Exhibit C-7:   Date Imputation for Three-Year Impact Study (Pooled PACE/HPOG Sample)

Lag Type	Modeled Variable	R-Squared (%)	Tested Variables (#)	Selected Variables (#)	Sample Size (#)	Missing Data Rate (%)	Average Lag/Duration	
							Reported (months)	Imputed (months)
1	Lag from randomization date to start of main spell #1 (always negative because spell #1 was activity at time of randomization)	15	1,071	18	8,994	9.7	−18.8	−18.6
2	Duration of main spell #1 (closed only)	79	3,625	3	7,377	7.3	25.9	28.0
3	Lag from start of main spell #1 to start of subspell	78	2,989	3	5,459	8.8	23.2	16.9
4	Duration of subspells of main spell #1 (closed only)	0	3,103	2	4,563	8.8	16.2	15.7
5	Lag from randomization date to start of main spell #2	7	1,089	2	3,863	7.0	6.7	6.7
6	Lag from randomization date to start of main spells #3 and higher	38	5,113	33	18,082	4.9	18.9	17.4
7	Duration of main spells #2 and higher (closed only)	16	4,760	23	13,509	5.4	8.3	8.3
8	Lag from start of main spell #2 and higher to start of subspell	43	4,105	11	4,270	6.3	6.0	4.2
9	Duration of subspells for main spells #2 and higher (closed only)	14	3,383	9	2,546	6.8	7.3	7.1

Source: National Directory of New Hires; National Student Clearinghouse; PACE and HPOG 1.0 three-year follow-up survey.

Note: Sample pooled across HPOG 1.0 and all nine PACE sites. Sample also pooled across treatment and control samples. A "main spell" is a spell that did not start in the middle of another spell. A "subspell" is a spell that did start in the middle of another spell.

The set of variables allowed into each model varied across the nine steps. Tested variables included program, randomized treatment group, the interaction of program with treatment group, elapsed time between randomization and follow-up interview (and its square), job/school status, next activity (work, school, or other), school control (three levels, nested within job/school

status), school level (three levels, nested within job/school status), open/closed status, life trajectory cluster (five levels), self-assessed goal progress, baseline covariates, two- and three-way interactions of these variables with program and treatment status, and other variables.

Model fit as measured by *R*-squared varied substantially across models, ranging from 0 percent to 79 percent. The reasons for this variation are not clear to us. Average imputed values were generally quite similar to average reported months.

Undercoverage of NSC-Reported Spells

As noted previously, we decided to supplement the histories of survey respondents who reported no training since randomization with any spells recorded for them in the NSC and then to impute the spell attributes collected in the survey beyond the simple start and end dates for the spells. Across the nine PACE sites, this edit changed the training history for 7 percent of the sample, switching them from a status of no training to some. In the I-BEST sample, there were 36 such respondents, accounting for 9 percent of the sample. We added these NSC-reported spells to the three-year follow-up survey history for those respondents and imputed the missing survey outcomes, such as earned credits and credentials.

This imputation proceeded by matching these 36 respondents to other I-BEST study participants and copying over the donors' outcomes. This matching was structured, not random. We constrained matches to be from the same treatment group and to have a similar predicted profile of four survey-reported spell-level variables:

- Received a diploma or certificate typically requiring less than a full year's worth of study during the spell.

- Received a diploma or certificate typically requiring a year or more worth of study, but less than an associate degree during the spell.

- Received an associate degree or higher during the spell.

- Total credits earned during the spell.

We formed linear models for each of these survey-reported spell-level outcomes in terms of baseline variables and NSC-reported spell- and person-level variables on enrollment and credential attainment. We fit these models on the pooled (treatment plus control) sample for the I-BEST program. Given that the matching was not random, we did not conduct multiple imputation. We instead conducted single imputation and have ignored the impact on variances.

C.3   Survey Nonresponse Analysis

As in any survey, nonresponse can lead to bias if nonresponse propensity is correlated with outcomes. In the context of a randomized experiment such as this evaluation of I-BEST, concern about nonresponse is heightened if the nonresponse rate is different in the treatment group than in the control group. Nonresponse can lead to biased impact estimates even without

differential nonresponse rates across study groups, but it is widely accepted that differential rates heighten concerns about biased impact estimates.[40]

The three-year follow-up survey for this PACE site obtained disparate response rates in the treatment (70 percent) and control (64 percent) groups. Such a difference suggests that there may be material differences in baseline characteristics and outcomes for respondents versus the full sample. We studied this matter further using administrative data and found weak evidence of nonresponse bias. (Illustrations of these biases are presented in Exhibit C-9 below). We developed a set of nonresponse adjustment weights that appears to remove most of this bias. This section first presents the evidence of nonresponse bias in unadjusted impact estimates and then documents the nonresponse adjustment weights that we created to mitigate this bias.

C.3.1 Evidence of Nonresponse Bias in Unadjusted Impact Estimates

We gauged the likelihood of nonresponse bias through two types of analysis, one involving baseline data and one involving post-randomization administrative data.

The first analysis takes baseline equivalence as an indication of the potential for bias. If randomization is correctly implemented, there should be no systematic differences between the treatment group and the control group. We directly tested that using complete data from the BIF (see Appendix Section A.2). This insight also provides a proxy for nonresponse bias and the ability of our weighting scheme to correct for it. In the absence of nonresponse bias, appropriately weighted tabulations of the BIF *among survey respondents* should also show baseline equivalence.

The second type of analysis looks directly at estimated impacts. We know who responded to the survey and we have administrative data outcomes for both survey respondents and nonrespondents. We can thus compute two impact estimates from the administrative data: one estimate from the unweighted full sample, which we treat as truth; and a second estimate from the weighted survey sample. In the absence of nonresponse bias (and with large enough samples), we should get the same (up to sampling variability) estimates of impact on the full sample and on the weighted sample of survey respondents. Theoretically, it is possible to test whether estimated differences between these two impact estimates are statistically significant, but we did not do this, relying instead on impressions of consistency across a collection of administratively measured outcomes.

Exhibit C-8 below considers baseline equivalence among survey respondents. In the first three columns reflecting all participants, there are two characteristics where we see statistically significant differences between the treatment and control groups.[41] The next three columns, which report statistics for survey respondents, do not show statistically significant differences for any characteristics. The last column, which reweights the survey respondents, has three

---

[40]   See for example, Deke and Chiang (2017). For a slightly contrarian view, see Hendra and Hill (2018).

[41]   Note that the numbers in the first three columns of Exhibit C-8 reflect baseline balance for the full sample following imputation, whereas Appendix A.2 presented pre-imputation numbers.

Exhibit C-8:   Baseline Balance on Full Sample, Unweighted Respondent Sample, and Weighted Respondent Sample

Characteristics	Treatment (Full Sample)	Control	p-Value	Treatment (Unweighted Sample)	Control	p-Value	Treatment (Weighted Sample)	Control	p-Value
Age (%)			.067			.125			.052
20 or under	23.2	21.2		22.0	18.4		23.7	19.3	
21-24	11.1	18.7		9.6	17.4		8.4	17.8	
25-34	31.4	28.2		30.3	29.4		30.4	29.3	
35+	34.3	32.0		38.1	34.8		37.5	33.6	
Gender (%)			.245			.411			.337
Female	55.2	59.8		58.3	62.2		56.4	61.3	
Male	44.8	40.2		41.7	37.8		43.6	38.7	
Race/Ethnicity			.465			.315			.150
Hispanic, any race	29.2	23.7		29.8	21.9		29.9	21.4	
Black, non-Hispanic	6.7	8.9		6.4	9.5		6.4	11.1	
White, non-Hispanic	54.3	56.3		51.4	54.2		53.8	53.9	
Another race, non-Hispanic	13.0	14.6		15.6	16.4		12.9	15.7	
Family Structure (%)			.640			.689			.436
Not living with spouse/partner and not living with children	49.2	46.2		43.6	43.8		47.6	44.7	
Not living with spouse/partner but living with children	14.6	17.7		15.6	16.9		14.2	17.4	
Living with spouse/partner and not living with children	18.1	16.5		20.6	16.4		18.7	14.4	
Living with spouse/partner and children	18.1	19.6		20.2	22.9		19.5	23.5	
Living with parents (%)	27.9	31.0	.398	27.1	30.4	.459	28.4	31.6	.479
One parent has at least some college (%)	47.3	43.0	.283	47.3	41.3	.221	49.5	40.3	.059
High School Grades (%)			.359			.172			.304
Mostly A's	7.6	6.3		7.8	8.0		6.6	6.1	
Mostly B's	33.3	29.1		35.8	27.4		36.4	29.6	
Mostly C's or below	59.1	64.6		56.4	64.7		57.0	64.3	

Characteristics	Treatment (Full Sample)	Control	p-Value	Treatment (Unweighted Sample)	Control	p-Value	Treatment (Weighted Sample)	Control	p-Value
Current Education (%)			.458			.945			.871
Less than a high school diploma	29.5	33.2		28.4	30.9		28.6	32.8	
High school diploma or equivalent	40.6	38.0		40.4	37.8		41.0	38.6	
Less than one year of college									
One or more years of college	10.8	8.5		11.0	10.5		10.8	9.3	
Associate degree or higher	7.3	9.8		8.3	9.5		8.1	9.1	
Received vocational or technical certificate or diploma (%)	19.4	19.0	.904	20.2	22.4	.583	21.3	19.1	.561
Career Knowledge Index (average of items)	0.41	0.41	.782	0.43	0.41	.592	0.43	0.41	.619
Psycho-Social Indices									
Academic Discipline Index	5.06	5.08	.651	5.09	5.14	.421	5.05	5.11	.343
Training Commitment Index	5.42	5.43	.865	5.43	5.41	.681	5.41	5.39	.730
Academic Self-Confidence Index	4.49	4.47	.751	4.46	4.49	.677	4.43	4.44	.852
Emotional Stability Index	4.96	4.94	.715	4.98	4.96	.823	4.95	4.93	.761
Social Support Index	3.22	3.20	.610	3.21	3.22	.845	3.20	3.21	.847
Stress Index	2.30	2.32	.702	2.31	2.28	.691	2.32	2.32	.964
Depression Index	1.61	1.59	.663	1.60	1.60	.975	1.61	1.63	.776
Income (%)			.182			.406			.328
Less than $15,000	45.4	47.8		44.5	49.3		45.6	51.1	
$15,000-29,999	26.7	20.9		27.5	21.4		27.5	21.2	
$30,000+	26.7	31.7		26.6	29.4		25.9	27.7	
Mean ($)	22,711	22,415	.869	22,292	21,673	.770	21,779	20,948	.693
Public Assistance / Hardship Past 12 Months (%)									
Received WIC or SNAP	55.6	61.7	.117	54.6	61.2	.172	55.2	63.5	.085
Received public assistance or welfare	17.5	23.1	.078	17.9	22.4	.252	18.4	24.5	.125
Reported financial hardship	50.2	46.8	.404	51.8	46.8	.301	52.5	48.8	.451

Characteristics	Treatment (Full Sample)	Control	p-Value	Treatment (Unweighted Sample)	Control	p-Value	Treatment (Weighted Sample)	Control	p-Value
Current Work Hours (%)			.927			.541			.539
0	67.9	65.8		68.8	62.2		69.9	66.4	
1-19	7.9	9.2		7.3	9.5		6.9	8.8	
20-34	11.8	11.1		11.0	10.5		11.5	8.9	
35+	13.0	13.9		13.3	16.9		11.6	14.6	
Expected Work Hours in Next Few Months (%)			.232			.212			.333
0	40.3	40.5		43.1	42.8		44.8	43.5	
1-19	9.2	11.7		8.3	11.0		7.7	10.4	
20-34	35.2	28.8		33.0	25.4		33.2	27.1	
35+	15.2	19.0		15.6	20.9		14.3	19.1	
Life Challenges Index (average in original units 1-5)	1.55	1.54	.811	1.57	1.55	.693	1.59	1.56	.488
Owns a car (%)	61.9	63.6	.659	64.7	65.2	.916	62.8	63.5	.886
Has both computer and internet at home (%)	69.8	74.1	.240	70.2	74.1	.370	68.3	73.2	.273
Ever arrested (%)	29.2	29.4	.951	25.7	27.4	.699	26.6	27.2	.894
Sample sizes:	315	316		218	201		218	201	

Source: PACE Basic Information Form; PACE Self-Administered Questionnaire; response status to the PACE three-year follow-up survey.

Note: SAS/SURVEYFREQ used to test for significant imbalances for categorical variables. SAS/TTEST used to test for significant imbalances for other variables. Weights are based on the dual raking system explained in Appendix Section C.3.2 below. Significant imbalances are highlighted in red, using a threshold for statistical significance of 10 percent.

statistically significant imbalances. As a result, baseline imbalances may make our sample prone to nonresponse bias. We will discuss how we correct for this in the next section. Furthermore, comparisons of impacts on administrative follow-up outcomes with and without survey weighting (shown in Exhibit C-9) show evidence of nonresponse bias, as well.

Exhibit C-9 presents evidence about the level of nonresponse bias with and without adjustment weights. The first two panels of Exhibit C-9 compare three sets of regression-adjusted impacts on earnings outcomes from National Director of New Hires (NDNH) records. The first set of impact estimates (column 1) is based on the full sample. The second set of impact estimates (column 3) excludes survey nonrespondents. Differences between the first and second set of impacts signal nonresponse bias. The third set of impact estimates (column 5) also excludes survey nonrespondents but weights survey respondents with nonresponse adjustment weights, which are explained in Section C.3.2 below. If the weights are good, then the differences between the first and fifth columns will be smaller than those between the first and third columns. Note that all three sets of impact estimates are regression-adjusted with the covariates discussed in Appendix Section A.3.

While we did not formally test the differences between the alternative estimates, given that the survey respondents constitute a very large subset of all participants, many of the differences would be statistically significant. For several follow-up administrative variables, there are troubling signs of bias. In particular, the impact of I-BEST on the confirmatory education outcome (receiving a credential requiring one or more years of college study) is not significant on the full sample but is significant on the respondent sample, roughly doubling from an impact of 2.4 percentage points to an impact of 5.1 percentage points, significant at the 5 percent level. However, the difference between these alternate estimates is itself not statistically significant. It appears that treatment group members who did not receive these credentials may have been a little less likely to respond to the survey than those who did receive them, whereas the opposite was true in the control group. Although one could build a theory about how educational success makes people more or less willing to respond, it does not make sense that this relation would be in the opposite directions for the two samples. Most likely, this is just noise arising from the small sample sizes.

Estimated impacts on earnings at Q9 also show signs of bias on the respondent sample—the estimated impact on earnings in Q9 is substantial and positive (+$805), whereas for the full sample the estimated impact is smaller and closer to zero (+$299). The estimated impact on having any earnings at Q9 is also much larger on the respondent sample, doubling from 6.2 percentage points on the full sample to 12.8 percentage points on the respondent sample. Again, the standard errors are large enough to be consistent with noise from small sample sizes.

Exhibit C-9:   Comparison of Selected Estimates of the Impact of I-BEST for the Unweighted and Weighted Survey Samples

Outcome (Data Source)	Impact (Full Sample)	Standard Error	Impact (Unweighted Sample)	Standard Error	Impact (Weighted Sample)	Standard Error
Confirmatory Outcome (NDNH)						
Average Q12-Q13 earnings ($)	404	344	401	425	205	411
Exploratory Outcomes (NDNH)						
Q5 earnings ($)	256	269	514*	331	415	330
Q9 earnings ($)	299	356	805**	439	617*	414
Q13 earnings ($)	348	365	345	447	133	439
Q17 earnings ($)	−346	403	−443	512	−517	499
Any earnings Q5 (%)	5.3*	3.8	6.1*	4.6	4.3	4.8
Any earnings Q9 (%)	6.2*	3.9	12.8***	4.6	11.1**	4.8
Any earnings Q13 (%)	2.7	3.8	2.4	4.6	0.9	4.8
Any earnings Q17 (%)	2.6	3.9	1.0	4.7	−1.4	4.8
Secondary Employment Outcomes (Survey)						
Employed at survey follow-up (%)			7.0*	4.8	6.7*	5.1
Employed at $14 per hour or above (%)			5.6*	4.3	7.5**	4.5
Employed in job requiring mid-level skills (%)			−4.0	3.2	−3.4	3.0
Confirmatory Education Outcome (SBCTC Records)						
Received credential taking 1+ year of college study	2.4	2.3	5.1**	2.9	5.9**	3.1
Secondary Education Outcomes (SBCTC Records)						
Number of workforce and academic credits	10.9***	2.6	12.2***	3.2	12.8***	3.5
FTE months enrolled at colleges (months)	2.4***	0.5	2.5***	0.66	2.6***	0.7
Receipt of any credential from a college (%)	31.0***	3.5	37.1***	4.19	33.5***	4.5
Secondary Education Outcomes (Survey)						
Receipt of an exam-based certification or license (%)[a]			10.2**	4.7	10.6**	5.0
Other Secondary Outcomes (Survey)						
Indicators of Independence and Well-Being						
Health insurance coverage (%)			−1.2	3.5	−0.8	3.8
Receives public benefits (%)			2.2	4.5	3.8	4.6
Personal student debt ($)			380	586	344	609
Any signs of financial distress (%)			−5.4	4.8	−5.2	4.9
Indices of Self-Assessed Career Progress (average)						
Confidence in career knowledge[b]			0.05	0.06	0.07	0.07
Access to career supports[c]			−0.01	0.03	−0.01	0.03
Sample size (treatment + control group):	SBCTC 631 NDNH 610		419		419	

Source: PACE 18-month survey; PACE three-year follow-up survey; SBCTC Records, National Directory of New Hires.

[a] Blended 18-month and three-year survey results.

[b] Seven-item scale tapping self-assessed career knowledge; response categories range from 1=strongly disagree to 4=strongly agree.

[c] Six-item scale tapping self-assessed access to career supports; response categories range from 1=no to 2=yes.

Statistical significance levels based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt): *** 1 percent level; ** 5 percent level; * 10 percent level.

Prior to viewing these estimates, the research team did find significant nonresponse biases in the impact of another PACE program on both earnings and educational progress. Given the centrality of earnings in the logic models for how PACE programs would affect a wide variety of life outcomes measured in the three-year follow-up survey, this relationship clearly implies some survey nonresponse adjustment was required for the other PACE program. We then applied the nonresponse adjustment at all sites out of an abundance of caution. (Section C.3.2 gives the details of how we created these nonresponse adjustment weights.) However, the final pair of columns in Exhibit 3-9 shows that this adjustment did little good for I-BEST.

In some cases, the nonresponse weights bring impact estimates based only on survey respondents back into better alignment with impact estimates on the full sample. For example, the impact on Q9 earnings for the full sample is +$299. The estimate impact for the weighted survey sample is +$617, which is much closer to the full sample estimate than the unweighted estimate of +$805 is. Though weighting reduced nonresponse bias for Q9 earnings, it failed to do so for six of the nine NDNH outcomes in Exhibit 3-9. Furthermore, the weighting increased nonresponse bias for three of the four education outcomes based on SBCTC records, including the confirmatory outcome.

Luckily, despite the marginal utility of the nonresponse weights for I-BEST, the use of weights appears to have little impact on estimated impacts of I-BEST on secondary outcomes based on the follow-up survey. For these survey-based outcomes, the third and sixth panels of Exhibit C-9 compares the unweighted and weighted impact estimates. There are only minor differences between the estimates.

C.3.2 Construction of Nonresponse Adjustment Weights

Construction of weights to reduce the biases discussed above was more complex than anticipated. At first, we tried a standard propensity scoring approach,[42] as was used in the short-term report on I-BEST (Glosser et al. 2018). However, that approach was not successful in removing the biases in estimated impacts based on administrative data for survey respondents at that other PACE site. Data storage arrangements posed a further challenge in developing a set of nonresponse adjustment weights. Contractual arrangements permitted the merging of survey data with either NDNH data or NSC data, but they did not permit the merging of NDNH and NSC data. In response to this challenge, we developed a new approach that we call dual-system raking.

"Raking" is the name for iterative procedures that create weights for a sample in such a manner that marginal tabulations of the sample agree exactly with pre-specified "control" totals in multiple dimensions. For example, raking can be used to create weights that will cause

---

[42] In the standard approach, a logistic model for response status is fit in terms of universally available covariates (baseline and administrative). The model is used to generate a predicted response propensity for each person (respondent and nonrespondent), then people are sorted on this prediction into strata. The empirical response rate is calculated for each stratum, and finally the inverse of this rate is applied to respondents as a nonresponse-adjustment weight.

tabulations by gender, tabulations by race, and tabulations by age all to agree with pre-specified totals for gender, race, and age. In this example, gender, race, and age are dimensions.

In the context of nonresponse, if tabulations are prepared from the full sample and raking is used on the respondents, then weighted tabulations of the respondent sample will be in perfect agreement with parallel tabulations of the full sample. This exact multi-dimensional agreement is referred to as "hyperbalance." In the context of an experiment, if this procedure is run separately for the treatment and control groups, then hyperbalance between respondents and nonrespondents means that the weighted balance between the treatment and control groups on the respondent sample should be just as good as on the full sample.

This hyperbalance by arm means that if we estimated treatment impact on just the respondent sample with these weights but without regression adjustment, the estimated program impact on each of these hyperbalanced variables would agree exactly with corresponding program impacts estimated on the full sample. The use of regression adjustment to estimate program impacts (rather than simple mean difference between arms) means that this agreement will not be exact, but agreement should still be very good for hyperbalanced variables. Theoretically, it should also improve agreement (between impact estimates based on the full sample and impact estimates based on just the respondent sample) for a variety of related parallel outcomes.

Key raking variables include both categorical variables (e.g., any NSC-reported enrollment) and interval-valued variables (e.g., number of months enrolled in college according to NSC records). Including these interval-valued variables seems particularly important because many educational outcomes are associated with the length of study.

The need to include continuous variables in the raking is challenging because traditional raking algorithms work only with categorical variables. In contrast, the generalized raking we propose and use here can handle a mix of categorical and continuous variables.[43] For categorical variables, the procedure guarantees perfect correspondence between the respondent sample and full sample by arm on the distribution of the sample across the categories of each variable; for continuous variables, the procedure induces perfect agreement on the marginal means of each of them.

The generalized raking procedure of Folsom and associates is available in the WTADJUST procedure of SUDAAN. A similar procedure that only works for categorical covariates is the SAS raking macro of Izrael, Hoaglin, and Battaglia (2000). It was necessary to use both software packages because the analyses had to be run on two servers, one that had SUDAAN installed (at Abt) and one that did not (at ACF). We refer to our system as dual-system raking because it permits raking both to NDNH information and to NSC information though the two types of data reside on two different systems.

---

[43] Generalized raking is most fully developed by Folsom and Singh (2000), who in turn draw on work originally proposed by Folsom (1991), Deville and Särndal (1992), and Folsom and Witt (1994). Dual raking is similar to the approach of Judkins et al. (2007) that involves the use of raking to construct weights in quasi-experimental designs.

The details of the dual-system raking procedure are as follows:

1. We used SUDAAN/WTADJUST to develop survey weights on the Abt server that induced hyperbalance by arm for the means of four NSC variables. Two of these NSC variables were counts on months: months with any enrollment and months of full-time-equivalent enrollment. Two of the NSC variables were binary flags: any enrollment and any completions (credentials). All four of these variables were constrained to enrollment and completions within 35 months of randomization.

2. We merged the weights from step 1 with baseline data and follow-up survey data on the Abt server. We then passed these merged data through to a secure ACF server, where third-party ACF contractors merged our data with NDNH earnings data, removing personal identifiers from the merged dataset. We had verified that this set of NSC-adjusted weights provides nearly unbiased impact estimates for survey-based education outcomes, but after merging the weights with NDNH data, we discovered that these NSC-adjusted weights did not remove bias in survey-based impact estimates for earnings outcomes.

3. To remedy this, we used the Izrael-Hoaglin-Battaglia macro on the ACF server to rake the weights from step 1 in such a manner as to attain hyperbalance by arm on three categorized versions of NDNH earnings. Specifically, we obtained hyperbalance for a six-level categorization of earnings at Q12 and Q13, a five-level categorization of earnings at Q9, and a five-level categorization of cumulative earnings from Q1 through Q12.[44] We verified that these weights removed most of the nonresponse bias on estimates of program impacts on NDNH earnings at the other PACE site when estimated from nonrespondents instead of from the full sample. This sensitivity analysis included the continuous versions of the variables used in the raking, as well as continuous earnings at Q5 and Q17 and binary indicators for any employment at Q5, Q9, Q13, and Q17.

4. We used the weights from step 3 on the ACF server to estimate (by arm) the distributions of survey-reported earnings. Specifically, we split Q12 earnings at $0, $6,000, and $9,000; Q9 earnings at $0, $6,000, and $9,000; and average quarterly earnings for Q1 through Q12 at $3,000 and $6,000. (The breaks for survey-reported earnings needed to be coarser than the breaks for NDNH earnings because of the smaller sample sizes in the respondent survey sample.)

5. We again used the Izrael-Hoaglin-Battaglia macro on the ACF server to rake the weights from step 1, but for this step we used the control totals from step 4 rather than the NDNH totals used in step 3. We then verified that these weights removed most of the nonresponse bias on estimates of program impacts on NDNH earnings when estimated from nonrespondents instead of from the full sample at the other PACE site. These

---

44   This process is also referred to as "binning." We used more bins for the confirmatory outcome than for the exploratory outcomes. Reducing the number of bins generally speeds convergence and reduces the frequency of extreme adjustments.

weights did not perform as well as the weights from step 3 in reducing nonresponse bias on the respondent sample, but the deterioration (not shown) was not very large.

6. We exported the 11 estimated totals from step 4 for each arm from the ACF server to the Abt server. (The data use agreement permitted the transfer of tabulations; only the export of microdata was prohibited.)

7. We again used the Izrael-Hoaglin-Battaglia macro to rake the weights from step 1 to the control totals from step 4, but this time we did the raking on the Abt server rather than on the ACF server. We then merged these with NSC data on the Abt server and verified that these weights removed most of the nonresponse bias on estimates of program impacts on NSC outcomes when estimated from nonrespondents instead of from the full sample at the other PACE site.

C.4 Quality and Completeness of Exam-Based Credentials Reported in the Survey

Earlier analyses for another PACE site identified a potential quality issue for reports on receipt of exam-based credentials in the three-year follow-up survey. Specifically, estimates of exam-based certifications and licenses for the San Diego Workforce Partnership's Bridge to Employment in the Healthcare Industry program were much lower than those based on the short-term survey at 18 months after randomization (Farrell and Martinson 2017). This points to a clear problem, since the percent who ever received these credentials cannot diminish over time.

A review of the survey's skip patterns and wording identified three features in the design of the three-year instrument that might have led to fewer credentials of this type being reported than were in the 18-month survey:

- First, the three-year instrument allowed only respondents with some formal schooling since randomization to report exam-based certifications and licenses. However, people who learn skills on the job or through independent online study (such as YouTube tutorials) can sit for the exams for many certifications and licenses.

- Second, the wording for the three-year instrument strongly emphasized that "school-issued certificates" were not the same thing as "exam-based certifications and licenses." We had introduced this language to ease confusion about the difference between credentials issued by schools and credentials issued by other authorities. However, because some schools serve as proxy administrators of exams for credentials that are actually issued by other authorities, it is possible that this wording led some people to report exam-based credentials as school-based credentials or to not report them at all.

- The third feature is just the greater passage of time. Respondents may not have renewed exam-based certifications and licenses or they might have discovered that the credentials are less useful than anticipated, either of which could have reduced respondents' inclination to report older exam-based credentials.

Given this review, we decided that the short-term follow-up survey reports of early exam-based credentials earned are probably more accurate than the reports from the three-year survey. Accordingly, we decided to combine reporting for the two time periods. The composite measure of receipt of any exam-based credential since randomization was set to yes if the respondent either reported it in the 18-month survey or reported receiving such a credential in the three-year survey at a time point after the date of the 18-month survey interview. For the 15 percent of the sample who did not respond at 18 months, we imputed a response. When receipt dates were not reported in the three-year survey, we also imputed them. Both of these imputations are discussed above in Section C.3.

C.5   Quality and Completeness of School-Issued Credentials Reported in the Survey

As the discovery of problems with reporting of exam-based credentials just discussed in C.4 raised the question of whether similar problems occurred for school-issued credentials that would justify also blending reports on these credentials from the two surveys together. Results from analyses for another PACE site, Pima Community College (PCC)'s Pathways to Healthcare program, argued against the latter (Judkins, Gardiner, and Litwok 2020). The PCC study offered college records to support the analysis, making it a good choice for investigating these survey outcomes.

For the Bridge to Employment report, we decided to use the three-year survey without blending with the 18-month survey for other types of credentials, and decided the same for all other PACE reports in which survey data are used as well.[45] This decision we based on analyses of data for yet another PACE site: Pathways to Healthcare. We chose this site for the research because we had Pima Community College (PCC) records and because the evaluation's processing of those records was further along (at the time of drafting the Bridge to Employment report) than was processing at other PACE sites for which we had negotiated access to college records.

Analysis of PCC records showed that the three-year survey was more accurate than the 18-month survey. We focused on Pathways to Healthcare respondents who reported a school-issued credential in only one of the two surveys, and then checked to see whether the PCC records confirmed issuance of that survey-reported credential. Among respondents who reported such a credential at 18 months but not at three years, PCC records confirmed this claim for just 35 percent. In contrast, among respondents who reported such a credential at three years but not at 18 months, PCC records confirmed this claim for fully 81 percent.

For some reason, the 18-month survey instrument seems to have generated many more unverifiable school-based credential claims than the three-year survey did. For this reason, we decided to rely on the three-year survey without blending for survey-based measures of school-issued credentials in all PACE sites.

---

[45]   This has no relevance to measures of educational progress based on college records, as is the mostly the case in this report.

Appendix D: National Student Clearinghouse Data

The National Student Clearinghouse (NSC) is a national database of college enrollment records designed to aid the administration of student loan programs, but it can be a useful tool for education researchers. In this report, we used NSC records for imputation of missing data and to prepare alternate estimates of the impacts of the Integrated Basic Education and Skills Training (I-BEST) program, as discussed in Appendix B. Section D.1 summarizes statistics on NSC coverage. Section D.2 provides details on how raw data from the NSC were recoded to make them more relevant to the evaluation of I-BEST. Finally, Section D.3 presents estimates of I-BEST impacts based on NSC data and contrasts them with the estimates presented in Chapter 3 of this report.

D.1   Coverage

Given the focus on loan administration, NSC only covers schools that are Title IV schools, the set of schools approved for federal student loans by the U.S. Department of Education. Moreover, although the NSC does include a few schools that are not colleges in the sense used elsewhere in this report (i.e., issuing degrees), the vast majority of the schools are colleges. Exhibit D-1 shows the percentage of colleges providing records to the NSC by year and by type of school. As shown, coverage of public two-year and four-year schools was more than 95 percent. Coverage was lower among private not-for-profit four-year schools, considerably lower among private for-profit four-year schools, and very low for private two-year schools (both for-profit and not-for-profit).

Exhibit D-1:   NSC College-Level Cooperation Rates by College Control and Level from 2013 through 2016

Type and Control of College	2013 (%)	2014 (%)	2015 (%)	2016 (%)
Public, four-year	99.2	99.4	99.5	99.6
Private, not-for-profit, four-year	93.6	95.2	95.8	96.1
Private, for-profit, four-year	74.4	79.9	81.7	81.0
Public, two-year	99.1	99.2	99.4	99.5
Private, not-for-profit, two-year	39.5	40.8	40.4	42.1
Private, for-profit, two-year	19.7	28.1	26.7	26.6

Source: National Student Clearinghouse, https://nscresearchcenter.org/wp-content/uploads/NSC_COVERAGE.xlsx.

Analyses of NSC data in this report are limited to enrollment records obtained from 2000 forward. All study participants gave their informed consent to have NSC share their records with the PACE research team. The team negotiated a contract with the NSC to match relevant NSC records to the study participants. The team sent both Social Security numbers and names to NSC to make the matching more accurate. The abstracted records were then sent by encrypted secure methods to the research team, who have used them under tight security conditions.

D.2 Data and Measures

Information on outcomes other than enrollment tends to be less reliable.[46] Notably, standards and practices governing credential reporting are inconsistent across schools. So our primary use of NSC data was to measure enrollment. Counting the quarter during which random assignment occurred as Quarter 0, we obtained an abstract from the NSC in October of 2018 covering enrollment through Quarter 16 for all 631 study participants (315 in the treatment group and 316 in the control group).

Records from the NSC are arranged in a spell format with starting and ending dates. We translated these first into a set of person-month-level records, reconciling multiple and conflicting spells as seemed most sensible. The team derived two variables for each person-month. The first was a simple binary indicator of "any enrollment." The second was a measure of full-time-equivalent (FTE) enrollment that took the values 1 (for full-time enrollment), 0.75 for three-quarter-time enrollment, 0.5 for half-time enrollment, 0.25 for some but less than half-time enrollment, and 0 for no enrollment.[47] To translate these to person-quarter-level outcomes, a student was counted as enrolled for the quarter if he or she was enrolled in any of the three months of that quarter, and FTE enrollment was calculated by summing the student's total FTE months for the quarter.

D.3 Program Impacts on NSC-Measured Outcomes

Exhibit D-2 compares a selection of estimated impacts of I-BEST using both NSC records and adjusted State Board for Community and Technical Colleges (SBCTC) records.[48] We included this table as a check on the impacts estimated in the main body of the report using college records. The use of college records allowed us to estimate impacts on variables not measurable with the NSC data (such as the number of earned credits and receipt of particular types of credentials), but it also required the use of imputation for experiences at colleges other than those under the SBCTC.

The pattern of effects of I-BEST based on the two records systems is broadly consistent, except for enrollment in the fourth quarter after random assignment and receipt of any college-issued credential. For enrollment, both data sources show substantial positive impacts in Q4, so the difference is not very important practically. On the other hand, the difference in impacts on receipt of a college-issued credential is substantially larger (nearly 15 percentage points). The

---

[46] Dundar and Shapiro (2016) indicate that schools that choose to submit information on type of credential pursued or earned do so voluntarily and with minimal processing by NSC staff. About 90 percent of students attend schools that do submit information on credential types, but there is no systematic classification scheme for credentials that are not degrees. Schools merely submit names of certificates and diplomas awarded. The authors also specifically note that information on earned credits is weak. In addition, Dynarski, Hemelt, and Hyman (2015) report that only about 80 percent of degrees from Michigan colleges were reported to the NSC in the 2008-2010 period.

[47] Because informed consent had been collected from all study participants, NSC shared full-/part-time status for everyone in the sample, something that is not otherwise shared with researchers.

[48] Refer to Section B.2 of Appendix B for details on the adjustment of SBCTC records.

source of this difference is unclear. It could be due to coordination issues between the SBCTC and the NSC. The SBCTC created several new credentials in preparation for the I-BEST program; any issues in communicating receipt of the new credentials would affect the recorded credential achievement among the treatment group, dampening the NSC-reported treatment impact.

Exhibit D-2:   Comparisons of Impacts of I-BEST Based on Adjusted SBCTC Records with Impacts Based on NSC Records

Outcome	NSC Records				Adjusted SBCTC Records				Difference in Impacts	Standard Error
	Treatment Group	Control Group	Impact (Difference)	Standard Error	Treatment Group	Control Group	Impact (Difference)	Standard Error		
Any College Enrollment (%)										
In Q4	33.8	20.6	+13.2***	3.5	37.0	27.2	+9.8***	3.7	+3.5*	1.9
In Q8	18.8	15.2	+3.6	3.0	20.1	19.0	+1.1	3.2	+2.4	1.8
In Q12	11.7	9.8	+1.9	2.5	13.8	10.1	+3.7	2.6	−1.8	1.6
Cumulative Number of FTE Months of College Enrollment										
Through Q12	7.3	4.7	+2.6***	0.6	6.3	3.8	+2.4***	0.5	+0.2	0.2
Any Completions from a College (%)										
Through Q12	27.1	10.8	+16.3***	3.0	48.1	17.1	+31.0***	3.5	−14.7***	2.7
Sample size:	315	316			315	316				

Source: National Student Clearinghouse; adjusted SBCTC records.

Note: Statistical significance levels, based on two-tailed tests of differences between research groups: *** 1 percent level; ** 5 percent level; * 10 percent level.

Appendix E: Sensitivity Analyses of Education Impacts

The report used local Washington State Board for Community and Technical Colleges (SBCTC) records as the primary source for measures of confirmatory and secondary education outcomes. As a check on the sensitivity of impact estimates to this measurement choice, here we present an alternative set of estimates based on survey data. Across the confirmatory outcome and three secondary outcomes, the conclusions that arise from each data system are significantly different for all but one outcome.

With respect to the confirmatory outcome, the survey data show a significantly larger impact than do the SBCTC records. A deeper investigation found two potential explanations for this disparity, both suggesting that the survey data are problematic. First, SBCTC records show that the survey respondents saw a slightly larger impact (5.1 percent) than did the full sample (2.4 percent) for receipt of a long-term college credential. That difference, though not statistically significant, does provide weak evidence of nonresponse bias (which the nonresponse weights failed to substantively correct).

In addition, inspection of individual responses provides some evidence that survey respondents may be misclassifying their credentials. We observed 16 respondents (4 percent of the respondent sample) with a survey-reported credential requiring one or more years of college study but no such SBCTC-reported credential. The SBCTC data did show receipt of a degree requiring less than one year of college coursework for 13 of these 16 respondents. In many of these cases, the name of the credential in the SBCTC data was similar to the name of the credential reported in the survey, and the credential was clearly shorter than one year in duration (e.g., Certified Nursing Assistant). These results suggest that the difference between survey-reported and SBCTC-reported receipt of long-term credentials is partly due to mischaracterizing of credentials by survey respondents.

We also observe impact differences in the number of credits earned and receipt of a college credential of any duration. Among survey respondents, those in the treatment group reported spending more full-time-equivalent months in college and earning more credentials than respondents in the control group. However, there was no corresponding impact on receipt of college credits. One plausible explanation is respondents may have been confused about what type of SBCTC credits counted as "college" credits, and they may not have included both academic and vocational credits in their survey responses. With respect to credential receipt, even though the difference between the systems in impacts is statistically significant, both systems agree that the treatment had a strong positive effect.

Exhibit E-1: **Comparisons of Impacts of I-BEST Based on Adjusted SBCTC Records with Impacts Based on the Three-Year Follow-up Survey**

Outcome	Three-Year Follow-up Survey				Adjusted SBCTC Records				Difference in Impacts[a]	Standard Error
	Treatment Group	Control Group	Impact (Difference)	Standard Error	Treatment Group	Control Group	Impact (Difference)	Standard Error		
Confirmatory										
Received credential taking 1+ year of college study (%)	15.0	5.9	+9.1***	3.1	10.7	8.2	+2.4	2.3	+6.6**	2.9
Secondary										
Number of workforce and academic credits	17.6	17.9	−0.3	3.9	26.6	15.7	+10.9***	2.6	−11.2***	4.0
FTE months enrolled in college	8.9	6.5	+2.4***	1.0	6.3	3.8	+2.4***	0.5	−0.0	0.9
Received any college credential (%)	43.2	23.3	+19.9***	4.7	48.1	17.1	+31.0***	3.5	−11.2**	4.7
Sample size:	218	201			315	316				

Source: PACE three-year follow-up survey; adjusted SBCTC records.

Note: Statistical significance is based on one-tailed tests, unless otherwise noted.

[a] Statistical significance for difference in impacts is based on two-tailed test.

Statistical significance levels based on tests of differences between research groups: *** 1 percent level; ** 5 percent level; * 10 percent level.

Appendix F: NDNH's Unemployment Insurance Wage Data

Through the 1990s, many social program evaluations relied on administrative earnings data provided by state Unemployment Insurance (UI) agencies. State agencies maintained these data, and privacy concerns sometimes precluded sharing with outside researchers. UI records have become more accessible since 1996 with the advent of a centralized national database—the National Directory of New Hires (NDNH). Among the NDNH's virtues is that, unlike state data, it captures earnings for study participants who move to another state during the follow-up period.

The federal Office of Child Support Enforcement (OCSE) in the U.S. Department of Health and Human Services' Administration for Children and Families (ACF) operates the NDNH.[49] The NDNH contains new hire, quarterly wage, and UI information submitted by State Directories of New Hires, employers, and state workforce agencies. OCSE also supplements the state reports with records about earnings from federal civilian and military jobs (which are otherwise not covered by state UI data). Given this supplementation, the most important sources of uncaptured earnings are from self-employment, firms' employment of independent contractors, unreported tips, and informal employment.[50]

F.1 Data Collection Process

The primary purposes of the NDNH are to assist state child support agencies to locate noncustodial parents, putative fathers, and custodial parents to establish paternity and child support obligations and to enforce and modify orders for child support, custody, and visitation. It is also used by state UI agencies and the federal Social Security Administration to identify overpayments of benefits. However, subject to federal law, regulation, guidance, and other requirements to protect data privacy and security,[51] OCSE may disclose certain information contained in NDNH to requesting local, state, or federal agencies for research likely to contribute to achieving the purposes of part A or part D of title IV of the Social Security Act. Part A governs the federal Temporary Assistance for Needy Families (TANF) program. Part D governs the state/federal child support program. Such disclosures may not include the names, Social Security numbers (SSNs), or other personally identifying information.

If the disclosure is approved, the agency and OCSE must work together on the operational issues surrounding the technical and procedural aspects of the disclosure, such as mitigating

---

[49] More detail is available at: https://www.acf.hhs.gov/css/training-technical-assistance/guide-national-directory-new-hires.

[50] According to the U.S. Bureau of Labor Statistics, about 10 percent of workers are self-employed: https://www.bls.gov/spotlight/2016/self-employment-in-the-united-states/home.htm.

[51] The legal authority for this disclosure for research purposes is contained in subsection 453(j)(5) of the Social Security Act and Section 5507 of the Patient Protection and Affordable Care Act. For more information, see: https://www.govinfo.gov/app/details/USCODE-2010-title42/USCODE-2010-title42-chap7-subchapIV-partD-sec653.

the risks of identifiability and establishing appropriate data retention and disposition schedules of data files.

ACF's Office of Planning, Research, and Evaluation (OPRE) and OCSE negotiated a memorandum of understanding allowing access to NDNH data for the Pathways for Advancing Careers and Education (PACE) evaluation. Among other provisions, the memorandum dictates what self-reported data from study subjects may be merged with NDNH data, the computing environment where these merges are conducted, and procedures for review of tables prior to release.

The PACE research team transmits match request files to OCSE quarterly. These match request files contain the names and SSNs of PACE study participants. OCSE verifies with the Social Security Administration that the reported SSNs belong to the named persons. For those SSNs that pass this test, OCSE copies NDNH records for that quarter and the preceding seven quarters to a secure folder on the ACF server.[52] (Ordinarily, these records would be destroyed after two years.) These copied records contain a pseudo-SSN; the records are stripped of all personal identifiers.

States are required to submit earnings records to OCSE within four months, but there are stragglers and corrections. To be safe, PACE analyses limit NDNH-based measures to time periods that ended at least six months prior to the extract date.

Once we are ready to analyze the collected data, we submit a "passthrough" file to OCSE containing a variety of PACE-assigned variables (such as treatment status and program ID) and self-reported variables (such as the baseline information described in Appendix A). OCSE then strips the personal identifiers out of the passthrough file and replaces the actual SSNs with the same pseudo-SSNs previously assigned to the archived wage records. The study then uses these pseudo-SSNs to merge program and self-reported data with NDNH quarterly wage data on ACF's secure server to estimate program impacts on earnings and employment.

F.2   Data and Measures

Random assignment for the Integrated Basic Education and Skills Training (I-BEST) program started in November 2011 and ended in September 2014. Given the lag of up to six months in processing of employer reports by the states and transfer of state data to OCSE, wage records from the NDNH were available through Q4 2018; this means that we had 28 post-randomization quarters of earnings data for the earliest randomized study participants and 17 post-randomization quarters of earnings data for the last randomized study participants. In addition, we had eight quarters of pre-randomization data for the entire sample (we included the four most recent pre-randomization quarters in our regression-adjustment models).

Of the 631 treatment and control group members randomized as part of the I-BEST evaluation, 610 study participants reported a name and SSN that OCSE deemed to be of sufficient quality

---

[52]   Those study participants who are not matched in the Social Security Administration database are considered "missing" for these purposes, because their employment records are not available.

for its matching purposes.[53] Analyses in this three-year report thus are based on the 97 percent of the sample the agency deemed suitable. This sample's earnings in each quarter were based on earnings records found for each sample member in matching. As usual in use of such data, we defined sample members as "not working" when there was no match to wage records in a given quarter.

Each quarter, we submitted a match request file to OCSE that contained the names and SSNs for everyone randomized to that date. For those where the SSNs and names aligned, OCSE returned earnings data for the eight most recent quarters in the NDNH, which is lagged by two quarters from the date of the match. This meant that we had up to eight wage reports for each quarter. We used the last version for each quarter within a window. For example, for earnings in the second quarter of 2014, we used reports from the match file for the third quarter of 2016 and discarded the seven earlier sets of earnings data for the second quarter of 2014.

When the earnings data for a quarter contained two or more reports for the same person from the state, we assumed that these reports reflected either different payments by the same employer or payments from different employers. Consistent with the logic discussed in Appendix H, we reviewed quarterly earnings for any values that were clearly impossible, but failing to find any such values, did not discard or top-code any large earnings amounts.[54]

We calculated two outcomes for each quarter: a binary indicator of "any earnings" (yes/no) and the total reported wages for the quarter ($). The result was two series of 22 measures for each person (employment and earnings for the four quarters before randomization, the quarter of randomization, and the 17 quarters after randomization). In addition, we formed a quarterly average for Q12 and Q13 after random assignment (the confirmatory earnings outcome, established to align with the I-BEST theory of change) and an annual average for Q10-Q13.

---

[53] The acceptability of the combination of a name and an SSN can vary over time. OCSE reviews the SSN ownership every quarter for the entire sample.

[54] Top-coding means values above a threshold are set equal to the threshold.

Appendix G: Comparing NDNH- and Survey-Based Employment and Earnings Estimates

This appendix examines the consistency between earnings impacts estimated for the Integrated Basic Education and Skills Training (I-BEST) program based on National Directory of New Hires (NDNH) and survey data.

Barnow and Greenberg (2015) review findings from evaluations including as data sources both administrative data on earnings (usually state Unemployment Insurance (UI) wage data) and surveys. Although average survey-reported earnings tend to be higher than average total UI earnings, impact estimates still may be nearly unbiased (Kornfeld and Bloom 1999). Barnow and Greenberg (2019) update their earlier work in their overview essay for a special edition of *Evaluation Review* that includes seven articles considering the relative strength of records-based and survey-based impact estimates in randomized trials and other evaluations. They find considerable variability across studies but perhaps a general trend for estimated impacts based on surveys to be larger than estimated impacts based on administrative data with no clear evidence of which is more credible. The same pattern also appears in our I-BEST findings, as shown in Exhibit G-1.

Exhibit G-1:   Impacts of I-BEST on Earnings and Employment around Follow-up Q12 Based on Wage Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Quarterly Earnings				
Average NDNH earnings in Q12 ($)	4,160	3,701	+460	360
Self-reported earnings in Q12 from survey ($)	4,828	3,623	+1,206**	470
Employment				
Average percentage with employer-reported wages in Q12	66.6	61.0	+5.6	3.8
Percentage working in the week prior to survey interview	59.8	53.1	+6.7	5.1
Sample sizes:				
NDNH	310	300		
Survey	218	201		

Source: National Directory of New Hires; PACE three-year follow-up survey.

Note: Self-reported earnings are calculated for the week prior to the survey interview, based on reported work hours and wages, and multiplied by 13 weeks for a quarterly estimate. A majority of survey interviews occurred in the 12th and 13th follow-up quarters.

Statistical significance levels based on two-tailed tests: *** 1 percent level; ** 5 percent level; * 10 percent level.

This exhibit contrasts estimates of employment and earnings impacts based on NDNH data and survey self-reports.[55] The top panel in Exhibit G-1 above shows the degree of agreement of impact estimates for I-BEST derived from the two sources. The difference between these two impacts is $746, with a standard error on the difference of $410,[56] which is statistically significant at the 10 percent level.

Barnow and Greenberg (2015; 2019) discuss how such discrepancies in estimates can arise from survey nonresponse, survey reporting errors, and administrative undercoverage. We consider each of these possible sources of the discrepancy in turn. With respect to survey nonresponse, as discussed in Appendix Section C.3.1, it is possible that survey nonresponse contributed to this discrepancy, but it does not appear to be the major cause. With respect to survey reporting errors, correlational analysis suggests measurement noise in one or both measurement systems. The correlation in person-level quarterly earnings between the two systems at Q12 is just 0.46 for the treatment sample and 0.63 for the control sample.[57]

With respect to administrative undercoverage, one obvious form would be self-employment income. The NDNH does not include self-employment income.[58] We explored whether earnings from self-employment could explain the difference between +$460 and +$1,206 if we were to treat the difference show in the top panel of Exhibit G-1 as real.

The second panel of Exhibit G-1 shows that NDNH-based employment estimates are slightly higher than survey-based estimates for both treatment group members (67 percent compared with 60 percent) and control group members (61 percent compared with 53 percent). Most of the within-group difference between the two data systems is probably due to the time frame. The percentage of respondents with any earnings over three months is bound to be higher than the percentage employed on a particular day. This appears to have affected the two groups nearly equally because the estimated impacts of I-BEST on employment at Q12 are very

---

[55] From the follow-up survey, we had a complete history of jobs, with the starting wage and hours for each job as well as the last wage and hours for each job. We combined these to establish weekly earnings for the first and last weeks of a job. We then interpolated to get wages for each intervening month. We then summed weekly wages across jobs for multiple-job holders to get weekly earnings for every week between randomization and interview. Finally, we summarized these to the person-quarter level.

[56] Assuming a correlation of 0.54 between the two person-level latent effects (the average of the correlations between NDNH- and survey-reported earnings for the two groups), the standard error of the difference between the two estimated impacts is $410.

[57] The survey figures convert the available survey measure—earnings in the prior week (calculated as hourly wage multiplied by number of hours worked)—to a calendar-quarter-level estimate by multiplying by 13 (the average number of weeks in a quarter).

[58] Washington State does not require the payment of Unemployment Insurance tax for independent contractors, but it appears to have an aggressive audit policy for false claims by employers of independent-contractor status for their employees (https://www.esd.wa.gov/employer-taxes/independent-contractors). It is impossible for us to know what proportion of "gig" jobs such as driving for Lyft or Uber are covered in these records.

similar.[59] This similarity of impact estimates on employment suggests that self-employment is not an important source of the difference in estimated impacts on earnings.

Of course, self-employment is not the only form of undercoverage in the NDNH. The NDNH does not cover most workers at small farms, at railroads, or at selected nonprofit organizations (particularly churches) and some casual or irregular jobs. Though we have no reason to expect it, if control group members are more likely to find employment of the type undercovered by the NDNH, then that could lead to positive bias in the NDNH-based impact. Hiding of tip income and income from household employment (such as childcare and cleaning) are additional potential sources of undercoverage. We also looked at differences in mean earnings among respondents who are classified as employed at Q12 in both systems. In the control group, the mean survey-reported earnings for those employment in both systems are *lower* than NDNH-reported earnings by $501 with a standard error of $452. In the treatment group the pattern is reversed. That is, the mean survey-reported earnings for those employment in both systems are *higher* than NDNH-reported earnings by +$982 with a standard error of +$398. The difference in differences is $1,483 with a standard error of $602, so the difference between the impacts in the two systems appears to be likely due to reporting differences for employed persons, rather than something to do with self-employment.
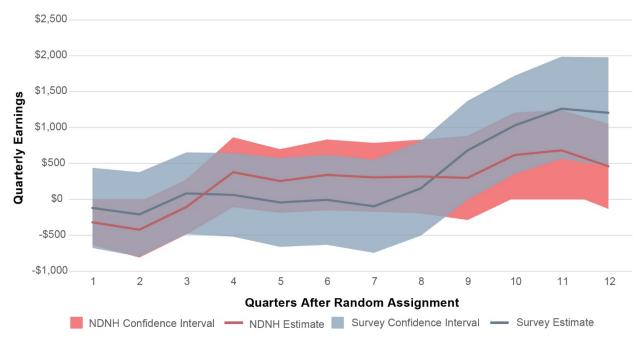
Looking at the whole history leading up to Q12, Exhibit G-2 below shows how estimated impacts based on NDNH-reported and survey-reported earnings varied over time.[60] The red line shows the NDNH-based estimated impacts by quarter, and the red-shaded area shows the 90 percent confidence intervals for those estimates. The blue line shows the survey-based estimated impacts by quarter, and the blue-shaded area shows the 90 percent confidence intervals for those estimates.

For Q1 through Q8, the systems were in very close agreement; then, starting in Q9, the two estimates start drifting apart. However, the difference in estimates is significant only in Q12 and only at a 10 percent level of significance. Thus, there is only weak evidence of any true difference in impacts between the two systems. It is also possible that the observed differences are due to random errors in the two systems. If the difference is real, it is impossible to say which data source is better.[61]

---

[59] We did not calculate correlations between NDNH- and survey-reported employment for the two groups; however, using the average correlation for the earnings as we did before, we obtain an approximate standard error for the difference between the two estimated impacts of 4.4 percentage points, which is larger than the difference between the two impacts. Therefore, this difference is not statistically significant.

[60] These are based on the maximal sample available for each system. For NDNH, that means everyone with a valid Social Security number. For the survey, that means all respondents at three years after randomization.

[61] One might try to argue that survey recall errors should be reduced for the later quarters, because the employment was closer to the time of the interview, but that would be speculative.

Exhibit G-2: **Impacts of I-BEST on Earnings over Time Based on NDNH Wage Records and Survey Self-Reports**

Source: National Directory of New Hires; PACE three-year follow-up survey.

Note: The red line shows NDNH-based estimated impacts and their associated 90 percent confidence intervals. The blue line shows survey-based estimated impacts and their associated 90 percent confidence intervals.

This report stresses the NDNH-based estimates because the study registration pre-specified NDNH-based estimates as confirmatory.[62] The team registered this measurement system as primary for several reasons already discussed but perhaps worth reiterating:

- No survey response errors in the NDNH.

- No danger of nonresponse bias in the NDNH.

- Reduced standard errors because of no sample loss to nonresponse.

- Reduced standard errors because of a lower population standard deviation for NDNH-reported earnings than for survey-reported earnings.

- Nothing known about the program design at the time of registration that would encourage self-employment.

- A sense that this is the standard for the field. We did this for all nine PACE sites. The NDNH is also the primary source for earnings impacts in the parallel evaluation of the Health Profession Opportunity Grants (HPOG) Program (Peck et al. 2019).

---

[62] See https://osf.io/kfyxc/?pid=wcus9 for the three-year report registration.

Appendix H: Treatment of Outliers

We took a conservative approach to outliers, retaining extreme values except where they were clearly impossible. This approach is based on the general difficulty of discriminating between errors and legitimate large values and on the fact that remedies require assumptions about true values that may not be correct.

Trimming observations could easily introduce non-ignorable nonresponse by making nonresponse a function of Y.[63]

Winsorizing observations (also known as top-coding, where values above a threshold are set equal to the threshold) could introduce bias if there is a treatment impact but the same threshold is used for treatment and control group members (and there is no reasonable basis for setting different thresholds for the two groups).

Furthermore, evidence suggests that results are generally robust to extreme values. In particular, research by Judkins and Porter (2016) and Lumley et al. (2002) indicates that for the sample sizes available in this evaluation, ordinary least squares inference on the reported data should be robust to outliers.

Outcomes assessed for extreme values included instructional hours (by type of instruction), credits, and National Directory of New Hires earnings. We found no values that were clearly impossible, and thus retained all reported values in the analysis.

---

[63] Trimming by definition creates item nonresponse because the provided response is discarded. If trimming is a function of observed Y, as is standard, and if there is some relationship between observed Y and true Y, then item nonresponse becomes a function of true Y, which is known as "non-ignorable nonresponse." Because there is no known way to remove bias due to non-ignorable nonresponse, trimming is likely to create uncorrectable biases in estimated treatment effects.

Appendix I: Cost-Benefit Analysis Supplement

This appendix supplies methodological details and reports supplementary findings for the cost-benefit analysis (CBA) in Chapter 6. Methodological details include discussion of assumptions and approximations and descriptions of data sources. Supplemental findings augment, support, and provide additional detail to the higher-level findings reported in Chapter 6.

A cost-benefit analysis plan (Dastrup, Burnett, and Buron 2017) provided an in-depth conceptual overview of the purpose of cost-benefit analyses in the three-year evaluation, along with a plan to conduct CBAs in up to six of the nine PACE programs. That document included discussions of methodological details. It also sketched several alternative approaches to assumptions and data sources for the analysis. The preferred alternative would depend on the specific program contexts, the pending impact study findings, and data availability for the various programs.

This appendix gives a high-level overview of the approach to the CBA, but does not reprise the conceptual detail provided in the analysis plan. The sections of this appendix detail the assumptions and data sources used for the CBA of Washington State's Integrated Basic Education and Skills Training (I-BEST) program given the program's unique features, the realized impact study findings, and the data available for the analysis.

The main result of the CBA is the *net benefit* of the I-BEST program. Exhibit I-1 summarizes the findings of the CBA. For each component of the analysis, the exhibit reports costs, benefits, and the combined net benefit from each stakeholder group perspective of interest and for society as a whole.

Each cost or benefit component estimate is the difference between the treatment and control group averages for that component. **Net benefit** is the sum of all benefit components minus the sum of all cost components. In general, the difference is either (1) the treatment group average cost or benefit less the control group average for the same cost or benefit, or (2) a direct estimate from a statistical model of the average difference between treatment and control group of the particular cost or benefit. The CBA used observed study participant and program data wherever possible to measure and monetize inputs and outcomes. Complete data on all benefits and costs were not always available. In these cases, values for costs or benefits are imputed or approximated using benchmarks from external research or the CBA team's best estimate.

The CBA's focus is whether benefits are greater than costs for society as a whole. However, whether the benefits of an intervention justify the costs typically differs when benefits and costs are limited to a given stakeholder's perspective. So the CBA also estimates benefits and costs from the perspectives of treatment group members (relative to the control group), the federal government, Washington state/local government, and the rest of society (primarily the Open Society Foundations). These perspective estimates are presented in the columns of Exhibit I-1.

Exhibit I-1: **Costs and Benefits of I-BEST, by Perspective**

	Component	Participants	Government, Federal	Government, State/Local	Rest of Society	Society as a Whole (sum)
Costs ($)						
A	I-BEST services or alternatives available in the community, total	−859	972	2,774	2,815	5,702
B	Postsecondary education and training other than I-BEST	285	-343	-1,359	-7	-1,424
	Total Cost	**-574**	**629**	**1,415**	**2,808**	**4,278**
Benefits ($)						
C	Net present value of earnings after random assignment (Q1-Q16)	3,097[a]	0	0	0	3,097
D	Fringe benefits	744	0	0	0	744
E	Taxes (federal, payroll, sales)[b]	−270	510	130	0	370
F	Public assistance (SNAP/WIC, housing assistance)[d]	-750	854	0	0	104
	Total Benefits	**2,821**	**1,364**	**130**	**0**	**4,315**
Net Benefits = Total Benefits – Total Costs ($)						
	Net Benefit, per Participant Q1-Q16	**3,395**	**735**	**-1,285**	**-2,808**	**37**

Source: PACE cost data interviews and I-BEST program financial records; PACE 18-month and three-year follow-up surveys; research team approximations of costs of alternative services accessed by the control group; Delta Cost Project Database, Integrated Postsecondary Education Data System; research team investigation. National Directory of New Hires; National Bureau of Economic Research TAXSIM model (Feenberg and Coutts 1993); Department of Revenue, Washington State (for sales tax calculations); Consumer Expenditure Survey by Income Quintiles (Table 1203). Sources for estimating public assistance listed in Exhibit I-9.

[a] This impact estimate has standard error of $3,058, and an associated *p*-value of 0.546.

[b] Washington state does not have a state income tax.

Subsequent sections of this appendix detail the CBA's approach to estimating each of the costs and benefits listed in Exhibit I-1. Costs included in the analysis are for:

- **I-BEST services or alternatives available in the community.** For treatment group members, this includes costs for direct services and administrative expenses incurred in operating the I-BEST program, which comprised several core components: a structured career pathway, team-teaching instruction that combines basic skills and job training, and supportive services such as a dedicated program coordinator (as described in Section 1.1 and Section 1.2). The CBA includes the three I-BEST programs in the PACE project. Each of these three implemented the model somewhat differently to meet student needs and the local economy, but all include dedicated advising and "fill the gap" financial support for students beyond typical sources as well as the core components. Control group members were students enrolled in the community colleges, so although they did not have access to the I-BEST program, they were engaged in basic skills and other training at the colleges and could participate in advising, financial assistance, and job search assistance available at the college and elsewhere in the community. All services provided by a college or training institutions outside of I-BEST are included as a cost of postsecondary education and enrollment, (the next cost component). As documented in Section 3.1.2 of the *Implementation and Early Impact Report* (Glosser et al. 2018), treatment and control group members could access job search supports from other community providers (e.g., Goodwill Industries, American Job Centers). However, since most treatment and control group members were recruited into the PACE project from SBCTC colleges, there was relatively little use of such outside supports. The CBA approximates the average cost per control group member of such alternative services, as detailed below.

  Costs of the I-BEST program are distributed across stakeholder perspectives based on program and college revenue sources. The three I-BEST programs in the PACE study have two primary sources of funding: 1) state government funding, through the State Board for Community and Technical Colleges (SBCTC), of I-BEST programs throughout the SBCTC system, and 2) funding from the Open Society Foundations (included in the "rest of society" perspective) provided to the three programs for program enhancements and expansion to facilitate the PACE study. Additional costs are incurred through Pell Grants, as treatment group participation in I-BEST includes enrollment in the college, for which the college receives some Pell-Grant funded tuition and fees, and students receive some remitted Pell Grant balances (i.e., the amount of Pell Grants in excess of tuition and fees).

- **Postsecondary education and training other than I-BEST.** This includes direct costs associated with all postsecondary education and training (for which impacts are reported in Chapter 3), as measured by institution expenses.[64]

  Overall, assignment to the I-BEST program actually *reduced* the costs of education and training outside of the program. Details on how this is calculated in the CBA are provided in Exhibit I-3 below. Although it increased the amount of education and training obtained, primarily in the form of workforce credits earned, the additional courses were largely within the I-BEST program. The within-program costs are already included in the prior cost component.

  In assessing costs from various perspectives, this reduced enrollment results in participant costs that are higher for the treatment group because they have less remitted Pell Grant balances as compared to the control group. Federal and Washington State and local governments have lower costs due to lower enrollments other than I-BEST by the treatment group.

The bottom panel of Exhibit I-1 shows the primary benefits considered in the CBA. They are:

- **Earnings**. I-BEST's effect on earnings is the main expected benefit. As explained in this chapter's introduction, this CBA includes impacts on total earnings over the first 16 quarters after random assignment. Changes in a number of items that result from increased earnings are included in the CBA:

  o **Fringe benefits.** Increases in earnings and full-time work imply increases in fringe benefits such as health insurance, employer retirement contributions, paid vacation, and sick leave. These benefits represent a value to participants in addition to earnings. The value of fringe benefits is estimated by multiplying observed earnings by external estimates of average benefits as a share of income, adjusted for the proportion of participants that receive benefits.

  o **Taxes.** Increased earnings also generate increases in taxes for treatment group members. The analysis includes estimated amounts for income (federal only, as Washington has no state income tax), payroll, and sales taxes (assuming increased earnings increase taxable purchases). Because taxes represent a transfer from study participants and employers to federal, state, and local governments, the amounts cancel each other out in the society as a whole perspective. Increased employer contributions to payroll taxes based on participants' wages do represent a benefit to society.

---

[64] Direct costs are inclusive of expenditures on instruction, public service, academic support, student services, institutional support, operations and maintenance, scholarships and fellowships, and depreciation, as reported by institutions. The primary indirect cost that is not included here is potential foregone earnings during any training and education funded or induced by I-BEST. Rather than split earnings into during-training and after-training periods, we analyze total earnings over the follow-up period as a component of benefits.

      o **Public assistance.** Increases in earnings reduce treatment group members' eligibility for, and receipt of, means-tested public assistance. The loss to participants and benefit to government budgets largely cancel each other out, although there is a small net savings to society from reduced governmental costs for administering programs. Public assistance receipt is estimated by multiplying earnings impacts by external estimates of how increases in income affect benefits, and adjusting for survey-measured rates of benefit receipt by study participants. As detailed below, this CBA considers public assistance items that are included in the PACE surveys, which provide estimates of the incidence of benefit receipt. These items are food assistance (Supplemental Nutrition Assistance Program/SNAP or Special Supplemental Nutrition Program for Women, Infants, and Children/WIC), Temporary Assistance for Needy Families/TANF or other cash public assistance, Unemployment Insurance, housing assistance, and Medicaid (public health insurance).

The first three sections of this appendix provide details of the methodology, with supplementary findings related to the labeled rows of Exhibit I-1.

- Section I.1 details the approach to calculating the cost of the I-BEST program and approximating the cost of control group alternative services. (row A)

- Section I.2 details the approach to estimating the cost of education and training obtained after random assignment and provides supplementary analysis about this cost component. (row B)

- Section I.3 reports the approach to measuring differences in earnings and estimating the fringe benefit, tax, and means-tested public assistance implications of the differences. (rows C, D, and E)

The final two sections are relevant to the entire CBA.

- Section I.4 discusses the uncertainty in the analysis, including how sensitive the findings are to alternative values for assumptions made throughout the analysis.

- Section I.5 catalogues the various data sources used in the CBA.

I.1   Cost of I-BEST Program and Control Group Alternatives

This section describes methods for determining the first cost component of the CBA, the cost of the I-BEST program less the costs of similar services that control group members may have accessed (row A of Exhibit I-1). This section first details the data collection and estimation process used to estimate the average per-participant costs of the I-BEST program. It then turns to the approach for approximating costs of similar services used by the control group.

I.1.1   Cost of the I-BEST Program

To estimate the per-participant cost of I-BEST, the CBA analysis identified all program inputs and expenditures on the inputs. The sum of those expenditures are then divided by the number of treatment group members. To analyze the share of costs by type of input, the CBA also

assigned expenditures to cost categories: program activities (e.g., instructor and specialists salaries and course materials, advising and support staff); program assistance (e.g., tuition or transportation support, supplies, and fees); and administration and overhead. The resources listed in Exhibit I-2 below are used to develop this estimate of per-participant costs.

Exhibit I-2: **Sources Used to Estimate Per-Participant Cost of the I-BEST Program**

Resource	CBA Use
I-BEST program profile (Glosser et al. 2014) and *Implementation and Early Impact Report* (Glosser et al 2018)	• Preliminary identification of program inputs
Program cost interview (in 2015) with I-BEST program administrators	• Confirm comprehensive list of program inputs • Determine that program expenditure reports and ITA tracking log capture costs of all inputs • Assign expenditures to categories
I-BEST program expenditure reports (expenditures by category for fiscal year 2014 or 2015)	• Determine cost of all program inputs except ITAs
PACE study enrollment data	• Identify study participants enrolled during period covered by expenditure data

Costs by Perspective. The CBA used contracts and memos detailing the Open Society Foundations grants to determine the cost to the rest of society perspective. The remaining cost was incurred by SBCTC through the normal approach to funding I-BEST, with some of this cost determined to include Pell Grant and other federal revenue based on the revenue shares calculated from IPEDS data (detailed below) for the three I-BEST colleges in the PACE study.[65]

I.1.2 Cost of Similar Services Accessed by Control Group Members

The next element needed to calculate the cost of the I-BEST program is the cost of similar services accessed by control group members. The control condition in the evaluation did *not* prohibit access to services outside of I-BEST. Control group members did not have access to the I-BEST program, but they could participate in all other education and training, financial assistance, and job search assistance available either at the community colleges they were attending or in the community. Chapter 4 of the *Implementation and Early Impact Report* (Glosser et al. 2018) document that control group members accessed supportive and employment services, but at a lower rate than treatment group members did.

Control group members were primarily students enrolled in the community colleges, so although they did not have access to the I-BEST program, they were engaged in basic skills and other training at the colleges and could participate in advising, financial assistance, and job search assistance available at the college and elsewhere in the community. All services provided by a college or training institutions outside of I-BEST are included as a cost of postsecondary education and enrollment, (the next cost component). As documented in Section 3.1.2 of the *Implementation and Early Impact Report* (Glosser et al. 2018), treatment and control group

---

[65] As reported in Section 1.1.1, SBCTC reimburses colleges 1.75 times the regular rate for a full-time-equivalent student to help cover the costs associated with implementing I-BEST in the system-wide I-BEST initiative. However, this reimbursement does not directly determine costs or provide a complete revenue picture for the three colleges in the PACE study.

members could access job search supports from other community providers (e.g., Goodwill Industries, American Job Centers). The CBA approximates the costs of the control groups' use of such community provider services. This cost is approximated rather than estimated because, though the PACE survey includes information on the *incidence* of the use of such services, the CBA does not have information that details whether such services were received from community providers, or on the total *quantity* of services that control group members used.

In consultation with authors of the *Implementation and Early Impact Report*, the CBA team determined that control group are not likely to have used services provided by such external providers to a substantially greater degree than treatment group members. So, the CBA does not assess a substantial cost for control group use of alternative services. To allow for the possibility of such a difference, the CBA assumes that 5 percentage points more treatment group members used such services than did control group members. Use of such services is assigned an average cost of $1,143, which includes a bundle of services (individual training accounts, program activities, and supportive assistance) with cost estimates for AJCs from Fortson et al. (2017). Adjusting for a regional price parity value of 1.11 and multiplying by the 5 percentage point approximation results in the approximated value of $63 reported in Exhibit 6-2. This amount is assumed to be funded by federal grant programs in the perspective analysis.

I.1.3   Cost of the I-BEST Program

The cost of the I-BEST program is then the estimated cost per treatment group member of the program minus the approximated cost per control group member of alternative services. The CBA separates the non-ITA and ITA amounts as subtotals of this cost in row A of Exhibit I-1.

I.2   Cost of Education and Training Other than I-BEST

This section considers the second cost component of the CBA, the estimated effect of assignment to the I-BEST treatment group on costs of additional, education and training other than I-BEST enrollment after random assignment (row B of Exhibit I-1). The section first reports methods and data used for estimation. It then provides some supplementary analysis that buttresses the assertion that all costs of the I-BEST are realized within the first three years after random assignment.

The cost of the I-BEST intervention should include the costs of all education and training obtained after random assignment. The intervention is expected to affect enrollment in education and training, which results in a change in costs to society. Education and training (and associated costs) will increase if, for example, education and training obtained as an initial result of the I-BEST program is the first of multiple stages along a new or accelerated career trajectory.[66] Alternatively, I-BEST services could initiate a successful career for treatment group members that replaces less directed education and training and thus lowers costs to society of education and training.

---

[66]   This concept of career pathways is common to all PACE programs, and implies that I-BEST participants may obtain additional education and training later in their career pathway because of their participation in the PACE program.

Again, I-BEST program services costs reported above include the costs of I-BEST enrollment for treatment group members, included in the prior sections. Meanwhile, control group members were largely students attending the same colleges, but had enrollment limited to standard, non-I-BEST offerings. The cost of this enrollment is included in this section. As a result, control group members have *more* postsecondary education and training other than I-BEST than do treatment group members. This represents a cost savings to society as a whole for this component (a negative cost in row B of Exhibit I-1) that partially offsets the I-BEST program costs.

I.2.1   Methods and Data

The CBA estimates the cost of education or training as the product of a *quantity* measure of units of education and training received and a *unit cost* of the education and training. The quantity of education obtained is derived from the impact estimates reported in Chapter 3, which are measured using individual-level information from SBCTC records and the three-year follow-up survey. We estimate unit cost from external institution-level estimates.

To determine the quantity of education or training received, the CBA builds on the Chapter 3 analysis of impacts on postsecondary education or training. Exhibit 3-3 in Chapter 3 reports a quantity measure of education or training obtained: full-time-equivalent (FTE) months enrolled at any school. This estimate is reproduced in the first row of Exhibit I-3. Since this measure does not distinguish between I-BEST enrollment (for which the CBA has already measured costs) and non-I-BEST enrollment, the CBA relies on Exhibit 4-2 of the *Implementation and Early Impact Report* (Glosser et al. 2018) to estimate this enrollment. That exhibit reports the number of credits earned and average number of quarters attended for each I-BEST college and program in the study. Based on analysis of that exhibit, the CBA estimates that treatment group members had an average of 3.7 FTE months of I-BEST enrollment. The unit quantity estimate used for education and training other than I-BEST is then the impact estimate on FTE months enrolled in any school (2.6) minus the estimate I-BEST FTE months enrolled (3.7), which equals -1.1. That is, control group members had approximately 1.1 more FTE months enrolled of education and training other than I-BEST than did treatment group members.

Exhibit I-3:   Estimate of FTE Months Enrolled Other than I-BEST

Outcome	Treatment Group	Control Group	Difference
A. FTE months enrolled in any school (Exhibit 3-3)	6.9	4.3	+2.6[a]
B. CBA estimate of I-BEST enrollment	3.7	0	3.7[b]
Estimated enrollment other than I-BEST (A-B)	3.2	4.3	-1.1

Source: Exhibit 3-3, CBA estimations based on Exhibit 4-2 of Glosser et al. (2018)

[a] This impact estimate has a standard error 0.73, and an associated p-Value of <.001

[b] This difference is estimated based on analysis of Exhibit 4-2 (and not using individual-level data), and does not have an associated standard error.

The CBA determines the unit cost of each FTE month as the average total cost of expenditure per-FTE month of enrollment an institution's that study participants report attending. The average is calculated with weights for the number of FTE months enrolled at each institution (i.e., it is averaged over FTE months attended), so that the value is adjusted to reflect that

amount of enrollment at the various institutions. This value is calculated or imputed for each institution that I-BEST study participants reported attending using data from the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS) supplemented with additional study data.[67] The primary source of definitions used to develop unit costs is the Delta Cost Project Database (DCPD), an extract and analysis of educational institution finances from IPEDS (Hurlburt, Peek, and Sun 2017).[68] Exhibit I-4 details the variables the CBA constructs using DCPD definitions applied to IPEDS variables, along with a few additional IPEDS variable definitions. All dollar-denominated variables are expressed in terms of FTEs and adjusted for inflation to 2013 dollars using public access code downloaded with the DCPD.[69]

Exhibit I-4:   DCPD/IPEDS Variables Used in the CBA

DCPD/IPEDS Variable	Variable Description from DCPD/IPEDS Documentation	Variable Use in CBA
Total education and general expenditures–current year total (eandg01_fte_cpi)	Includes all core operating expenditures, including sponsored research, but excluding auxiliary enterprises. This variable was originally reported in IPEDS, but for recent years it is calculated by summing expenditures on instruction, research, public service, academic support, student services, institutional support, operations and maintenance, and scholarships and fellowships.	Primary component of total per-participant costs.
Depreciation–current year total (depreciation01_fte_cpi)	Allocation or distribution of the cost of capital assets, less any salvage value, to expenses over the estimated useful life of the asset in a systematic and rational manner. Depreciation for the year is the amount of the allocation or distribution for the year involved. This field is used if the institution has not allocated all depreciation to other functions.	Added into total per-participant costs where reported.
Expenditures for other non-operating–current year total (othernon01_fte_cpi)	Other non-operating expenses and deductions. Total expense is the sum of all expenses incurred other than interest that are not classified as operating expenses.	Added into total per-participant costs where reported.

---

[67] Study participants attended a small number of institutions that do not submit data to IPEDS (e.g. for-profit non-degree granting training providers). However, most I-BEST study participants (treatment and control) were enrolled in the three I-BEST colleges, over 90 percent of reported FTE months enrolled (in the PACE survey data) were at the three colleges, and the impact estimates in Exhibit 3-3 imply a difference of only 0.01 FTE months enrolled at non-SBCTC institutions between the treatment and comparison groups. As a result, the CBA does not develop separate unit price estimates for FTE months enrolled at institutions not reporting data to IPEDS.

[68] We developed an FTE month count for each spell in the follow-up survey responses based on responses about whether students were attending full-time or part-time. FTE months were calculated for DCPD data by dividing the IPEDS variable on annualized full-time-equivalent undergraduate enrollment by 12. That variable is defined in IPEDS documentation as follows: "For institutions with a semester, trimester, or 4-1-4 plan, the number of FTE undergraduates is the sum of undergraduate credit hours divided by 30 and contact hours divided by 900. For institutions with a quarter plan, undergraduate credit hours divided by 45 and contact hours divided by 900. For institutions with continuous enrollment over a 12-month period, undergraduate credit hours were divided by 30 and contact hours were divided by 900."

[69] The code uses a CPI-U inflation adjustment variable that is included with the data download.

DCPD/IPEDS Variable	Variable Description from DCPD/IPEDS Documentation	Variable Use in CBA
Non-educational share of expenses (noneducation_share)	Research and public service portion of spending on instruction, student services, research, and public service.	Used to prorate total per-participant costs to education share of expenses only. Imputed as state-level sector by highest-degree-offered mean where missing.
Net tuition and fees revenue (nettuition01)	Amount of money the institution takes in from students after institutional grant aid is provided (this is not the same as the net tuition number available in IPEDS, which is net of all discounts and allowances applied to tuition and fees).	Numerator when calculating tuition and fees share of cost for perspective analysis.
Published in-district tuition and fees (may be lower than in-state)	The tuition charged by institutions to those full-time undergraduate students who meet the state's or institution's residency requirements.	Used to estimate PACE participant out of pocket expenditures.
Average federal grant aid awarded (state_grant_avg_amount)	Average amount of federal grants (grants/educational assistance funds) received by first-time, full-time undergraduate students. Includes Pell Grants, Supplemental Educational Opportunity Grants, and other federal agency grants.	
Average state grant aid awarded (state_grant_avg_amount)	Average amount of state/local grants (grants/scholarships/waivers) received by first-time, full-time undergraduate students. Includes state and local monies awarded to the institution under state and local student aid programs.	
Net tuition directly from students (net_student_tuition)	Net tuition revenue coming directly from students (not including Pell, federal, state, and local grants).	Numerator when calculating student out-of-pocket costs share for perspective analysis.
Pell grants (grant01_fte_cpi)	Gross amount of Pell grants disbursed or otherwise made available to recipients by the institution.	Numerator when calculating the Pell and other federal grant awards cost share for perspective analysis.
Other federal grants (grant02_fte_cpi)	Expenditures for scholarships and fellowships, excluding Pell grants, which were funded from federal government agencies. This includes Supplemental Educational Opportunity Grants (SEOG) and State Student Incentive Grants (SSIG), but not loans or College Work Study Program.	
Revenue from federal appropriations, grants, and contracts (federal10)	The total amount of revenue coming from federal appropriations, grants, and contracts.	Less Pell and other federal grants is numerator for calculating federal non-grant cost share for perspective analysis.
Revenue from state and local appropriations (state_local_app)	The total amount of revenue from state and local appropriations.	Combined to determine the numerator for the state and local cost share for perspective analysis.
Revenue from state and local grants and contracts (state_local_grant_contract)	The total amount of revenue from state and local grants and contracts. Grants by state government include expenditures for scholarships and fellowships that were funded by the state.	
State grants (grant03) and local grants (grant04)	Grants by local government are for scholarships and fellowships that were funded by local government.	

DCPD/IPEDS Variable	Variable Description from DCPD/IPEDS Documentation	Variable Use in CBA
Revenue from affiliated entities, private gifts, grants, and contracts; investment returns; and endowment earnings (priv_invest_endow)	The total amount of revenue coming from affiliated entities, private gifts, grants and contracts, investment returns, and endowment earnings. Endowment earnings stopped being reported to IPEDS in 1997 for FASB-reporting institutions and 2002 for GASB-reporting institutions.	Numerator when calculating the other cost share for the perspective analysis.

Key: FASB = Financial Accounting Standards Board. GASB = Governmental Accounting Standards Board. IPEDS = Integrated Postsecondary Education Data System

Source: Integrated Postsecondary Education Data System and Delta Cost Project Database data documentation.

The CBA uses a five-year average, 2011-2015, of annual DCPD unit cost estimates to estimate net costs of education and training. DCPD data are only available through the 2014-2015 school year, whereas education spells recorded in the PACE three-year follow-up survey responses extend into the 2016-2017 school year. So the analysis must impute unit cost data for 2015-2016 and 2016-2017 to calculate total costs. Rather than applying observed cost estimates in earlier years and imputed unit costs in later years, the CBA uses a five-year average of inflation-adjusted, observed costs for all periods. Trends in unit costs during the study period are not anticipated to contribute to *net* costs, and so the CBA does not attempt to model or approximate trends in costs for any institutions.

Exhibit I-5 reports the total *unit cost* estimate of $1,251, which is the FTE-month weighted average over all attended institutions, and the institution-level costs for the most attended institutions.

Exhibit I-5:   Per-FTE Monthly Total Costs at Most-Attended Institutions and Total

Institution	Share of All FTE Months Attended in Q1-Q13 (%)		Estimate of Per-FTE Monthly Total Cost of Instruction ($)	Estimated PACE participant out of pocket costs
	Treatment Group	Control Group		
Bellingham Technical College	55.3	34.9	1,428	-228
Everett Community College	20.9	26.8	1,157	-281
Whatcom Community College	19.0	26.5	974	-223
Edmonds Community College	2.1	2.4	1,304	-221
Skagit Valley College	0	4.8	1,181	-176
Total (weighted by FTE Months for all attended institutions)	*100*	*100*	*1,251*	*-230*

Source: PACE three-year follow-up survey; Integrated Postsecondary Education Data System; Delta Cost Project Database.

The CBA then multiplies this weighted average cost of enrollment by the estimated FTE months enrolled other than I-BEST (Exhibit I-3) to estimate the cost of education and training other than I-BEST. The resulting estimated impact on education and training costs is reported in Exhibit 6-2 and incorporated into Exhibit I-1.

Costs by Perspective. To assign costs of education and training to different perspectives, the CBA relies on two types of estimates. First, the CBA estimates out-of-pocket expenses for study participants. Second, it estimates the share of an institution's revenues that are from various

sources that correspond to each stakeholder perspective. Overall revenue shares are used to estimate cost shares for stakeholder perspectives (except participants themselves) because institution educational expenditures (realized costs) are typically not directly linked to specific revenue sources. That is, the CBA assumes an institution's revenues are fungible across all educational expenditures.[70]

The CBA estimates out-of-pocket expenses of education and training for study participants using the same data sources used to approximate unit costs. It first calculates the average net student out-of-pocket tuition and fee amount (which is often negative, meaning students are remitted Pell and other grants). The following details of education and training providers and study participant characteristics informed the estimation.

For student out-of-pocket expenses, the CBA assumes the following (for both treatment and control groups). In each case, IPEDS amounts are weighted averages over institutions that PACE participants report attending, weighted by FTE months enrolled.

- Study participants are assessed the published charged tuition and fees (on a per-FTE month enrolled basis) reported to IPEDS.

- Study participants receive the average federal (primarily Pell) and state grant amounts received by first-time full-time students receiving grants reported to IPEDS.

- The received grants are first applied to charged tuition and fees, and any remaining funds are dispersed to the study participant.

The costs of education and training from the remaining perspectives are estimated using institution-specific revenue shares calculated from IPEDS variables (using DCPD definitions), also weighted across institutions by reported FTE months enrolled. Specifically, the CBA calculates average per-participant costs of education and training for each subgroup perspective as follows:

- For the federal government perspective, total costs are multiplied by the share of institutions' non-tuition and fee revenues that are from federal sources. These are primarily student-level grants (Pell and others), but also include other federal appropriations.

- For the state and local government perspective, total costs are multiplied by the share of institutions' non-tuition and fee revenue that is from state and local appropriations and grants.

- For the rest of society perspective, total costs are multiplied by the revenue share from other sources (e.g., endowment revenue and private donations).

Exhibit I-6 reports the breakout of revenue across these sources for the same most-attended institutions included in Exhibit I-5. Annual shares for student out-of-pocket revenue can be

---

[70] The analysis does, however, assume that other students' out-of-pocket tuition and fees support their own costs of attendance, rather than subsidize PACE participant's costs.

negative because an institution's total payments to students (e.g., as disbursements of Pell and other grants) exceed tuition and fees received from students.
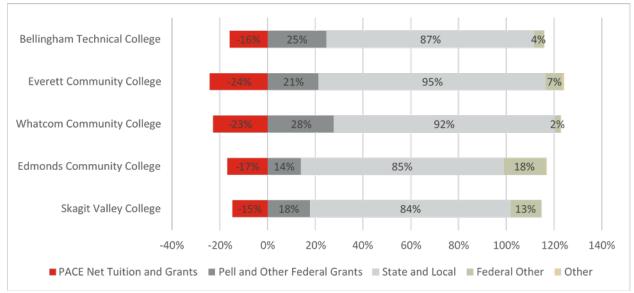
Exhibit I-6:   Per-FTE Monthly Total Costs at Most-Attended Institutions

Source: PACE three-year follow-up survey; Integrated Postsecondary Education Data System; Delta Cost Project Database;
Note: Shares for PACE participant net tuition and grants can be negative because institution's total payments to students can exceed tuition and fees received from students.

I.2.2   Potential for Future I-BEST Costs

The CBA estimates education and training from the three-year follow-up survey. As a result, any impact on education training past three years after random assignment is not included. This section discusses the likelihood of differences in education and training cost between treatment and control group members that persist beyond the three-year follow-up survey.

Chapter 6 argues that essentially all costs of the I-BEST program were incurred in those first three years. Two facts support a conclusion that costs are not understated due to the three-year data window for education and training because all costs of the I-BEST program were incurred in those three years.

First, all study members have completed their engagement with the I-BEST program itself.

Second, the quarter-by-quarter analyses reported in Exhibits 3-2 and 3-5 suggest that any meaningful differences in education and training enrollment between the treatment and control groups occurred before Q12. Exhibit 3-2 shows that differences in college enrollment for treatment and control group members occurred primarily in the first five quarters after random assignment. Exhibit 3-5 shows that impacts on cumulative credits earned grew from Q1 through Q9 but then levelled off at approximately 11 credits for the last three quarters. These trends together lead the CBA to conclude that differences in total costs of enrollment have also stabilized and will not continue to accrue.

I.3 Earnings Impacts, Fringe Benefits, Taxes, and Means-tested Assistance

Earnings impacts reported in the CBA (row C of Exhibit I-1) are estimated with the same impact model as earnings impacts in Chapter 4, except that individual-level earnings are discounted to the time of random assignment to account for inflation and the time value of money. The CBA calculated the NPV at the time of random assignment as the sum of all discounted earnings. The nominal discount rate of 5 percent aligns with the 3 percent used for the inflation-adjusted education and training costs under an assumption that inflation is approximately 2 percent.

Increases in earnings and full-time work imply increases in fringe benefits such as health insurance, employer retirement contributions, paid vacation, and sick leave (row D, Exhibit I-1). Estimates in this analysis multiply earnings gains by external estimates of average fringe benefit value. This approach follows Schaberg and Greenberg's (2020) CBA of WorkAdvance. To develop assumptions about appropriate multipliers and the value of fringe benefits as a percentage of earnings, the CBA research team consulted the Employer Costs for Employee Compensation (ECCC), Compensation Percentiles, from the National Compensation Survey, produced by the U.S. Bureau of Labor Statistics and national averages of fringe employment benefit receipt from Solis and Galvin (2012).

Following Schaberg and Greenberg, the ratio of the value of total fringe benefits to total wages for private industry workers reported in the ECCC from 2016 to 2018 averaged 42.5 percent. I-BEST participants' median wages were comparable to the median wages over this period in the ECCC, so we assume that the 42.5 percent value of benefits applies to treatment group participants earnings impacts. A comparison of figures reported in Solis and Gavin (2012) to averages for I-BEST treatment group participants indicates that treatment group participants receive fringe benefits at a lower rate than all workers (similarly to WorkAdvance participants in Schaberg and Greenberg, who receive benefits at approximately 60 percent of the prevailing rates). Additional analysis of PACE three-year follow-up survey responses shows that fringe benefit receipt rates are below national averages (with no statistically significant differences between the treatment and control group), with rates of benefit receipt that average roughly 56.6 percent of the national average. So, to estimate increases in the value of fringe benefits, the CBA multiplies earnings impacts by 0.425*0.566=0.240.

Exhibit I-7 catalogues the details of the CBA's estimation of changes in taxes and public assistance that result from earnings increases. To impute individual-level tax liabilities, the CBA relies on the National Bureau of Economic Research taxsim model (Feenberg and Coutts 1993). The model calculates tax liabilities based on a variety of inputs, including the following that the team uses in this analysis: earnings, marital status, and number of dependents by age. The PACE surveys report whether participants live with a spouse or partner and have one or more children at home (but not the number of children). We assume that half of participants with children have one child and half have two.

An important caveat in imputing tax liability is that, because of data privacy and security restrictions, individual-level earnings data from the NDNH (used to generate Chapter 3 earnings impacts) cannot be sent to the taxsim model, which is a cloud-based program. Instead, the study team groups participants into small (at lease 10, based on data use restrictions) bins

Exhibit I-7:   Estimating Marginal Effective Taxes Associated with Earnings Impacts

Element Estimated	Sources	Estimate Details
Income Taxes (federal, state, and payroll)		
Effective tax amounts	NBER Internet TAXSIM and analysis of NDNH earnings data	Tax profiles for small bins used to impute individual-level tax liability estimates.
Effective Sales Tax		
Average share of income spent on taxable purchases	Consumer Expenditure Survey by Income Quintiles (Table 1203) DRWS[b]	37% of income spent on sales taxable goods and services
Sales tax rate	DRWS	8.89%
Implied Multiplier		*3.29%*
Marginal Effective Tax – SNAP/WIC		
Share of study participants affected	Analysis of baseline and three-year survey	51.2%
Estimate of marginal effective tax rate for food assistance	Marginal rate effect for those affected (*weighted by earnings*)	11.2%
Implied Multiplier		*5.8%*
Marginal Effective Tax – TANF or other cash public assistance		
Share of study participants affected	Analysis of baseline and three-year survey	15.9%
Estimate of marginal effective tax rate for cash public assistance	Calculated based on Hanson and Andrews (2009)	48.1%
Implied Multiplier		*7.6%*
Change in receipt of Unemployment Insurance		
Change in households receiving unemployment or workers compensation	Analysis of three-year survey	5.3 percentage points
Value of Unemployment Insurance	Estimate from Vroman (2018)	$4,200 per instance
Implied Savings		*$223*
Marginal Effective Tax – Housing		
Share of study participants affected	Analysis of baseline and three-year survey	11.9%
Marginal effective tax rate for housing assistance	Most federal housing assistance programs require recipients to pay 30% of income in rent	30%
Implied Multiplier		*3.6%*

Key: DRWS = Department of Revenue Washington State. NBER = National Bureau of Economic Research. SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] The actual number of dependents is not available in PACE data. Households with children are assumed to have one or two dependents in equal proportion.

[b] DRWS: Retail Sales and Use Tax Exemptions https://dor.wa.gov/taxes-rates/retail-sales-tax/retail-sales-and-use-tax-exemptions, last retrieved July 15, 2020; Tax rate lookup: https://webgis.dor.wa.gov/taxratelookup/SalesTax.aspx, last retrieved July 15, 2020.

based on treatment status, earnings, marital status, and the presence of children. Average earnings for each bin are calculated in NDNH. These averages are submitted to taxsim, and calculated tax liabilities are matched back to all participants in the bin, and used as imputed individual-level liabilities. Treatment and control group averages are then calculated for the

individual liabilities, with the difference used to compute the federal and payroll tax implications that go in row E of Exhibit I-4 (Washington state has no state income tax).

The second panel of Exhibit I-7 shows the approach to accounting for sales tax. The first resource is a tabulation of average expenditure shares by category for low-income households from the Consumer Expenditure Survey reported by the Bureau of Labor Statistics.[71] These shares are cross-tabulated by a list of categories of purchases subject to sales tax published by the Washington government, and determine that approximately 37 percent of low-income households' income is spent on items subject to sales tax. The sales tax rate in the jurisdictions where the three colleges are located is approximately 8.9 percent. Together, this implies a sales tax offset on income gains of 3.29 percent.

The CBA also calculates marginal effective taxes for each year for changes in sales taxes, and public assistance measured in the 36-month follow-up survey. These include food assistance (SNAP and WIC), TANF or other cash public assistance, unemployment insurance, and housing assistance (row E of Exhibit I-1). The bottom section of Exhibit I-7 provides the details to the estimation approach for each of these. The CBA estimates average marginal effective taxes of 3.29 percent for sales taxes, 5.8 percent for SNAP/WIC benefits, and 3.9 percent for housing assistance.

I.3.1   Earnings, Taxes, and Means-tested Assistance by Perspective

Assigning earnings and associated taxes and means-tested assistance to the various perspectives is a relatively straightforward exercise of noting which perspective is receiving money and, for elements that are transfers, which perspective is paying it. Transfers are amounts that flow from one perspective to another and so net out for society as a whole.

Exhibit I-8 summarizes the assignment of net costs and benefits to each perspective; it includes earnings, taxes, and means-tested assistance in its second panel. The benefit of earnings accrues to participants. Taxes are transfers from participants to state/local government, meaning higher tax amounts are a net negative benefit for participants and a net positive benefit for government. A small transfer of the employer portion of payroll taxes from the rest of society to the federal government is also included. Means-tested assistance are transfers from governments to participants.

---

[71]   https://www.bls.gov/cex/tables.htm.

Exhibit I-8:   Summary of Net Costs and Net Benefits by Perspective

Component	Treatment Group (relative to control)	Government, Federal	Government, State/Local	Rest of Society	Society as a Whole (sum)
Costs (treatment – control)					
Cost of the I-BEST program or alternative services available in the community[a]	Net benefit	Net cost	Net cost	Net cost	Net cost
Cost of no-BEST education or training in three years after random assignment	Net cost	Net benefit	Net benefit	Net benefit	Net benefit
Total Net Cost	**Net benefit**	**Net cost**	**Net cost**	**Net cost**	**Net cost**
Benefits (treatment – control)					
NPV of earnings in Q1-Q19 after random assignment	Net benefit				Net benefit
Net taxes (federal, state, payroll, including credits)	Net cost	Net benefit	Net benefit	Net cost	Transfers offset
Public assistance (SNAP/WIC, housing assistance)[a]	Net cost	Net benefit			Transfers offset
Total Net Benefits	**Net benefit**	**Net benefit**	**Net benefit**	**Net cost**	**Net benefit**
Summary					
Total Net Benefits – Total Net Costs, per Participant Q1-Q19	**Net benefit**	**Net cost**	**Net cost**	**Net cost**	**Net cost**

Key: NPV = net present value. SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] To avoid confusion in terminology in this chapter, we use the term "public assistance" when referring to what other chapters call "public benefits" (e.g., TANF, SNAP, WIC, Medicaid, etc.).

I.4   Uncertainty in Components of the Cost-Benefit Analysis

Estimated NPVs of PACE programs based on all measured costs and benefits are subject to three types of uncertainty:

(7) Sample variability;

(8) Measurement error in a single observation of I-BEST costs; and

(9) A multiplicity of options for elements that cannot be estimated from observed data but must instead be assumed from estimates available in the CBA literature.

This section discusses each of these sources of uncertainty. After describing each type of uncertainty, we discuss how sensitive the CBA findings likely are to the particular uncertainty. In brief, the uncertainty associated with earnings impact estimates is large enough such that definitive CBA findings are not possible, and other sources of uncertainty do not materially affect this conclusion. In addition to the uncertainty discussed in this section, Section 6.4 discusses two additional sources of uncertainty—intangible costs and benefits that are not monetized and items that could affect the CBA that are not included in the analysis.

I.4.1 Sample Variability and Measurement Error in Participant-level Data

Sample variability and measurement error in participant-level data[72] cause virtually all impact estimates and other parameters calculated using statistical analysis to be subject to some uncertainty. This is true even for parameters found to be statistically significant. Standard errors associated with each estimate provide a measure of the extent of this uncertainty. Larger standard errors indicate greater parameter uncertainty.

Confidence intervals are one way of expressing these standard errors. The CBA reports plausible ranges of values based on 90 percent confidence intervals in Chapter 6 for the two elements of the CBA for which estimates are based on participant-level data that use a statistical model of impacts. The costs of education and training other than I-BEST estimated at the end-points of the 90 percent confidence interval for the education and training impacts result in plausible values spanning from −$3,218 to +$371. The 90 percent confidence interval for the estimate of the net benefit of the NPV of quarterly earnings after random assignment spans from −$2,938 to +$9,132.[73]

Section 6.6 discusses how sensitive the CBA findings are to the uncertainty of these estimates. That discussion concludes that given the imprecision in the estimates of earnings and education and training impacts, definitive CBA conclusions are not possible.

Uncertainty also exists for the elements in the perspective analysis calculated using the same data on costs of education and training used to calculate total costs of education and training. These elements are program participants' estimated out-of-pocket costs and the share of revenue that institutions derive from a variety of sources.[74]

I.4.2 Measurement Error in Single Program and Site Estimates

The CBA estimates I-BEST program costs based on cost data collected for three sites. Inaccuracies in observed values result in measurement error for this cost estimate. Because costs are incurred (and observed) at the program level for each of the three programs, no standard error or comparable characterization of the resulting uncertainty is meaningful for point estimate based on three sites. However, because actual ITA costs and program expenses for treatment group members are observed for a relatively time-limited program, it is unlikely that the size of error in the measurement of program costs is materially relative to the standard

---

[72] Sample variability and measurement error both result in chance variation in outcomes not due to the I-BEST program that is in part a result of error in measuring the outcome.

[73] These two estimates are based on the same sample of individuals and thus are likely correlated. We nevertheless do not attempt to characterize any correlation in uncertainty between these outcomes. Doing so would require joint estimation of the impacts, which is not straightforward given our general approach to estimation of impacts.

[74] Unlike overall costs, we estimate statistically significant differences in the share revenues coming from student tuition and fees, state and local appropriations, and Pell and other federal grants. This analysis is correlational because it is weighted by participant FTE months attended at each institution and is thus conditional on any attendance.

errors of the estimates measured with individual-level data. This means that measurement error of I-BEST program costs is unlikely to affect the CBA conclusions.

Additionally, the CBA calculates the cost of the program by subtracting an approximation of the cost of control group member use of similar alternative services. The next section discusses the uncertainty associated with this approximation as uncertainty in assumed parameters.

I.4.3   Error in Assumed Parameters

Many parameters in the CBA must be assumed rather than estimated from available data. These assumed parameters are documented in the prior sections of this appendix or in Chapter 6. Such assumed parameter values include elements of the calculation of the marginal effective taxes in the perspective analysis, and the choice of a discount rate to account for the time value of money.

Because the calculation of marginal effective taxes is based on fractions of observed earnings, error in the approach to these calculations is some fraction of the margin of error of the earnings impact estimates. Additionally, this estimation process largely identifies transfers that affect the perspective analysis only, but not the main cost-benefit result from the perspective of society as a whole.

The CBA uses a 3 percent real discount rate to calculate the NPV at random assignment of costs of education and training undergone in the subsequent three years. A real discount rate is used because the DCPD data amounts are already adjusted for inflation to 2013 dollars. Two (2) percentage points are added to this discount rate for the earnings analysis to account for inflation, resulting in a 5 percent nominal rate. It is standard practice in CBAs to recalculate NVPs using higher and lower alternative discount rates, such as 2 and 6 percent (real) (Boardman et al. 2011).[75] The CBA recalculated earnings impacts using 3 and 8 percent nominal discount rates. Results are presented in Exhibit I-9. The impact estimate is about $150 higher when using the lower discount rate, and $200 lower when using the higher discount rate. Thus, the choice of discount rate is not material for the CBA conclusions.

Exhibit I-9:   Net Present Value of Quarterly Earnings after Random Assignment

Outcome	Discount Rate	Treatment Group	Control Group	Impact (Difference)	Standard Error	Relative Impact	p-Value
Net present value of total earnings after random assignment (Q1-Q16)	3%	$50,189	$46,902	+$3,287	$3,235	6.4%	0.31
	5%	$47,720	$44,623	+$3,073	$3,073	6.4%	0.31
	8%	$44,355	$41,517	+$2,837	$2,854	6.4%	0.32
Sample size		310	300				

Source: National Directory of New Hires.

In principle, it is possible to assess the likely combined effect of these many sources of uncertainty by conducting a Monte Carlo analysis. For such a Monte Carlo analysis, the total net

---

[75]   This recalculation is less relevant in our CBA because we examine earnings for 16 quarters after random assignment, a time frame over which choice of discount rates has relatively small consequences.

benefits minus total net costs CBA conclusion would be calculated a very large number of times (more than 10,000) with the parameters subject to substantive uncertainty all drawn with each calculation from distributions of their probable values. In the Monte Carlo framework envisioned in Dastrup et al. (2017), NPV values for all parameters that could materially alter CBA conclusions would be replaced simultaneously with each calculation. However, the uncertainty associated with earnings impacts (and to a lesser extent education and training costs) dominates other sources of uncertainty in the analysis. As a result, the characterization of uncertainty for this variable presented in Section 6.6 is an adequate characterization of the uncertainty for the entire CBA.

I.5   Data Sources

This section catalogues the information and data sources used to develop the estimates reported in Chapter 6.

I.5.1   Program Profile and Implementation and Early Impact Report

Profiles of each of the nine PACE programs, including I-BEST, (Glosser et al. 2014), were compiled as part of the PACE implementation study.[76] The I-BEST profile gives a high-level overview of the program; details program goals, target population, and structure; and describes the program's career pathways components. The information on program structure and career pathways components was used to conduct background research that identified program inputs. This research was completed in preparation for the cost data collection interviews we conducted for the CBA.

The *Implementation and Early Impact Report* (Glosser et al. 2018) included program descriptions and analysis of 18-month follow-up survey data. The research team consulted the report to develop a proxy for control group member access to similar services.

I.5.2   Qualitative Data from the PACE Implementation Study

The PACE program profiles were based on site visits in 2012 and 2013 that included interviews with program leadership and staff, review of documents related to the program, and observation of program activities. Site visit teams also had monthly monitoring calls in which they discussed study enrollment and program implementation. The research staff who conducted the site visits and maintained contact with the programs kept and organized notes documenting these visits, which were used to produce the profiles.

The research team found these notes useful in preparing for cost data collection interviews. In particular, these notes sometimes included additional detail on program components and structure that had not been included in the program profiles. For I-BEST, the person who had conducted the program's site visit remained as a research team member, so the CBA research team met with that person to review the program's structure and context.

---

[76]   Program profiles are available at http://www.acf.hhs.gov/opre/research/project/pathways-for-advancing-careers-and-education.

I.5.3 Cost Data Collection Interviews and I-BEST Program Expenditure Reports

In 2015, we conducted interviews with I-BEST staff to gather information on I-BEST program operations and costs. Program directors and financial officers participated in a one- to two-hour phone interview. The research team used this interview and subsequent follow-ups to determine a comprehensive list of program inputs and associated costs. The program also provided annual, audited expenditure reports to estimate the cost of the program.

I.5.4 SBCTC enrollment records

Washington's State Board for Community and Technical Colleges (SBCTC) provided student-level records to the PACE project. These records were used to estimate education and training other than I-BEST (including by estimating the portion of all enrollment that was I-BEST training).

I.5.5 DCPD and Other IPEDS data

The Delta Cost Project Database (DCPD) is a longitudinal database derived from the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS). The CBA calculated five-year averages of the expenditure, revenue, enrollment, and student aid variables used in the analysis, from the 2011-2015 school years using DCPD variable definitions applied to IPEDS data. IPEDS data added additional information to determine key institution characteristics (such as type of institution and undergraduate share). The database includes information on more than 6,000 public, private not-for-profit, and private for-profit institutions, including revenues, sources of revenues, spending, and total operating expenditures (Hurlburt et al. 2017).

I.5.6 Follow-up Surveys

The CBA used follow-up survey data (from 18-month and three-year surveys) to estimate unit and develop unit cost estimates using IPEDS data and to inform estimates of fringe benefits and public assistance.

I.5.7 NDNH Wage Data

Derived from state Unemployment Insurance records, the National Directory of New Hires contains quarterly employment/earnings data for all covered workers. The CBA used NDNH data for the 16 quarters after random assignment—the longest time period used in the Chapter 4 earnings impact analyses.

I.5.8 TAXSIM Simulation Output

Internet TAXSIM (v27) is a cloud-based program that "calculates federal and state income tax liabilities from typical survey data" available and documented via the National Bureau of Economic Research website (Feenberg and Coutts 1993). Internet TAXSIM takes as input a data file of taxpayer profiles, including tax year, state, marital status, number of dependents of various ages, and wage and salary income. For each submitted profile, TAXSIM returns a listing of the applicable federal and state tax liabilities, including both income and payroll taxes. The CBA uses the TAXSIM analysis to estimate the tax implications of earnings impacts.

Appendix References

Barnow, B. S., and D. Greenberg. 2015. "Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments." *Evaluation Review* 39 (2): 179-228.

Barnow, B. S., and D. H. Greenberg. 2019. "Special Issue Editors' Essay." *Evaluation Review*. https://doi.org/10.1177/0193841X19865076.

Beicht, Ursula, and Michael Friedrich. 2008. "Anlage und Methode der BIBB-Übergangsstudie." In *Ausbildungschancen und Verbleib von Schulabsolventen,* edited by Ursula Beicht, Michael Friedrich, and Joachim Gerd Ulrich, 79-99. Bielefeld, Germany: W. Bertelsmann.

Betz, N. E., and K. M. Taylor. 2001. *Manual for the Career Decision Self-Efficacy Scale and CDMSE—Short Form*. Columbus, OH: The Ohio State University.

Bühlmann, P., and S. van de Geer. 2011. *Statistics for High-Dimensional Data.* Berlin, Heidelberg, Germany: Springer.

Ciolino, Jody D., Hannah L. Palac, Amy Yang, Mireya Vaca, and Hayley M. Belli. 2019. "Ideal vs. Real: A Systematic Review on Handling Covariates in Randomized Controlled Trials." *BMC Med Res Methodol* 19: 136. Doi:10.1186/s12874-019-0787-8

Cohen, S., R. Kamarck, and R. Mermelstein. 1983. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24 (4): 385-396.

Cutrona, C. E., and D. W. Russell. 1987. "The Provisions of Social Relationships and Adaptation to Stress." *Advances in Personal Relationships* 1 (1): 37-67.

Dastrup, Samuel, Kimberly Burnett, and Larry Buron. 2017. *Career Pathways Intermediate Outcomes Study: Plan for Cost-Benefit Analyses*. OPRE Report 2017-68. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Career Pathways Intermediate Outcomes Study: Plan for Cost-Benefit Analyses | The Administration for Children and Families (hhs.gov).

Deke, J., and H. Chiang. 2017. "The WWC Attrition Standard: Sensitivity to Assumption and Opportunities for Refining and Adapting to New Contexts. *Evaluation Review* 41: 130-154. https://journals.sagepub.com/doi/10.1177/0193841X16670047.

Deville, J. C., and C. E. Särndal. 1992. "Calibration Estimation in Survey Sampling." *Journal of the American Statistical Association* 87: 376-382. https://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475217.

Duckworth, Angela L., C. Peterson, M. D. Matthews, and D. R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6): 1087-1101. https://psycnet.apa.org/record/2007-07951-009.

Dundar, A., and D. Shapiro. 2016. *The National Student Clearinghouse as an Integral Part of the National Postsecondary Data Infrastructure.* Retrieved from the National Student Clearinghouse Research Center website: https://nscresearchcenter.org/wp-content/uploads/NSC-as-an-Integral-Part-of-the-National-Postsecondary-Data-Infrastructure.pdf.

Dynarski, S. M., S. W. Hemelt, and J. M. Hyman. 2015. "The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes." *Educational Evaluation and Policy Analysis* 37(1s): 53S–79S. https://journals.sagepub.com/doi/pdf/10.3102/0162373715576078.

Farrell, Mary, and Karin Martinson. 2017. *Pathways for Advancing Careers and Education (PACE). The San Diego County Bridge to Employment in the Healthcare Industry Program: Implementation and Early Impact Report*. OPRE Report -2017-41. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/san-diego-county-bridge-employment-healthcare-industry-program-implementation-and-early.

Fein, David J. 2012. *Career Pathways as a Framework for Program Design and Evaluation: A Working Paper from the Pathways for Advancing Careers and Education (PACE) Project.* OPRE Report 2012-30. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/career-pathways-as-a-framework-for-program-design-and-evaluation-a-working.

Folsom, R. E. 1991. "Exponential and Logistics Weight Adjustments for Sampling and Nonresponse Error Reduction. In *Proceedings of the American Statistical Association, Social Statistics Section*, 197-202. Alexandria, VA: American Statistical Association.

Folsom, R. E., and A. C. Singh. 2000. "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Post-Stratification." In *Proceedings of the American Statistical Association, Survey Research Methods Section*, 598-603. Alexandria, VA: American Statistical Association.

Folsom, R. E., and M. Witt. 1994. "Testing a New Attrition Nonresponse Adjustment Method for SIPP." In *Proceedings of the American Statistical Association, Section on Survey Research Methods,* 428-433. Alexandria, VA: American Statistical Association.

Glosser, Asaph, Jill Hamadyk, Karen Gardiner, and Mike Fishman. 2014. Pathways for Advancing Careers and Education Career Pathways Program Profile: Washington's Integrated Basic Education and Skills Training (I-BEST) Program in Three Colleges. OPRE Report 2014-38, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/pace-career-pathways-program-profile-integrated-basic-education-and-skills-training-i.

Glosser, Asaph, Karin Martinson, Sung-Woo Cho, and Karen Gardiner. 2018. *Washington State's Integrated Basic Education and Skills Training (I-BEST) Program in Three Colleges: Implementation and Early Impact Report*. OPRE Report 2018-87. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/washington-states-integrated-basic-education-and-skills-training-i-best-program-three.

Goldrick-Rab, S., and K. Sorensen. 2010. "Unmarried Parents in College." *Future of Children* 20 (2): 179-203.

Hendra, Richard, and Aaron Hill. 2018. "Rethinking Response Rates: New Evidence of Little Relationship between Survey Response Rates and Nonresponse Bias." *Evaluation Review*. https://doi.org/10.1177/0193841X18807719.

Holland, Paul W. (1986). "Statistics and Causal Inference". *J. Amer. Statist. Assoc.* 81 (396): 945–960. doi:10.1080/01621459.1986.10478354

Hoven, M. R. 2012. *Investigating the Relationship between Perceived Social Support and Parent Self-Efficacy in Parents of Preschool-Aged Children*. Master's Thesis. University of British Columbia. https://circle.ubc.ca/bitstream/handle/2429/43343/ubc_2012_fall_hoven_michaelyn.pdf?sequence=3. Last accessed 8/28/2015.

Hurlburt, Steve, Audrey Peek, and Jie Sun. 2017. *Delta Cost Project Database 1987-2015: Data File Documentation.* Delta Cost Project at American Institutes for Research. https://deltacostproject.org/sites/default/files/database/DCP_Database_Documentation_1987-2015.pdf.

Izrael, David, David C. Hoaglin, and Michael P. Battaglia. 2000. "A SAS Macro for Balancing a Weighted Sample." In *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference,* Paper 275. Cary, NC: SAS Users Group International. https://pdfs.semanticscholar.org/f777/e121632ccc23bc2332efa8d1d2b4a5a311d3.pdf.

Judge, T. A. 2009. "Core Self-Evaluations and Work Success." *Current Directions in Psychological Science* 18 (1): 58-62.

Judge, Timothy, and Joyce E. Bono. 2001. "Relationship of Core Self-Evaluation Traits—Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability–with Job Satisfaction and Job Performance: A Meta-Analysis." *Journal of Applied Psychology* 86 (1): 80-92.

Judge, Timothy, Edwin A. Locke, and Cathy C. Durham. 1997. "The Dispositional Causes of Job Satisfaction: A Core Evaluations Approach." *Research in Organizational Behavior* 19: 151-188.

Judge, Timothy, Edwin A. Locke, and Cathy C. Durham. 1998. "Dispositional Effects on Job and Life Satisfaction: The Role of Core Evaluations." *Journal of Applied Psychology* 83 (1): 17-34.

Judkins, D., D. Morganstein, P. Zador, A. Piesse, B. Barrett, and P. Mukhopadhyay. 2007. "Variable Selection and Raking in Propensity Scoring." *Statistics in Medicine* 26: 1022-1033. https://onlinelibrary.wiley.com/doi/10.1002/sim.2591.

Judkins, David, and Kristin Porter. 2016. "Robustness of Ordinary Least Squares in Randomized Clinical Trials." *Statistics in Medicine* 35 (11): 1763-1773. https://www.statisticsviews.com/details/journalArticle/9169971/Robustness-of-ordinary-least-squares-in-randomized-clinical-trials.html.

Judkins, David, David Fein, and Larry Buron. 2018. *Analysis Plan for the PACE Intermediate (Three-Year) Follow-up Study*. OPRE Report 2018-95. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/opre/pace_three_yearanalysisplan_mainreport_508.pdf.

Judkins, David. 2019. "Covariate Selection in Small Randomized Studies." Presentation at the Joint Statistical Meetings, Denver, Colorado. https://ww2.amstat.org/meetings/jsm/2019/onlineprogram/AbstractDetails.cfm?abstractid=307372

Judkins, David, Daniel Litwok, and Karen Gardiner. 2020. *Pima Community College's Pathways to Healthcare Program: Appendices for Three-Year Impact Report,* OPRE Report 2020-43. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Pima Community College's Pathways to Healthcare Program: Three-Year Impact Report | The Administration for Children and Families (hhs.gov)

Kessler, R. C., G. Andrews, D. Mrocek, B. Ustun, and H. U. Wittchen. 1998. *"*The World Health Organization Composite International Diagnostic Interview Short-form (CIDI-SF)." *International Journal of Methods in Psychiatric Research* 7 (4): 171-185. https://onlinelibrary.wiley.com/doi/abs/10.1002/mpr.47.

Koch, Gary G., Catherine M. Tangen, Jin-Whan Jung, and Ingrid A. Amara. 1998. "Issues for Covariance Analysis of Dichotomous and Ordered Categorical Data from Randomized Clinical Trials and Non-parametric Strategies for Addressing Them." *Statistics in Medicine* 17: 1863-1892. https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-0258(19980815/30)17:15/16%3C1863::AID-SIM989%3E3.0.CO%3B2-M.

Kornfeld, R., and H. Bloom. 1999. "Measuring Program Impact on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17 (1): 168-197. https://www.journals.uchicago.edu/doi/pdfplus/10.1086/209917.

Le, H., A. Casillas, S. Robbins, and R. Langley. 2005. "Motivational and Skills, Social, and Self-Management Predictors of College Outcomes: Constructing the Student Readiness Inventory." *Educational and Psychological Measurement* 65 (3): 482-508. https://www.academia.edu/527739/Motivational_and_skills_social_and_self-management_predictors_of_college_outcomes_Constructing_the_Student_Readiness_Inventory.

Lin, W. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7: 295-318. https://projecteuclid.org/download/pdfview_1/euclid.aoas/1365527200.

Lumley, T., P. Diehr, S. Emerson, and L. Chen. 2002. "The Importance of the Normality Assumption in Large Public Health Data Sets." *Annual Review of Public Health* 23: 151-169. https://www.annualreviews.org/doi/pdf/10.1146/annurev.publhealth.23.100901.140546.

Peck, Laura R., Daniel Litwok, Douglas Walton, Eleanor Harvill, and Alan Werner. 2019. *Health Profession Opportunity Grants (HPOG 1.0) Impact Study: Three-Year Impacts Report*. OPRE Report # 2019-114. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/health-profession-opportunity-grants-hpog-10-impact-study-three-year-impacts-report

Peterson, C. H., A. Casillas, and S. B. Robbins. 2006. "The Student Readiness Inventory and the Big Five: Examining Social Desirability and College Academic Performance." *Personality and Individual Difference* 41 (4): 663-673. https://isiarticles.com/bundles/Article/pre/pdf/76798.pdf.

Research Triangle Institute. 2012. *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Author.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: Wiley.

Schaberg, Kelsey, and David Greenberg. 2020. *Long-Term Effects of a Sectoral Advancement Strategy.* New York: MDRC. https://www.mdrc.org/sites/default/files/WorkAdvance_5-Year_Report-Final.pdf.

Schaberg, Kelsey, and David Greenberg. 2020. *Long-Term Effects of a Sectoral Advancement Strategy.* New York: MDRC. https://www.mdrc.org/sites/default/files/WorkAdvance_5-Year_Report-Final.pdf.

Solis, Hilda L., and John M. Galvin. 2012. "Labor force characteristics by race and ethnicity, 2011." *Bureau of Labor Statistics*. https://www.bls.gov/opub/reports/race-and-ethnicity/archive/race_ethnicity_2011.pdf

Stumpf, S. A., S. M. Colarelli, and K. Hartman. 1983. "Development of the Career Exploration Survey (CES)." *Journal of Vocational Behavior* 22 (2): 191-226. https://www.sciencedirect.com/science/article/abs/pii/0001879183900283.

Tukey, John W. 1991. "Use of Many Covariates in Clinical Trials." *International Statistical Review* 59(2):123-137. https://www.jstor.org/stable/1403439?seq=1.

Walker, Joan M. T., Andrew S. Wilkins, James R. Dallaire, Howard M. Sandler, and Kathleen V. Hoover-Dempsey. 2005. "Parental Involvement: Model Revision through Scale Development." *The Elementary School Journal* 106 (2): 85-104.

Williams, R. L., and R. E. Folsom. 1981. "Weighted Hotdeck Imputation of Medical Expenditures Based on a Record Check Subsample." In *Proceedings of the American Statistical Association, Survey Research Methods Section,* 406-411. Alexandria, VA: American Statistical Association.