Washington State's Integrated Basic Education and Skills Training (I-BEST) Program in Three Colleges: Implementation and Early Impact Report

Appendices

Pathways for Advancing Careers and Education

OPRE Report No. 2018-87

September 2018

PACE
Pathways for Advancing Careers and Education

Washington State's Integrated Basic Education and Skills Training (I-BEST) Program in Three Colleges: Implementation and Early Impact Report

Pathways for Advancing Careers and Education (PACE)

OPRE Report No. 2018-87

September 2018

Authors
Karin Martinson, Sung-Woo Cho, and Karen Gardiner, Abt Associates
Asaph Glosser, MEF Associates

Submitted to:
Nicole Constance
Federal Project Officer
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Contract No. HHSP2332007913YC

Project Director: Karen Gardiner
Abt Associates
6130 Executive Blvd.
Rockville, MD 20852

Contents

List of Exhibits

Appendix A: Baseline Characteristics and Adjustments

This appendix describes specifications for baseline covariates—including the approach to missing values in Section A.1. It then compares distributions for treatment and control group members on these measures (Section A.2). Finally, Section A.3 explains how the analyses control for these covariates in estimating impacts.

A.1. Details on Baseline Covariates

Exhibit A-1 details the specifications and data sources for baseline covariates. Item nonresponse rates on these covariates were generally low. Across all nine PACE sites, item nonresponse rates were under four 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).

The research team imputed values for missing covariates using SUDAAN/IMPUTE, a weighted hot-deck imputation procedure (Research Triangle Institute 2012). This imputation step entailed a single computer run on the combined sample from all nine PACE sites. With this process, each missing value was replaced with an observed response from a similar case. Each case with a missing value was randomly matched to a single case with a non-missing, reported value within the same stratum; the reported value was then copied over to the case where the value was missing. The strata represented a cross-classification of the following: treatment-control status, site, NSC-reported enrollment status (some or none), NSC-reported credential award (some or none), and number of months of NSC-reported enrollment.[1]

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

Variable Description	Operationalization Details	Data Source(s) Instrument & Item Number
Age	Categorical measure: Under 21 21-24 25-34 35+*	BIF: B2_dob RABIT: R_RA_Date_Assigned
Female	Binary variable 1 if female 0 if male	BIF: B7

[1] In instances where this level of matching was too restrictive because no matched case with a reported value was found, then the procedure was re-run matching only on treatment status and NSC-reported enrollment status.

Variable Description	Operationalization Details	Data Source(s) Instrument & Item Number
Race-ethnicity	Categorical measure: Any race, Hispanic Black, non-Hispanic White, non-Hispanic* Other, non-Hispanic	BIF: B9
Family structure	Categorical measure: Spouse/partner, with children Spouse/partner, without children Single, with children* Single, without children (Only biological and adopted children of randomized participant considered here. Step children, 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
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*	BIF: B23
Highest level of education completed	Categorical measure: No college* Under 1 year's college credit 1 year+ of college credit Associate's degree or above	BIF: B17
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
Academic discipline[a]	Average of 10 items (scale ranging 1-6) after reversing responses to negatively phrased items. Missing if seven or more of 10 responses blank	SAQ: S11a
Training commitment[a]	Average of 10 items (scale ranging 1-6) after reversing responses to negatively phrased items. Missing if seven or more of 10 responses blank	SAQ: S11b
Academic confidence[a]	Average of 12 items (scale ranging 1-6) after reversing responses to negatively phrased items. Missing if nine or more of 12 responses blank	SAQ: S11d
Emotional stability[a]	Average of twelve items (scale ranging 1-6) after reversing responses to negatively phrased items. Missing if nine or more of twelve responses blank	SAQ: S11e
Family income in past 12 months	Categorical measure: Less than $15,000 $15,000-29,999 $30,000+*	BIF: B27
Received food assistance (WIC/SNAP) in past 12 months	Binary variable: 1 if yes 0 if no	BIF: B26b

Variable Description	Operationalization Details	Data Source(s) Instrument & Item Number
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 if 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 otherwise	SAQ: S8, S9
Current work hours	Categorical measure: 0-19* 20-34 35+	BIF: B24
Expected work hours in next few months	Categorical measure for covariate: 0-19* 20-34 35+	SAQ: S2
Expecting to attend school part-time if accepted	Binary variable: 1 if yes 0 if no	SAQ: S1
Frequency of situations interfering with school, work, job search, or family responsibilities	Average of six items of frequency of problems (scale ranging 1-5). Missing if four or more of six responses blank	SAQ: S15
Stress[b]	Average of four items (scale ranging 1-5) after reversing responses to negatively phrased items. Missing if three or more of four responses blank	SAQ: S14

Data source abbreviations: RABIT (Random Assignment and Baseline Information Tool), BIF (Basic Information Form), SAQ (Self-Administered Questionnaire).
* = category omitted in creating binary (dummy) variables for regression-adjustment models.
[a] 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).
[b] 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 it is said that "baseline balance" was achieved. The list expands on the characteristics in Chapter 2, Exhibit 2-2.

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

Exhibit A-2: Baseline Balance for I-BEST Program

Characteristic	All Participants	Treatment Group	Control Group	p-Value
Age (%)				.067
20 or under	22.2	23.2	21.1	
21 to 24	14.9	11.1	18.6	
25 to 34	29.8	31.4	28.1	
35 or older	33.2	34.3	32.2	
Female (%)	42.5	44.9	40.1	.231
Race/Ethnicity (%)				.346
Any race, Hispanic	26.0	28.9	23.1	
Black, Non-Hispanic	7.6	6.2	9.1	
White, Non-Hispanic	54.9	53.1	56.7	
Other, Non-Hispanic	14.1	13.4	14.8	
Family Structure (%)				.591
Neither Spouse/Partner nor Children	47.2	48.7	45.8	
No Spouse/Partner, 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/Both Parents Had Some College (%)	45.3	45.5	45.2	.955
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	
Educational Attainment (%)				.497
Less Than a High School Degree	30.7	28.2	33.1	
High School or Equivalent	40.0	42.0	38.0	
Less Than 1 Year of College	11.1	12.1	10.2	
1 or More Years of College	9.5	10.1	8.9	
Associate's 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 (%)				.551
Less than $15,000	47.3	46.5	48.1	
$15,000-$29,999	23.9	26.0	21.9	
$30,000 or More	28.8	27.6	30.0	
Family Income (mean)	22,110	23,002	21,240	.378
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

Characteristic	All Participants	Treatment Group	Control Group	p-Value
Current Work Hours (%)				.993
0	66.6	66.9	66.3	
1 to 19	8.5	8.5	8.5	
20 to 34	11.7	11.7	11.6	
35 or more	13.2	12.8	13.6	
Expected Work Hours in Next Few Months (%)				.228
0	41.1	41.4	40.8	
1 to 19	9.9	8.6	11.2	
20 to 34	32.0	35.2	28.9	
35 or more	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: Abt Associates calculations based on data from PACE Basic Information Form (BIF) and Self-Administered Questionnaire (SAQ).
NOTES: Tests for statistically significant imbalance were based on SAS/TTEST procedure. The research team did not run baseline tests for each line of a multi-category construct. Instead, the procedure called for determining imbalance within the construct generally, and not for specific categories.

A.3. Regression Adjustment

In this appendix, the team describes the regression adjustment approach used to improve precision and minimize effects of sampling error on impact point estimates.

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

$$Y_i = X_i\beta + \delta T_i + e_i \quad \text{(A.1)}$$

where Y_i is the outcome, T_i is a 0/1 dummy variable indicating treatment group membership, X_i is a row vector of baseline covariates, β is the vector of parameters indicating the influence of each covariate on the outcome, δ is the effect of treatment, and e_i is an error term. 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 covariates (Lin, 2013) even when the outcomes are not normally distributed (Judkins and Porter 2016). Estimates of the treatment effect are "asymptotically unbiased" and for adequately large sample sizes, under most conditions, $\text{var}\left(\hat{\delta}\right) \approx \left(1-R^2\right)\text{var}(\bar{y}_T - \bar{y}_C)$, where R^2 is proportion of the variance in Y_i that can be explained by X_i, in Equation A.2 below.

The team's analyses of results from simulations and the first few PACE sites to complete data collection showed that the method can perform poorly when the number of baseline covariates is relatively large compared with the number of observations. Specifically, when the ratio *n/p* is

not very large, it can happen that $\text{var}\left(\hat{\delta}\right) > \text{var}(\bar{y}_T - \bar{y}_C)$, meaning that the variance on the estimated treatment effect using the regression adjustment in Equation A.1 is actually larger than the variance of the simpler randomization-based estimate of the treatment effect, formed by simply contrasting the mean outcomes in the two groups. Unpublished simulations show that the variance penalty increases as the ratio of non-significant to significant covariates grows.[2] There is a lack of good research on how large the ratio of cases to variables needs to be in order to guarantee that $\text{var}\left(\hat{\delta}\right) < \text{var}(\bar{y}_T - \bar{y}_C)$, but it appears that values of *n/p* less than 30 may be problematic. Eight of nine of the PACE sites have values of *n/p* in this potentially problematic range even after trimming the number of baseline predictors to 34. These baseline predictors were selected based their theoretical importance, their correlations with other baseline variables—to avoid selecting multiple measures of the same underlying construct—and their ability to predict important postsecondary outcomes captured by the National Student Clearinghouse (Fein 2016).

Based on this research, the team applied a slightly different approach to estimation for this report. The approach involved first estimating the influences of the 34 baseline characteristics on the outcome under the control condition (Equation A.2 below). The next step was to calculate how different each program and control group member's outcome was from what would have been expected under control conditions, as in Equation A.3. These differences between actual and predicted outcomes are called "residuals." The team then calculated the difference between the average residual in the program group and the average residual in the control group, as in Equation A.4.

Equation A.5 gives the formula used to estimate standard errors on these impact estimates.

$$Y_i = X_i\beta + e_i \tag{A.2}$$

$$\hat{r}_i = Y_i - X_i\hat{\beta} \tag{A.3}$$

$$\hat{\delta} = \hat{\mu}_T - \hat{\mu}_C = \frac{\sum_i T_i\hat{r}_i}{\sum_i T_i} - \frac{\sum_i (1-T_i)\hat{r}_i}{\sum_i (1-T_i)} \tag{A.4}$$

[2] For example, with a sample size of 1,000, when there are three covariates that explain 57 percent of the variation in the outcome and 97 covariates are uncorrelated with the outcome—and thus explain none of its variation—the standard error of the estimated impact is 11 percent higher with OLS than with Koch's method (Austin Nichols, Abt Associates, unpublished simulations, 2016). The standard error for each method was estimated by taking the standard deviation of the impact estimates across the simulated samples.

$$se(\hat{\delta}) = \sqrt{\frac{\sum_i T_i \left(\hat{r}_i - \hat{\mu}_T\right)^2}{\sum_i T_i - 1} + \frac{\sum_i \left(1-T_i\right)\left(\hat{r}_i - \hat{\mu}_C\right)^2}{\sum_i \left(1-T_i\right) - 1}} \quad \text{(A.5)}$$

For survey-based outcomes subject to nonresponse, the team used a weighted version of this estimator, as in Equation B.6.

$$\hat{\delta} = \frac{\sum_i w_i T_i \hat{r}_i}{\sum_i w_i T_i} - \frac{\sum_i w_i \left(1-T_i\right)\hat{r}_i}{\sum_i w_i \left(1-T_i\right)} \quad \text{(A.6)}$$

where w_i is the nonresponse-adjustment weight for survey-reported outcomes.

This method is similar to the method developed by Koch and coauthors (1998), who referred to it as nonparametric ANCOVA. Since then, most authors have referred to it as Koch's estimator. The difference between Koch's estimator and the method applied in this report is that Koch and coauthors fit Equation A.2 on the entire sample rather than just on the control sample. The main advantage of fitting Equation A.2 just on the control sample is that the parameters are more easily interpretable when the null hypothesis is rejected. A secondary advantage is that, as Lesaffre and Senn (2003) demonstrated, Koch's estimator can produce overly liberal significance tests, meaning that the null hypothesis of no program effect is rejected too often. This occurs because the estimated standard errors on the estimated treatment effect using Koch's method are too small. Our goal in fitting Equation A.2 to the control sample instead of the entire sample was to correct for the Koch method's tendency to underestimate the standard error while still producing more precise impact estimates than OLS.

In the eight PACE sites other than I-BEST, analysis confirmed that use of the modified Koch's estimator improved precision for the confirmatory outcome relative to OLS (see Equation A.1). However, for the evaluation of the I-BEST program, the modified Koch's estimator increased the variance on the estimate of the impact of the program on the confirmatory outcome (academic and workforce credits earned through 24 months) by 12 percent compared to what it would have been with the OLS approach. The variance on the modified Koch's estimator for the impact on the confirmatory outcome was still two percent smaller than what would be obtained from pure randomization, but this was still a disappointing result for the method. Across a wider of collection of confirmatory and secondary outcomes, the modified Koch's estimator sometimes decreased the variance substantially, but there was practically no net difference in average variances between OLS and the modified Koch's estimator for the set of confirmatory and secondary hypothesis tests at I-BEST. The team decided to report estimates based on the modified Koch's estimator, despite the lack of variance reduction, since changes

to the analysis after the analysis plan was published could be misconstrued as an attempt to alter the study findings.

Exhibit A-3 shows the regression coefficients from Equation A.2 for the confirmatory outcome, the number of academic and workforce credits earned through 24 months at colleges. These covariates were selected based on a pooled analysis across all nine PACE sites of factors that predict various measures of success reported to the National Student Clearinghouse. Note that of the 34 baseline covariates allowed into the model, only three of these (highlighted in red) are predictive of future credits for the control group sample. Specifically, being under 21 years of age and having parents who attended college themselves are positively associated with future credits while current work hours at the time of randomization between 20 and 34 is negatively associated with future credits. Nonetheless, all 34 variables were retained in the model fit for A.2.

Exhibit A-3: Coefficients for Baseline Characteristics as Predictors of Academic and Workforce Credits Earned at Colleges through 24 Months: I-BEST Control Group Members

Baseline Covariate	Estimate	Standard Error	*p*-Value
Intercept	0.60	0.53	.259
Age			
20 or under	0.26	0.15	.080
21 to 24	0.00	0.12	.980
25 to 34	-0.07	0.09	.404
35 or older	0	0	na
Sex			
Female	-0.13	0.11	.276
Male	0	0	na
Race/Ethnicity			
Hispanic	0.19	0.12	.126
Black, Non-Hispanic	0.00	0.09	.959
White, Non-Hispanic	0	0	na
Other, Non-Hispanic	0.02	0.11	.889
Living Arrangements			
Neither spouse/partner or children	-0.11	0.10	.253
No spouse/partner, living with children	-0.13	0.12	.282
Spouse/partner, no children	0	0	na
Spouse/partner and children	0.10	0.09	.287
Living with Parents	-0.10	0.10	.316
One/Both Parent(s) Had Some College	0.19	0.08	.016
High School Grades			
Mostly Got A's	0.04	0.12	.747
Mostly Got B's	-0.10	0.08	.252
Mostly got C's or Below	0	0	na

Baseline Covariate	Estimate	Standard Error	*p*-Value
Current Education			
High School Degree or Less	0	0	na
Less Than 1 Year of College	-0.12	0.10	.231
1 or More Years of College	-0.06	0.11	.591
Associate's Degree or Higher	0.03	0.12	.788
Career Knowledge Index	-0.03	0.11	.816
Family Income			
Less than $15,000	0.16	0.12	.202
$15,000-$29,999	0.20	0.13	.135
$30,000 or More	0	0	na
Psycho-Social Indices			
Academic Discipline Index	0.04	0.08	.649
Training Commitment Index	0.05	0.09	.554
Academic Self-Confidence Index	-0.02	0.07	.789
Emotional Stability Index	-0.06	0.08	.466
Stress Index	-0.06	0.06	.316
Life Challenges Index	-0.06	0.07	.418
Public Assistance/Hardship Past 12 Months			
Received WIC or SNAP	0.03	0.10	.728
Received Public Assistance or Welfare	-0.05	0.08	.578
Reported Financial Hardship	-0.05	0.08	.497
Current Work Hours			
0 to 19	0	0	na
20 to 34	-0.17	0.10	.093
35 or more	0.10	0.17	.550
Expected Work Hours in Next Few Months			
0 to 19	0	0	na
20 to 34	0.04	0.09	.661
35 or more	-0.09	0.08	.298
Plan to attend school only part-time if admitted to the I-BEST program	0.05	0.09	.571

SOURCE: Abt Associates calculations based on data from SBCTC records, the PACE Basic Information Form (BIF), and the PACE Self-Administered Questionnaire (SAQ).
NOTES: Model estimated with SAS/SURVEYREG procedure. Sample size=249.

Exhibit A-4 shows impacts on selected confirmatory and secondary outcomes before and after regression adjustment. The similarity in estimates further attests to the high level of balance achieved through random assignment. There are no changes in the characterization of the strength of evidence against the null hypothesis (i.e., the number of stars attached to estimates did not change). Because of this good balance, the adjustment had little impact on standard errors. All the standard errors are either modestly smaller or unchanged with adjustment.

Exhibit A-4: Comparison of Selected Impact Estimates with and without Adjustment for Baseline Imbalances

Outcome	Unadjusted Est (StdErr)	Adjusted Est(StdErr)
Academic and workforce credits earned through month:		
6	5.54***	5.65***
	(0.52)	(0.52)
12	9.44***	9.73***
	(1.16)	(1.14)
18	10.71***	11.21***
	(1.69)	(1.67)
24	12.51***	13.12***
	(2.25)	(2.23)
Earned college credential through 24 months: (%)		
Workforce completion	3.51**	3.75**
	(1.96)	(1.92)
Workforce award	33.36***	33.34***
	(3.17)	(3.10)
Associate's degree or higher	0.00	0.00
	(0.45)	(0.45)
Any	32.10***	32.13***
	(3.35)	(3.27)
Any enrollment through 24 months in: (%)		
Academic courses	7.07**	6.14**
	(3.61)	(3.52)
Vocational courses	39.81***	40.91***
	(3.58)	(3.57)
Any courses (including basic and developmental)	20.53***	21.92***
	(3.20)	(3.16)
Earned credential at 18 months: (%)		
From a college	30.51***	30.75***
	(3.24)	(3.15)
Another education and training institution	-3.78	-3.41
	(1.44)	(1.44)
A licensing/certification body	13.91***	15.52***
	(3.99)	(3.86)
Any issuer	28.02***	29.74***
	(4.38)	(4.22)

SOURCE: Abt Associates calculations based on data from the administrative records and the PACE short-term follow-up survey.
NOTES: Standard errors on estimated impacts shown in parentheses. Adjusted impact estimates and associated standard errors were prepared with the modified Koch's estimator, as defined in Equations (A.4) and (A.5). Statistical significance levels, based on one-tailed *t*-tests tests of differences between research groups, are summarized as follows: *** statistically significant at the one percent level; ** at the five percent level; * at the 10 percent level.

Appendix B: College Records Data

The research team used records from the Washington State Board for Community and Technical Colleges (SBCTC) to measure college outcomes in the study. Such administrative data offer strong advantages over survey data—notably in avoiding loss of sample to nonresponse and any effects on data quality from survey response errors. This appendix has two major sections. The first concerns course classification in SBCTC records. The section concerns imputation of experiences at colleges not covered by the SBCTC records.

B.1. Course Classification in SBCTC Records

The research team received course enrollment records from SBCTC, including all course enrollments for the study sample between 2010 and 2017. The records included the academic year and quarter of enrollment, grade and passing status, number of credits earned, course title, department, institutional intent, and a six-digit classification code for the course from the Classification of Instructional Programs (CIP).[3]

The team selected courses taken in quarters that ended within 24 months after random assignment.[4] The schools included in the study all use a quarter system with four 10-week quarters per academic year, so the follow-up period includes eight consecutive quarters of potential enrollment. Programs offered at the study schools require a relatively high number of credits; most certificate programs carry requirements of 50 to 90 credits across three quarters and workforce degrees carry requirements of more than 120 credits across six to eight quarters. There are two types of enrollment measures the team analyzed: percent enrolled (includes any course enrollment, regardless of whether the student passed the course); and average number of credits earned (only includes courses that were passed).

In addition to overall course enrollment and credits earned, the team also analyzed enrollment within three types of courses: academic, workforce, or basic/developmental. The SBCTC data do not directly categorize courses this way, so the team created the categories based on two related variables: institutional intent and CIP code. Each course has an assigned institutional intent code that is specific to SBCTC administrative data, and is used to identify the primary intent of the class from the perspective of the college and not the student. For example, if a course is intended as an academic course but a student enrolls in the course as an elective, the

[3] The CIP is a taxonomy that organizes courses according to the field of study. The system was created by the U.S. Department of Education's National Center for Education Statistics (NCES) to provide a standardized coding scheme across colleges.

[4] Random assignment was conducted from November 2011 through September 2014.

course is still identified as academic in regards to institutional intent. There are five main categories of institutional intent: academic, vocational, basic skills, developmental, and community service (the last of which do not count toward any degree or credential).

Some of the courses with a vocational or academic institutional intent are actually remedial courses, and SBCTC data documentation included guidance on recoding these courses as developmental or basic skills courses based on their CIP code.[5] The SBCTC administrative data includes the six-digit CIP code for each course, in additional to the institutional intent code.[6] Where the CIP code reflected a developmental course (regardless of the institutional intent), the course was recoded as basic skills (CIP code beginning with 32) or developmental (CIP code beginning with 33).

Using a combination of institutional intent and CIP codes in the SBCTC data, the team categorized every course as academic, workforce, basic, or developmental, and analyzed course enrollment and average credits earned within each category. However, for analysis purposes, enrollment in basic skills and developmental classes were collapsed, and only credits from vocational and academic courses were included in total credit counts. Additional analyses looked only at reading and writing courses, and these were identified using CIP codes only. Specifically, courses with CIP codes 330101, 330102, and 330103 were identified as developmental reading, writing, and math courses, respectively. Courses with CIP codes between 230000 and 240000 were identified as college-level reading and writing courses, and courses with CIP codes between 270000 and 280000 were identified as college-level math courses.

B.2. Imputation of Credits, Credential and Detailed Enrollment at Colleges Outside of SBCTC

An important concern in considering whether to use SBCTC data for this evaluation was whether SBCTC data could support sufficiently broad measures of college enrollment given that some sample members also attended colleges not covered by the state board (four-year colleges, private colleges, and out-of-state colleges). To assess the extent of enrollment at other colleges, the team matched the sample to college records maintained in the NSC. Covering 96 percent of college enrollments nationwide, the NSC data provide an excellent frame for these purposes. As shown in Exhibit B-1, the vast majority of sample members enrolling in college

[5] Accessed online August 9, 2017 here (see definition of Instit_Intent_Recat): https://www.sbctc.edu/resources/documents/colleges-staff/data-services/data-warehouse/class.docx

[6] It should be noted that SBCTC's classification of CIP codes varies slightly from the standard scheme, particularly in the area of developmental and basic skills courses. The team relied on SBCTC's crosswalk of CIP codes, accessed online August 8, 2017 at: https://www.sbctc.edu/resources/documents/colleges-staff/data-services/data-warehouse/CIP2010EPCCrosswalk.pdf

(over the period from randomization to November 2015) enrolled only at colleges covered by the SBCTC and few enrolled only at other colleges. Rates of enrollment at other colleges are slightly higher in the control group.

Exhibit B-1: National Student Clearinghouse–Reported Enrollment at Colleges Covered by the SBCTC and Other Colleges for the Integrated Basic Education Skills Training Evaluation Sample, by Study Group

Group	Enrollment Documented in National Student Clearinghouse				
	Enrolled at SBCTC College			Enrolled Only at Colleges Outside the SBCTC (%)	Total Ever Enrolled (%)
	At SBCTC College (%)	Only at SBCTC College (%)	At SBCTC and Other College (%)		
Treatment	99.23	97.30	1.93	0.77	100.0
Control	95.18	90.96	4.22	4.82	100.0

SOURCE: Abt Associates calculations based on data from the National Student Clearinghouse (NSC).
NOTE: NSC data cover the period from randomization (for the person) to November 2015, a period that varies in length from 23 months to 48 months, depending on how early the student was randomized.

Based on these results, the team devised an imputation approach that used NSC data to adjust SBCTC data for enrollment spells[7] at other colleges.[8] The goal of imputation was to fill in the missing earned credits, and earned credentials for every spell that the NSC showed the student spent at another college by pairing every other-college spell with a similar SBCTC spell, and then copying the information over.

The imputation strategy involved several steps:

1. Find a SBCTC record for as many NSC-reported SBCTC spells as possible. This step filled in instructional hours, earned credits, and earned credentials for most NSC-reported SBCTC spells. The team referred to this step as the *exact matching* process because there is a single correct match in the SBCTC system for almost all the NSC-reported SBCTC spells.
2. Resolve NSC-reported spells that did not match to a SBCTC record. For such NSC-reported spells, the team assumed that no instructional hours were received, no credits were earned, and no credentials were earned.

[7] A spell was defined as a period of enrollment with no gaps longer than three months unless the gap included one of the summer months (June, July, and August), in which case a gap of seven months was required to initiate a new spell.

[8] Though useful for this purpose, the NSC data were not in themselves sufficient because they exclude some key study outcomes (e.g., academic and workforce credits) and do not completely cover others (e.g., certificates and degrees).

3. Summarize the available data for each NSC-reported spell and the student to whom the spell belonged. The team summarized these data by developing statistical models that predict four critical SBCTC–reported outcomes for each spell.[9]
4. Match each NSC-reported other-college spell with a "similar" NSC-reported SBCTC spell in terms of the predicted four critical outcomes. The team referred to this step as *statistical matching* because there are many possible NSC-reported SBCTC spells that could be matched to every NSC-reported other-college spell. The team only matched other-college spells of students in the treatment group to SBCTC spells or other students in the treatment group. A parallel restriction was placed on the matching of other-college spells of students in the control group. The team imposed these restrictions to avoid "washing out" any effects by making control experiences artificially more similar to treatment experiences.
5. Lastly, copy the information from steps 1 and 2 that were associated with every NSC-reported SBCTC spell over to the statistically matched NSC-reported other-college spell.

The following sections give more information for each step.

Details on Step 1 (Exact Matching)

A total of 564 spells at SBCTC colleges were found in NSC. The team conducted the exact matching of each of these NSC-reported spells at SBCTC 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 SBCTC, 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 SBCTC, 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, then the SBCTC–reported spell was broken into pieces that better matched the NSC-reported spells. The team then transcribed the training hours, credits, and credentials associated with the SBCTC–reported spell in the SBCTC record system over to the NSC-reported spell.

Details on Step 2 (Unmatched NSC-Reported SBCTC Records)

A total of 17 of the 564 NSC-reported spells at SBCTC did not overlap with any SBCTC–reported spells. Lacking any further information on these spells, and following procedures established for other PACE reports, the team assumed that the student never showed up for any classes or

[9] Each of the four predicted outcomes is a linear function of the larger collection of available covariates, and therefore "summarizes" the larger collection.

withdrew too early to be included in reports sent by colleges to the SBCTC. Accordingly, when two of these 17 spells were selected as donors for NSC-reported spells at colleges not part of the SBCTC, the team assumed that those spells had the same zero outcomes.

Details on Step 3 (Data Summarization)

The available data about each spell that could be used included NSC-reported spell duration and timing, NSC-reported credentials awarded in connection with the spell, and self-reported baseline variables. Follow-up survey data could not be used because the team wanted to do the matching both on survey respondents and on non-respondents. To facilitate matching, the team developed statistical models for four SBCTC–reported outcomes on the set of exactly matched records in terms of these variables. The SBCTC–reported outcomes were:

- Academic credits earned within the intersection of the spell (as reported by SBCTC) with the 12-month window after randomization
- Academic credits earned within the intersection of the spell (as reported by SBCTC) with the 18-month window after randomization
- Workforce credits earned within the intersection of the spell (as reported by SBCTC) with the 12-month window after randomization
- Workforce credits earned within the intersection of the spell (as reported by SBCTC) with the 18-month window after randomization

These SBCTC–reported outcomes are important components of the confirmatory outcome and some secondary outcomes.

The procedure involved first fitting models for NSC-reported SBCTC spells and then using estimated coefficients to predict values for both SBCTC spells and spells at other colleges. The models (not shown) involved 4 to 21 characteristics from NSC and from the baseline.

Details on Step 4 (Statistical Matching)

For each spell at a college other than SBCTC, the team calculated the weighted Euclidean distance from that spell to every spell at SBCTC as:

$$D_{ij} = \sqrt{\sum_{\ell=1}^{4} c_\ell \left(\hat{z}_{\ell i} - \hat{z}_{\ell j} \right)^2} \qquad \text{(B.1)}$$

where $\hat{z}_{1i}, \ldots \hat{z}_{4i}$ are the predicted outcomes for the spell at a college other than SBCTC, $\hat{z}_{1j}, \ldots \hat{z}_{4j}$ are the predicted outcomes for a spell at SBCTC, and $c_1, \ldots, c_4$ are emphasis

weights.[10] The team selected the SBCTC spell *j* within the same study group (treatment/control) that minimized D_{ij} as the matched SBCTC spell for the *i*-th other-college spell. The same SBCTC spell could be matched to more than one other-college spell. The procedure is known in the literature as "predictive mean matching" (van Buren, 2012).

Details on Step 5 (Propagating SBCTC Values)

The final step entailed copying matched data on earned credits (by type and timing) and earned credentials (by type and timing) from SBCTC records to serve as values for a spell at the other college. The procedure involved both the exact matching and the statistical matching. The outcomes of 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. The team did this separately for every outcome based on SBCTC records, including enrollment by class type, credits by class types, and credentials within six-month anniversaries of the randomization date for the person out through 24 months.

Exhibit B-2 provides summary statistics on selected outcomes reflecting the quality of resulting imputations. Imputation involved a total of 27 NSC-reported spells at colleges outside the SBCTC. The team found SBCTC matches for each of the 27 spells from the 564 NSC-reported SBCTC spells. The matches were generally of high quality. For example, in the control group, the correlation in predicted earned workforce credits through 18 months across matched pairs was 0.98. In the treatment group, it was only slightly lower at 0.96.

Exhibit B-2: Descriptive Statistics Related to Imputation of Academic and Workforce Credits for Spells at Colleges Outside SBCTC

Statistic	Control	Treatment
Number of spells at colleges outside the SBCTC	17	10
Number of spells at college covered by the SBCTC	216	348
Correlation across matched pairs (other-college spell and SBCTC spell) in predicted:		
Academic credits through 12 months	0.98	1.00
Workforce credits through 12 months	0.86	0.86
Academic credits through 18 months	1.00	1.00
Workforce credits through 18 months	0.98	0.96

SOURCE: Abt Associates calculations based on data from the National Student Clearinghouse, SBCTC records, the PACE Basic Information Form (BIF), the PACE Self-Administered Questionnaire (SAQ), and the PACE short-term follow-up survey.

[10] The team gave slightly larger weights to the 18-month predictions.

Appendix C: Survey Data Recoding and Adjustments

This appendix documents key technical detail for impact estimates for outcomes based on 18-month follow-up survey data. 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. Section C.3 analyzes survey nonresponse and documents the decision to apply nonresponse weights in the impact analysis.

C.1. Measures Based on Follow-up Survey Data

Exhibit C-1 provides details on specifications for the process outcomes analyzed in the implementation analysis of Chapter 4.

Exhibit C-1: Details on Specifications for Survey-Based Outcomes in Chapter 4

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Receipt of Education or Training		
Entire Study Sample		
Received education or training since random assignment		
In any subject/field	Two-question format with slightly different wordings to try to get all training spells reported.	A1, A1a
In a healthcare occupation	Open-ended responses about name of target occupation and understanding of future duties were coded by staff from the U.S. Census Bureau into the U.S. Department of Labor's Standard Occupational Classification (SOC) codes. Those in programs designed to train them for jobs as Healthcare. Practitioners/Technicians (SOC 29-xxxx) or Healthcare Support Workers (SOC 31-xxxx) were counted for this outcome. This does not include office workers in the healthcare industry or personal care aides in nursing homes.	A19a, A20, A21, A27a, A27c, A27d
Since random assignment, ever attended	The team looked up place names reported in question A4 in the Integrated Postsecondary Education Data System and used the IPEDS classification to edit self-reports in question A5. Private for-profit colleges were not counted as proprietary schools. Only places not classified as degree-granting in IPEDS and that are privately run for profit were classified as proprietary schools.	A4, A5
Two-year college	Community or technical college (two-year college)	
Four-year college	Four-year college/university	
Proprietary school	Private school/company that provides training	
Adult high school/education	Adult education / adult high school / community school / night school	
Community/non-profit organization		
Other	State unemployment/employment office, One-Stop career center, your place of employment, or somewhere else.	

Of Those Who Attended Any Education or Training		
Time spent at school and work at first place attended Full-time school and full-time work Full-time school with no or part-time work Part-time school and full-time work Part-time school with no or part-time work	Question was asked about each place attended since randomization, but only information on first place was analyzed. Enrollment dates were used to determine first place attended since randomization.	A7
Views of classes at first place attended	Questions about career relevance and learning methods were asked only about first place attended. This was done to reduce respondent burden. First place was chosen rather than last place because PACE programs put particular emphasis on innovative teaching methods for basic education classes, which would typically be the first classes taken.	
Strongly agrees relevant to life/career	Strongly agrees that "These classes were relevant to my career interests" or strongly disagrees that "These classes did not relate to much of anything else in my life."	A46c, A46d
Used active learning methods most/all of the time	Responses to six-item battery were reverse scaled (1=none of the time, 4=all the time) and then averaged. Anyone with an average of 2.5 or larger was counted.	A47a-A47f
Perceived strong emphasis on community	People who responded "a great deal" were counted.	A37
Basic Skills Instruction and Tests		
Received basic skills instruction since random assignment Academic skills English as a Second Language		 A10b A10a
Took college placement exam English Math		 A57 A58
Passed college placement exam English Math		 A57a A58a
Life Skills Instruction		
Received life skills instruction since random assignment		A10e
Receipt of Various Supports		
Received assistance from any organization since random assignment (%) Career counseling Help arranging supports for school/work/family Job search or placement	This was asked of everyone, even those with no training since randomization.	A62
Cited financial support as challenge in enrollment or persistence	Reported money troubles as reason for not continuing studies, not currently studying, or never starting studies; or reported that it was very or somewhat difficult to obtain adequate financial support to continue their studies.	A11a, A14a, A23a, A26a, A35, A59, A60

Of Those Who Attended Any Education or Training		
Received supports at first place of instruction attended (%)	Question was asked about first and second places attended since randomization, but only information on first place was analyzed. Enrollment dates were used to determine first place attended since randomization.	
Career counseling		A36d
Ever		
Three or more times		
Academic advising		A36a
Ever		
Three or more times		
Financial advising		A36b
Ever		
Three or more times		
Tutoring		A36d
Ever		
Three or more times		
Help arranging supports for school or work		A36f
Ever		
Three or more times		
Job search/placement assistance		A36e
Ever		
Three or more times		
Received financial assistance at first place of instruction (%)[a]	Question was asked about each place attended since randomization, but only information on first place was analyzed. Enrollment dates were used to determine first place attended since randomization.	
Grants/scholarship	A Pell grant or other government grant or scholarship—not counting loans you have to pay back, Must indicate in question A31 that funds were used for tuition, other school-related expense, or living expenses.	A30g, A31
Loan	Loans in your own name or loans in your parents' names. Must indicate in question A31 that funds were used for tuition, other school-related expense, or living expenses.	A30e, A30f
Offered opportunities for related work experience as part of training at first place of instruction (%)	Question was asked about each place attended since randomization, but only information on first place was analyzed. Enrollment dates were used to determine first place attended since randomization.	
Clinical internship		A38b
Visits to local employer		A38c
Work-study job		A38a
Apprenticeship		A38e
Any related work experience (including other)		A38f

Chapter 5, Exhibit 5-1 provided descriptions of outcomes in the impact analysis of the I-BEST program. Exhibit C-2 provides details on the operationalization of each measure and the underlying survey questions.

Exhibit C-2: Details on Specifications for Survey-Based Outcomes in Chapter 5

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Secondary		
Education		
Credential receipt from another type of education-training institution	The survey had separate questions about credentials awarded for regular credit-bearing courses and for non-credit occupational courses. It the respondent indicated receiving either type of credential, then this variable was coded as 1 (for yes); otherwise, it was coded as 0 (for no). The survey did not ask for credentials awarded as a result of ESL, Adult Basic Education, or life-skills courses.	A22, A23, A27e, A27f
Credential receipt from a licensing/certification body	The survey asked about the highest level of occupational training completed. One of the possible answers was "a professional, state or industry certification, license or credential." If the respondent picked this level, then there was a follow-up question about the year of award. If the year of award was the same as the year of randomization or later, then the person was coded as having earned such a credential.	A56
Received a credential from any source	Combined SBCTC records on credentials from colleges with survey-reported credential receipt from other education-training institutions and from licensing/certification bodies, as defined above.	
Career Progress		
Employment and earning $12 or per hour	Analyzed response to survey question for control group. Selected 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 programs such as I-BEST.	E2
Employment in job requiring mid-level skills	Three open-ended questions about the kind of work done, the usual activities completed, and the job title were coded into one of the U.S. Department of Labor's Standard Occupational Classification (SOC) codes. The team then looked up the Job Zone[11] for each SOC code in the Bureau of Labor Standards O*NET system.[12] 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 of 3—occupations that need medium preparation—seemed a reasonable goal for all programs involved in PACE. This Job Zone 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."	E3, E4, E5
Perceived career progress	This was a new scale created for PACE. 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 an individual's sense of progress in a career pathways program as described in Fein (2012).	C5, C6

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

[12] https://www.onetonline.org/ [last accessed September 12, 2016]

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
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). Items d-f were adapted from the Career Exploration Survey. Two items (c and g) were new and written specifically for the PACE BIF. Response categories range from 1=strongly disagree to 4=strongly agree.	C3
Access to career supports	This was a new scale created for PACE. It is a six-item scale, counting number of types of career-supportive relationships in workforce and education settings; response categories range from 1= no to 2=yes. 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 Sorenson 2010).	C2
Exploratory		
Psycho-Social Skills		
Grit	Existing scale from Duckworth et al. (2007). The eight-item scale captures persistence and determination; response categories range from 1=strongly disagree to 4=strongly agree.	B3
Academic self-confidence	Existing scale from Le et al. (2005). This scale was used for a second time in the follow-up survey. It was used initially in the BIF. The 12-item scale includes response categories that range from 1=strongly disagree to 6=strongly agree.	B4
Core self-evaluation	Existing scale from Judge (2009). The 12-item scale response categories range from 1=strongly disagree to 4=strongly agree.	B6
Social belonging in school	Shorter version of an existing scale by Walton and Cohen (2007, 2011). The five-item scale captured sense of belonging; response categories range from 1=strongly disagree to 4=strongly agree.	B7
Life Stressors		
Financial hardship	This was a new scale created for PACE. This scale was used for a second time in the follow-up survey. It was used initially in the BIF. The two-item scale asked about financial hardship, reported as either an inability to pay rent/mortgage or not enough money to make ends meet. Response categories were 0=no or 1=yes.	D1, D2
Life challenges	This was a new scale created for PACE. It was adapted from a longer instrument by Kessler et al. (1998). This scale was used for a second time in the follow-up survey. It was used initially in the BIF. The seven-item scale captured life challenges that interfered with school, work, or family responsibilities. The response categories range from 1=never to 5=very often.	D3
Perceived stress	Existing scale from Cohen et al. (1983). This scale was used for a second time in the follow-up survey. It was used initially in the BIF. The four-item scale captured perceived stress. The response categories range from 1=never to 4=very often.	D4

C.2. Imputation of Some Item Nonresponse in the Follow-up Survey

This section documents the research team's response to two sources of missing data affecting survey outcomes. First, initial data quality assessment revealed that a small fraction of respondents who initially indicated receiving some education and training did not answer subsequent questions on the nature of these experiences. Second, all outcomes were affected

by at least some missing data where respondents either declined to answer a question or gave an answer of "don't know."

Concerning the first issue, the discrepancy affected fewer than eight percent of respondents and occurred at a higher rate for control (13 percent) group than for treatment (four percent) group members. Specifically, the missing data involved responses to a filter question (A10) ascertaining participation in each of a series of types of education and training activities (ESL, Adult Basic Education, courses for college credit, non-credit occupational training, life-skills classes). Checks against SBCTC confirmed education and training receipt for 16 of the 24 people with inconsistent responses on A1 (any school) and A10 (any classes by type). This suggested misunderstanding survey questions as a likely source of the missing data.

To adjust for these missing data, the team imputed new responses for A10 using a custom-written nearest neighbor hot deck procedure along the lines suggested by Andridge and Little (2010). The hot deck involves "binning" and sorting. Within a bin, the procedure matches each case that is missing an outcome to the nearest complete case with respect to the sort. This hot deck imputation procedure matched spells with consistent responses to A10 (*consistent spells*) to spells with inconsistent responses to A10 (*inconsistent spells*). The team used site and treatment status to define the bins and the modeled propensity of a spell being consistent to define the sorting variable. To model the propensity that a spell would be consistent, the team searched a large potential set of predictor variables from baseline variables and from sections of the follow-up survey for which A10 was not a filter question. The team included interactions as well as main effects. The team conducted this search and fit the final model on a pooled dataset including observations from SBCTC, as well five other PACE sites to increase power.[13] The final imputation model used 24 variables and interactions from the survey.

In the course of imputing A10, the team kept track of the ID of the consistent spell that was matched to each inconsistent spell. After imputation of A10 was complete, the team then filled in responses to the detailed questions (A11-A29) filtered by A10 by copying the responses for the consistent spell that had been matched to the inconsistent spell.[14]

In response to the second issue—the common problem of small fractions missing on most questions due to refusals and don't knows—the team for the most part simply omitted those respondents from the relevant analyses. This was done separately for each outcome, meaning that the maximum number of usable responses was used for estimating the impact of each

[13] Data collection was completed at three sites sooner than at the other six. Processing was kept separate for the two batches.

[14] If A10e was answered "no" or was not answered, then items A49-A51 were skipped. The team decided not to impute values for these items in the cases where A10e was imputed to have a value of "yes," as A49-A51 do not provide important outcomes for PACE impact analyses.

outcome. For training hours, however, the team imputed responses for each type of course at each school the respondent attended. This imputation allowed the team to sum training hours across schools and types of courses without having high missing data rates on the sums because of scattered item missingness. To carry out this imputation, the team used SUDAAN/IMPUTE, as discussed in Section A.1 for missingness of baseline covariates. This random matching was constrained to occur within strata defined by treatment status, site, type of training, and self-reported completion status of the spell.

C.3. Survey Nonresponse Analysis

The 18-month follow-up survey obtained a lower response rate in the control group (71 percent) than in the treatment group (76 percent). This gap raises concerns about the possibility of differential nonresponse bias. Further heightening this concern, survey response rates were significantly higher for those with any SBCTC-reported enrollment (75 versus 66 percent), indicating that school engagement boosts willingness to respond to the survey. This section assesses the implications of nonresponse for the study's impact findings.

Exhibit C-3 compares distributions on baseline characteristics for all sample members and survey respondents. Nonresponse increased the number of significant imbalances across the two arms (using a threshold of .10 for statistical significance) from three to four statistically significant differences. (Significant differences are in red italic font.)

Exhibit C-4 compares regression-adjusted impacts on college outcomes from administrative records for the full and respondent samples. Ignoring for the moment the weighted column on survey respondents, impact estimates on credits and credentials are consistently larger when based on unweighted survey respondents than on the full sample while estimates on enrollment are consistently smaller. Despite these deviances, the conclusions are generally consistent, suggesting that survey nonresponse only mildly distorts impact estimates.

Despite the lack of evidence of serious bias associated with survey nonresponse, the evaluation team developed and applied weights to adjust for nonresponse, based on statistical models of the association between baseline characteristics and response probabilities within each of the two randomly assigned groups. Covariates also included several measures of college enrollment and credential receipt over the follow-up period. These methods are common in survey research.

The main steps in constructing weights follow:

1. Winnow the list of potential covariates that are statistically significant in a logistic regression model for response status.[15] Do this separately for treatment and control cases. This approach identified living with spouse/partner and children, living with parents, receiving WIC/SNAP, owning a car, having ever been arrested, planning to work more 20 to 34 hours per week (all as of at baseline), and college persistence (per NSC records) as significant predictors of response status in the treatment sample. The team found fewer significant predictors of response status in the control arm: living with spouse/partner and children, having earned a vocational technical certificate, and financial hardship (all as of at baseline).
2. Using the winnowed list of potential covariates, estimate the response propensity separately for each member of the treatment and control sample—both for respondents and nonrespondents.
3. Sort the sample in each study arm by the estimated response propensity, and then divide the sample into five equal-size groups (quintiles).
4. Within each arm and quintile, calculate the empirical response rate. Invert it to calculate the nonresponse-adjusted weight.

The last column in Exhibit C-3 shows that the weighting eliminated two of the four statistically significant baseline imbalances among survey respondents but also caused a new imbalance on presence of both computer and internet at home.[16] Switching attention to Exhibit C-4, the benefits of nonresponse weighting appear to be mixed. The weighting sharply reduced bias on the estimated impact of I-BEST on workforce awards, receipt of any college credential and credits earned through 6 and through 12 months. On the other hand, weighting aggravated bias on the estimated impact of I-BEST on enrollment in academic courses and in vocational courses. Moreover, for the confirmatory outcome of academic/workforce credits earned, the weighting reversed a positive bias on the estimated impact of 1.2 credits into a negative bias of 1.4 credits.

Thus, the investigation of bias in estimated impacts on administrative outcomes when estimated from survey respondents did not provide the team with strong guidance on whether to use the nonresponse-adjusted weights for analyses of survey-reported outcomes or to

[15] The team used the stepwise search option in SAS/LOGISTIC for this purpose with a *p*-value to enter the model of .20 and a *p*-value to stay in the model of .10.

[16] Not shown in this table, the adjustment was effective in making the weighted treatment respondent sample resemble the full treatment sample more closely and making the weighted control respondent sample resemble the full control sample more closely. In fact, 20 of 25 within-arm imbalances were removed by weighting. However, given that the paramount focus of this study is on treatment/control differences, the team did not believe that this improvement should be an important consideration in whether to use nonresponse adjustment weights.

estimate I-BEST outcomes on these outcomes without weights. In the end, the team decided to use the weights in estimated impacts based on survey data because of the strong tradition in the field of using nonresponse-adjusted weights in such analyses. The consensus opinion in the field appears to be that even when nonresponse adjustment is unnecessary, it does little harm—the principal disadvantage is slightly higher variances. This seemed to the team to be a reasonable trade-off to increase protection against nonresponse bias.

Exhibit C-3: Baseline Balance on Full Sample, Unweighted Respondent Sample, and Weighted Respondent Sample

SBCTC Baseline Characteristics	All Participants			Survey Respondents Unweighted			Survey Respondents Weighted		
	Treatment	Control	*p*-Value	Treatment	Control	*p*-Value	Treatment	Control	*p*-Value
Age			*.067*			*.048*			*.018*
20 or under (%)	23.2	21.1		23.5	20.0		23.8	21.0	
21 to 24 (%)	11.1	18.6		9.7	18.7		9.2	19.7	
25 to 34 (%)	31.4	28.1		29.0	26.2		30.1	26.0	
35 or older (%)	34.3	32.2		37.8	35.1		36.9	33.3	
Gender			*.231*			*.023*			*.032*
Female (%)	55.1	59.9		53.6	64.0		53.1	63.1	
Male (%)	44.9	40.1		46.4	36.0		46.9	36.9	
Race/Ethnicity			.346			.272			.417
Hispanic (%)	28.9	23.1		28.8	23.0		27.5	22.8	
Black non-Hispanic (%)	6.2	9.1		5.5	10.0		6.5	10.1	
White non-Hispanic (%)	53.1	56.7		51.2	53.1		52.7	53.5	
Other non-Hispanic (%)	13.4	14.8		15.2	18.0		13.8	17.5	
Living Arrangements			.591			.449			.470
Neither spouse/partner nor children (%)	48.7	45.8		47.2	41.9		49.2	45.0	
No spouse/partner, living with children (%)	14.6	18.6		14.4	18.6		14.5	17.8	
Spouse/partner, no children (%)	18.2	16.3		19.2	17.2		18.6	15.7	
Living with Spouse/Partner and Children (%)	18.5	19.3		19.2	22.3		17.7	21.5	
Living with Parents (%)	27.2	30.1	.412	29.3	29.3	.992	27.5	31.5	.357
One/Both Parents Had Some College (%)	45.5	45.2	.955	46.5	44.0	.626	46.3	44.5	.728
High School Grades			.170			.204			.343
Mostly Got A's (%)	7.7	6.2		7.5	8.0		7.2	8.0	
Mostly Got B's (%)	36.8	29.5		39.9	30.5		39.1	31.2	
Mostly got C's or Below (%)	55.6	64.3		52.6	61.6		53.7	60.9	
Educational Attainment			.497			.860			.892
Less Than a High School Degree (%)	28.2	33.1		24.9	28.6		25.6	29.3	
High School or Equivalent (%)	42.0	38.0		43.1	38.7		43.3	39.4	
Less Than 1 Year of College (%)	12.1	10.2		12.4	11.5		11.4	11.2	
1 or More Years of College (%)	10.1	8.9		10.2	11.1		10.9	10.3	
Associate's Degree or Higher (%)	7.7	9.8		9.3	10.1		8.7	9.8	
Received Vocational or Technical Certificate or Diploma (%)	19.7	19.0	.853	20.9	22.4	.711	20.3	18.1	.543
Career Knowledge Index (average of items)	0.41	0.41	.934	0.43	0.39	.303	0.42	0.39	.334

SBCTC Baseline Characteristics	All Participants			Survey Respondents, Unweighted			Survey Respondents, Weighted		
	Treatment	Control	*p*-Value	Treatment	Control	*p*-Value	Treatment	Control	*p*-Value
Income			.551			.867			.828
Less than $15,000 (%)	46.5	48.1		43.0	43.1		45.8	43.6	
$15,000-$29,999 (%)	26.0	21.9		26.4	24.3		25.3	24.5	
$30,000 or More (%)	27.6	30.0		30.6	32.6		28.9	31.9	
Mean ($)	23,002	21,240	.378	24,653	22,865	.447	23,336	22,562	.735
Psycho-Social Indices									
Academic Discipline Index	5.05	5.08	.528	5.09	5.14	.404	5.07	5.13	.366
Training Commitment Index	5.42	5.43	.815	5.40	5.42	.729	5.40	5.42	.703
Academic Self-Confidence Index	4.47	4.48	.960	4.43	4.50	.337	4.42	4.50	.282
Emotional Stability Index	4.95	4.94	.943	4.98	5.00	.772	4.99	5.00	.896
Social Support Index	3.21	3.21	.898	3.21	3.24	.388	3.19	3.24	.235
Stress Index	2.30	2.31	.887	2.27	2.29	.749	2.31	2.29	.788
Depression Index	1.61	1.59	.729	1.59	1.58	.800	1.63	1.58	.390
Life Challenges Index (averages in original units 1-5)									
Public Assistance/Hardship Past 12 Months									
Received WIC or SNAP (%)	*55.0*	*62.1*	*.092*	*50.2*	*60.1*	*.045*	55.4	61.0	.252
Received Public Assistance or Welfare (%)	*18.1*	*24.3*	*.094*	*17.0*	*24.1*	*.081*	19.3	24.3	.250
Reported Financial Hardship (%)	49.8	47.1	.499	45.9	43.2	.571	47.6	46.0	.740
Current Work Hours			.993			.946			.940
0 (%)	66.9	66.3		67.1	67.0		67.7	66.9	
1 to 19 (%)	8.5	8.5		9.4	8.6		9.7	8.7	
20 to 34 (%)	11.7	11.6		9.9	11.5		9.6	11.3	
35 or more (%)	12.8	13.6		13.6	12.9		13.1	13.1	
Expected Work Hours in Next Few Months			.228			.468			.327
0 (%)	41.4	40.8		42.4	43.3		41.3	44.0	
1 to 19 (%)	8.6	11.2		8.3	10.2		8.2	10.3	
20 to 34 (%)	35.2	28.9		33.2	27.0		34.6	26.7	
35 or more (%)	14.8	19.1		16.1	19.5		16.0	19.1	
Owns a Car (%)	62.1	63.4	.733	66.7	66.4	.946	62.5	65.8	.467
Has Both Computer and Internet at Home (%)	70.2	73.7	.338	70.6	75.8	.218	67.7	75.8	.065
Ever Arrested (%)	28.4	29.6	.740	26.0	27.4	.733	28.9	27.2	.712

SOURCE: Abt Associates calculations based on data from the PACE Basic Information Form (BIF), the PACE Self-Administered Questionnaire (SAQ), and response status to the PACE short-term follow-up survey.

NOTES: SAS/SURVEYFREQ used to test for significant imbalances for categorical variables. SAS/TTEST was used to test for significant imbalances for other variables. The research team did not run baseline tests for each line of a multi-category construct. Instead, the procedure called for determining imbalance within the construct generally, and not for specific categories.

Exhibit C-4: Comparison of Selected Impact Estimates for the Full Sample and Unweighted and Weighted Survey Samples

Outcome	Full Sample	Survey Respondents	
		Unweighted Est (StdErr)	Weighted Est (StdErr)
Academic and workforce credits earned at college through month:			
6	5.65***	6.23***	5.62***
		(0.62)	(0.62)
12	9.73***	10.71***	9.28***
		(1.36)	(1.30)
18	11.21***	12.23***	10.19***
		(2.01)	(1.90)
24	13.21***	14.31***	11.77***
		(2.70)	(2.56)
Earned college credential through 24 months: (%)			
Workforce completion	3.75**	4.32**	3.43*
		(2.33)	(2.29)
Workforce award	33.34***	37.17***	33.75***
		(3.62)	(3.59)
Associate's degree or high	-0.04	-0.66	-0.76
		(0.47)	(0.49)
Any	32.13***	35.70***	32.11***
		(3.87)	(3.85)
Any enrollment through 24 months in: (%)			
Academic courses	6.14*	4.39	1.73
Vocational courses	40.91***	39.31***	36.59***
Any Courses (including basic and developmental)	21.92***	21.49***	20.04***
Earned a credential at 18 months from: (%)			
A college	30.75***	33.70***	30.74***
		(3.75)	(3.72)
Another education and training institution		-3.41	-3.22
		(1.44)	(1.44)
A licensing/certification body		15.52***	15.14***
		(3.86)	(3.84)
Any issuer		29.74***	27.55***
		(4.22)	(4.25)

SOURCE: Abt Associates calculations based on data from the administrative records and the PACE short-term follow-up survey.
NOTES: Standard errors on estimated impacts are shown in parentheses. All estimates are regression adjusted, as discussed in Section A.3 of Appendix A. Statistical significance levels, based on one-tailed *t*-tests tests of differences between research groups, are summarized as follows: *** statistically significant at the one-percent level; ** at the five-percent level; * at the 10-percent level.

Appendix D: Treatment of Outliers

The team 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*. 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.

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) and credits. The research team found no values that were clearly impossible, and thus retained all reported values in the analysis.