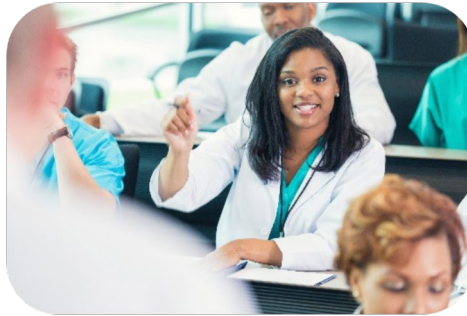
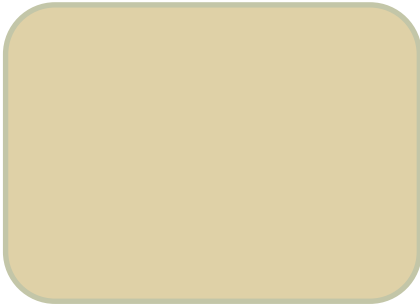


Madison Area Technical College's Patient Care Pathway Program

Appendices for Three-Year Impact Report



OPRE Report 2020-161

November 2020



PACE
Pathways for Advancing
Careers and Education

Madison Area Technical College's Patient Care Pathway Program: Appendices for Three-Year Impact Report

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

OPRE Report 2020-161

November 2020

Douglas Walton, Daniel Litwok, Gabriel Durham, and David Judkins, Abt Associates

Submitted to:

Nicole Constance and Amelia Popham, Project Officers
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP23320095624WC, Task Order HHSP23337019T

Project Director: Larry Buron

Principal Investigator: David Fein; Director of Analysis: David Judkins
Abt Associates

6130 Executive Boulevard
Rockville, MD 20852

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Walton, Douglas, Daniel Litwok, Gabriel Durham and David Judkins. 2020. *Madison Area Technical College's Patient Care Pathway Program: Appendices for Three-Year Impact Report*. OPRE Report 2020-161. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

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](https://twitter.com/OPRE_ACF)



Like OPRE's page
on Facebook
[OPRE.ACF](https://www.facebook.com/OPRE.ACF)



Follow OPRE
on Instagram
[@opre_acf](https://www.instagram.com/opre_acf)



Connect on
LinkedIn
[company/opreacf](https://www.linkedin.com/company/opreacf)



Contents

Appendix A: Baseline Characteristics and Adjustments	1
A.1 Details on Baseline Covariates	1
A.2 Comparing Treatment and Control Groups at Baseline	4
A.3 Regression Adjustment.....	6
Appendix B: College Records Data	15
B.1 Rationale for Use of Madison College Records.....	15
B.2 Imputation of Enrollment, Credits Earned, Credentials Earned at Colleges Other Than Madison College	16
Appendix C: Three-Year Survey Data	24
C.1 Measures Based on Follow-up Survey Data	25
C.2 Imputation in the Three-Year Survey	30
C.3 Survey Nonresponse Analysis	44
C.4 Quality and Completeness of Exam-Based Credentials Reported in the Survey ...	53
C.5 Quality and Completeness of School-Issued Credentials Reported in the Survey	54
Appendix D: National Student Clearinghouse Data.....	56
D.1 Coverage	56
D.2 Data and Measures	57
D.3 Program Impacts on NSC-Measured Outcomes	57
Appendix E: Sensitivity Analyses of Education Impacts	60
Appendix F: NDNH's Unemployment Insurance Wage Data.....	62
F.1 Data Collection Process.....	62
F.2 Data and Measures	63
Appendix G: Comparing NDNH- and Survey-Based Employment and Earnings Estimates	65
Appendix H: Treatment of Outliers	67
Appendix References	68

List of Exhibits

Exhibit A-1:	Operationalization of Baseline Measures Used as Covariates in Regression-Adjusted Impact Estimates.....	2
Exhibit A-2:	Baseline Balance	5
Exhibit A-3:	Covariates Selected, by Outcome Domain.....	12
Exhibit A-4:	Comparison of Confirmatory and Secondary Impact Estimates Unadjusted and Adjusted for Baseline Imbalances	14
Exhibit B-1:	NSC-Reported Enrollment at Madison College and Other Colleges by Research Group	16
Exhibit B-2:	Predictive Power of Models for Madison College-Reported Education Outcomes	21
Exhibit B-3:	Spell Counts and Matched-Pair Correlations for Each of the Four Match Variables.....	23
Exhibit C-1:	Details on Specifications for Survey-Based Education Outcomes in Chapter 3	25
Exhibit C-2:	Details on Specifications for Survey-Based Employment/Earnings Outcomes in Chapter 4	26
Exhibit C-3:	Details on Specifications for Survey-Based Family Economic Well-Being Outcomes in Chapter 5	27
Exhibit C-4:	Details on Specifications for Survey-Based Intermediate Outcomes in Chapter 5	28
Exhibit C-5:	Details on Specifications for Survey-Based Family Structure Outcomes in Chapter 5	29
Exhibit C-6:	Imputation Rates among Survey Respondents in PCPP	31
Exhibit C-7:	Comparison of Selected Impact Estimates of PCPP	39
Exhibit C-8:	Date Imputation for Three-Year Impact Study (Pooled Sample).....	42
Exhibit C-9:	Baseline Balance on Full Sample, Unweighted Respondent Sample, and Weighted Respondent Sample.....	46
Exhibit C-10:	Comparison of Selected Estimates of the Impact of PCPP for the Unweighted and Weighted Survey Samples	49
Exhibit D-1:	NSC College-Level Cooperation Rates by College Type, 2012 through 2016.....	56
Exhibit D-2:	Comparisons of Impacts of PCPP Based on Madison College Records with Impacts Based on NSC Records	59
Exhibit E-1:	Comparisons of Impacts of PCPP Based on Madison College Records with Impacts Based on the Three-Year Follow-up Survey	61
Exhibit G-1:	Impacts of PCPP on Earnings and Employment around Follow-up Q12 Based on Wage Records and Self-Reports.....	66

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 the Patient Care Pathways Program (PCPP) (Cook, Hamadyk, Zeidenberg, Rolston, and Gardiner 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 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 Other, 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 six responses blank.	SAQ: S15
Stress ^g	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, Casillas, Robbins, and Langley (2005). Further validation in Peterson, Casillas, and Robbins (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).

^gCohen, Kamarck, and Mermelstein (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 that no systematic differences exist 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. The *p*-values shown in Exhibit A-2 indicate that there were statistically significant differences between the randomly assigned treatment and control groups in four of the baseline characteristics. 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 such chance differences might have on the impact estimates.

Exhibit A-2: Baseline Balance

Characteristic	All Participants	Treatment Group	Control Group	p-Value
Age (%)				.238
20 or under	23.4	19.9	26.9	
21-24	21.2	23.1	19.3	
25-34	30.0	32.3	27.7	
35+	25.4	24.7	26.1	
Female (%)	84.3	86.4	82.3	.203
Race/Ethnicity (%)				.021
Hispanic, any race	8.8	12.4	5.3	
Black, non-Hispanic	20.8	19.8	21.8	
White, non-Hispanic	67.3	64.2	70.4	
Other, non-Hispanic	6.4	7.8	4.9	
Family Structure (%)				.530
Not living with spouse/partner and not living with children	43.5	40.7	46.5	
Not living with spouse/partner but living with children	18.3	20.3	16.2	
Living with spouse/partner and not living with children	19.7	19.9	19.5	
Living with spouse/partner and children	18.5	19.1	17.8	
Living with parents (%)	26.5	22.8	30.3	.060
One parent has at least some college (%)	51.8	50.4	53.2	.542
Usual High School Grades (%)				.255
Mostly A's	6.1	6.3	5.9	
Mostly B's	43.1	46.7	39.5	
Mostly C's or below	50.8	47.1	54.6	
Highest Level of Education (%)				.211
Less than a high school diploma	3.0	4.8	1.2	
High school diploma or equivalent	44.4	44.8	43.9	
Less than one year of college	24.8	24.4	25.2	
One or more years of college	21.6	20.4	22.8	
Associate degree or higher	6.3	5.6	6.9	
Received vocational or technical certificate or diploma (%)	39.5	41.1	38.0	.481
Career Knowledge Index (mean)	0.41	0.43	0.39	.196
Psycho-Social Indices (means)				
Academic Discipline Index	5.06	5.05	5.07	.786
Training Commitment Index	5.59	5.61	5.58	.628
Academic Self-Confidence Index	4.43	4.46	4.39	.361
Emotional Stability Index	4.99	5.04	4.95	.203
Social Support Index	3.31	3.30	3.32	.493
Stress Index	2.22	2.21	2.24	.748
Depression Index	1.55	1.52	1.57	.310

Characteristic	All Participants	Treatment Group	Control Group	p-Value
Family Income in Past 12 Months (%)				.038
Less than \$15,000	25.6	27.6	23.5	
\$15,000-\$29,999	29.9	33.5	26.1	
\$30,000+	44.6	38.9	50.4	
Family income (mean)	\$33,165	\$31,694	\$34,694	.280
Public Assistance/Hardship Past 12 Months				
Received WIC or SNAP (%)	35.6	32.9	38.2	.193
Received public assistance or welfare (%)	4.4	5.1	3.7	.600
Reported financial hardship (%)	34.3	35.5	33.2	.528
Current Work Hours (%)				.949
0	27.9	27.4	28.5	
1-19	11.5	11.7	11.4	
20-34	32.6	31.9	33.3	
35+	27.9	29.0	26.8	
Expected Work Hours in Next Few Months (%)				.227
0	18.3	16.4	20.3	
1-19	15.1	18.1	12.1	
20-34	47.4	47.8	47.0	
35+	19.2	17.7	20.7	
Life Challenges Index (mean)	1.44	1.45	1.43	.710
Owns a car (%)	84.9	86.7	83.1	.207
Has both computer and internet at home (%)	84.1	84.3	83.9	.900
Ever arrested (%)	20.8	24.0	17.7	.082
Sample sizes	499	250	249	

Source: PACE Basic Information Form and 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.

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

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.

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 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 PACE evaluation of PCPP) 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 used 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

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

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:

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

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 δ and can accommodate the required nonresponse-adjustment weights for survey-measured outcomes. (See Appendix C, 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 $\text{var}(\hat{\delta}) \approx (1 - R^2)\text{var}(\bar{y}_t - \bar{y}_c)$, where R^2 is the proportion of the variance in Y_i explained by the baseline characteristics (X_i) in the OLS estimation of Equation A.2 below:

$$Y_i = X_i\beta + e_i \quad (\text{A.2})$$

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.

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.

Simulation research (Judkins 2019) showed that dropping (with “backward selection”) or downweighting covariates⁷ 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 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 in order 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 $\bar{Y}_i - Y_i$ is measured for each of the candidate values of the constraint.

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, Fein, and Buron 2018; Koch, Tangen, Jung, and Amara 1998).

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 PCPP 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.

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 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

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.

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 main report, for PCPP, the most salient education outcome is receipt of a credential requiring a year or more of college classes. 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 or by virtue of their being out of balance (OOB) at baseline. Some covariates were selected both by the LASSO and by virtue of being out of balance at baseline. These covariates are flagged as such in the table. 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. In contrast, the out-of-balance test selected all or none of the levels of a categorical variable. The table shows all possible levels of categorical variables and indicates which specific categories we selected as covariates.

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 (OCSE) 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 properly match the Social Security Administration's records.

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).

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

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

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

Note: "LASSO" flags that the covariate was selected by the LASSO for variance reduction. "OOB" flags that the covariate was selected because it was significantly out of balance.

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 substantially for two of the three targeted outcomes (earnings and receipt of means-tested benefits), but otherwise did not appear to be beneficial. In order to get variance reduction on every estimate or at least avoid increasing standard errors, it would probably be necessary to run a separate LASSO for each.

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-10 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	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings (NDNH)				
Full Sample				
Average quarterly earnings Q12-Q13 after randomization (\$)	-47	336	-175	313
Secondary Outcomes: Employment (Survey)				
Survey Respondents without Weights				
Employed at survey follow-up (%)	-0.6	4.8	-0.1	4.9
Employed at \$13 per hour or above (%)	8.9**	5.4	9.7**	5.5
Employed in a job requiring at least mid-level skills (%)	-9.5	4.2	-9.5	4.2
Confirmatory Outcome: Education (Madison College Records)				
Full Sample				
Received 1+ year college credential from a college (%)	-3.6	2.1	-3.4	2.1
Secondary Outcomes: Education (Madison College Records)				
Full Sample				
Number of college credits earned from a college (%)	1.4	1.7	1.4	1.7
Full-time-equivalent months enrolled at a college (%)	1.0*	0.7	1.0*	0.7
Received any credential from a college (%)	3.9	3.5	3.9	3.5
Secondary Outcomes: Education (Survey)				
Survey Respondents without Weights				
Received an exam-based certification or license (%) ^a	-3.9	5.4	-4.0	5.5
Secondary Outcome: Other (Survey)				
Survey Respondents without Weights				
Indicators of Independence and Well-Being				
Has health insurance coverage (%)	5.4*	3.5	6.2**	3.5
Receives means-tested public benefits (%)	3.0	6.0	4.3	5.2
Personal student debt (\$)	932	1,286	1,244	1,284
Financial hardship (%)	-2.5	5.6	-0.3	5.4
Indices of Self-Assessed Career Progress (average)				
Confidence in career knowledge ^b	0.06	0.07	0.01	0.07
Access to career supports ^c	0.02	0.03	0.01	0.03
Sample sizes (across treatment and control groups): NDNH and Madison College Records 499 Survey 326				

Source: Madison College records; PACE 18-month follow up survey; PACE three-year follow-up survey; 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), are summarized as follows: *** 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 education outcome (*receipt of a college credential that typically requires at least a year of college*) as well as other important college outcomes. It discusses our decision to base most such measures on records from Madison College, rather than on potential alternative sources such as the three-year follow-up survey and the National Student Clearinghouse (NSC). This decision is explained in Section B.1. Section B.2 discusses the specification of outcomes constructed from the college records.

In addition to preparing impact estimates with the college records data, we also prepared alternative impact estimates using NSC data and survey data. See Appendices D and E, respectively, for these sensitivity analyses.

B.1 Rationale for Use of Madison College Records

The Madison College records offer several advantages over other potential sources for defining college education outcomes. For study participants who enrolled in Madison College, the records provide complete data on their enrollment and credentials. This is a key advantage over the follow-up survey, which only covers respondents 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 from credentials requiring less time. This is an important advantage over the NSC credential data, which does not have sufficient information to categorize credentials as taking less than or more than a year to complete.

A limitation of the Madison College records is that they do not cover credentials earned at other schools. If a large share of study participants attended other schools, then outcomes based on Madison College records would underestimate credential attainment. Importantly, if attendance at other schools differed between the research groups, then our impact estimates could be biased.

Analysis of NSC records showed that the majority of study participants who enrolled in college attended Madison College (Exhibit B-1 below). More than 95 percent of those with any college attendance attended Madison College; more than 80 percent attended only Madison College; about 14 percent attended both Madison College and another school. Less than 5 percent attended only another school, although control group participants were 3 percentage points more likely to attend only another school than were the treatment group (4.3 for the control group, 1.3 for the treatment group).

These results suggest that Madison College records provide sufficient coverage to use as the primary source for defining college education outcomes, but that an imputation procedure would be necessary to ensure that attendance at other colleges is also captured in the outcomes, in order to avoid any potential bias in the results.

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

Group	Enrollment Documented in NSC				
	Enrolled at Madison College			Enrolled Only at Colleges Other Than Madison College	Total Ever Enrolled
	At Madison College	Only at Madison College	At Madison College and Other College		
At Three Years					
Treatment (%)	98.7	84.6	14.0	1.3	100.0
Control (%)	95.7	81.3	14.4	4.3	100.0
T-C difference (percentage points)	3.0	3.3	-0.4	-3.0	

Source: NSC.

Note: NSC data are available for a longer period, but for this table we used only NSC data on the 35 months following the month of random assignment.

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

The prior section explained our decision to base most college outcomes on Madison College records, supplemented with imputed experiences at other colleges. This section documents the imputation procedure we used for this purpose. The imputation process is based largely on information from the NSC, but also uses information from baseline forms and both rounds of follow-up surveys. Briefly, our approach entails matching NSC-reported spells at colleges other than Madison College to spells at Madison College, and then copying *Madison College-reported variables* associated with that Madison College spell to the person who attended the other college.

For example, if NSC reports that “Bill” attended the University of Wisconsin for a certain number of months and was (or was not) awarded a credential from the University of Wisconsin, the procedure identifies a person (“Susan”) who attended Madison College as a match. The match is based on statistical procedure that assesses the similarity between Bill and Susan in their length of attendance and educational outcomes. (Detailed matching criteria are found below. This is step #5 in the process.) The procedure then duplicates Susan’s spell at Madison College as a spell for Bill (step #6 below). Continuing the example, if Madison College shows that Susan received an associate degree from Madison College, then the procedure imputes that Bill received the same at the University of Wisconsin. 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 Madison College.

This imputation yields unbiased impact estimates if, conditional on the information used in the matching, a spell at Madison 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 Madison 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 for this three-year report used NSC-reported spell duration, NSC data on credential receipt, follow-up survey data on credential receipt, and baseline information to identify suitable Madison College spells.

One potential shortcoming of the procedure is that it does not use information on receipt of support services. Because PCPP did boost some forms of supports, such as receipt of academic advising (Cook et al. 2018, Exhibit 4-7), 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 academic advising to be a problem, the advising would need to have had important effects other than boosting receipt of credentials. An example might be an advising-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 think it is unlikely.

At a more detailed level, the imputation strategy involved six steps:¹⁷

- (1) Find a Madison College record for as many NSC-reported Madison College spells as possible. This step filled in credits and credentials earned for most NSC-reported Madison College spells. The research team referred to this step as the *exact matching* process, because there is a single correct match in the Madison College records for almost all of the NSC-reported Madison College spells.
- (2) Resolve NSC-reported Madison College spells that did not match to a Madison College record.¹⁸ 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 data collected at baseline and at the first follow-up survey.
- (4) Summarize the available data for each NSC-reported Madison College spell and the student to whom the spell belonged. The team summarized these data by using statistical models that predict four critical Madison College–reported outcomes for each spell (credits earned by 12 months after randomization; credits earned by 24 months after randomization; receipt of a college certificate by 36 months after randomization; receipt of a college degree by 36 months after randomization).
- (5) Match each NSC-reported other-college spell with a similar NSC-reported Madison College spell in terms of the four predicted critical outcomes. The team referred to this step as *statistical matching*, because many possible NSC-reported Madison College spells could be

¹⁷ We used a similar procedure to model Madison College outcomes for the short-term impact report, but step 3 was not part of the prior procedure, and step 4 at 18 months did not use survey data.

¹⁸ There were also Madison College records that did not match to NSC. We ignored these records in this process since they did not affect the imputation. We were focused on finding spells that were comparable to NSC-reported spells.

matched to every NSC-reported other-college spell. The team only matched other-college spells of students in the treatment group to Madison College 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. The team imposed this restriction to avoid “washing out” any effects by making control experiences artificially more similar to treatment experiences.

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

The following sections give more information for each step.

Details on Step 1 (Exact Matching)

We conducted the exact matching of each NSC-reported spell at Madison College with a Madison College–reported spell by determining the amount of overlap between the spells, based on the start and end dates of each spell.

If only one Madison College–reported spell overlapped with an NSC-reported spell at Madison College, then we considered the two spells to be matched without regard to how well start and end dates aligned between the two systems. If multiple Madison College–reported spells overlapped with one NSC-reported spell at Madison College, then we considered the Madison College–reported spell with the most months of overlap to be matched to the NSC-reported spell. If one Madison College–reported spell overlapped with multiple NSC-reported spells at Madison College, then the Madison College–reported spell was broken into pieces that better matched the NSC-reported spells.

We then transcribed the outcomes associated with the Madison College–reported spell in the Madison College record system over to the NSC-reported spell. Outcomes included FTE months enrolled, enrollment by quarter, credits earned, and credentials earned.

Details on Step 2 (Unmatched NSC-reported Madison College Records)

In a small number of cases, NSC-reported spells at Madison College did not overlap with any Madison College–reported spells. Data investigation with Madison College 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 Madison College provided to the evaluation. However, the college does appear to include these records in extracts it sends to the NSC. For these spells, we assumed that the student had experienced zero hours of instructional credits and earned no credits or credentials.

Details on Step 3 (Impute Survey Outcomes for Survey Nonrespondents)

We reasoned that survey data on credential attainment would help us find better Madison College matches for NSC-reported spells at colleges other than Madison College. However, this strategy works only if the survey data are available on the full sample, including survey nonrespondents. 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 PACE programs, we imputed eight 3-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 academic credential.
- Earning of a long-term academic credential less than a degree.
- Earning of an associate degree.
- Earning of a bachelor's or graduate degree.
- 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 three steps: modeling the probability of earning a credential requiring a year or more of training (including degrees) in terms of baseline variables and variables from the short-term follow-up survey;¹⁹ using the predictions to form propensity strata; and then imputing by randomly selecting a value from a record within the same cell defined by site, treatment group, and propensity stratum (i.e., hotdeck imputation).

Details on Step 4 (Data Summarization)

We needed to create a parsimonious set of key variables on which to statistically match Madison College 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 spell, self-reported baseline variables, and the actual or imputed three-year follow-up survey variables from step 3. To facilitate matching, we used statistical models for four Madison College–reported outcomes on the set of exactly matched records in terms of these variables. These variables were selected to reflect both early and late educational experiences, as well as easier and harder goals.

The Madison College–reported outcomes (and their associated variable names) were:

- Credits earned by 12 months of randomization (Credits12);
- Credits earned by 24 months of randomization (Credits24);
- Completion of an associate degree by 36 months after randomization (Degree36); and
- Completion of a certificate by 36 months after randomization (Certificate36).

¹⁹ 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.

The procedure involved first fitting models for these four outcomes on NSC-reported Madison College spells and then using estimated coefficients to predict values for both Madison College spells and spells at other colleges. The models involved up to 32 characteristics from NSC, the baseline forms, and the three-year follow-up survey.

The baseline variables fell into several categories:²⁰

- Age: less than 20 years old; age 21 to 24; age 25 to 34, age 35 and older (omitted category).
- Race/ethnicity: Hispanic, any race; White, non-Hispanic (omitted category); Black, non-Hispanic; other race, non-Hispanic.
- Educational attainment: High school diploma or less (omitted category); less than one year of college; one or more years of college; associate degree or higher.
- Grades: mostly A's; mostly B's; mostly C's or lower (omitted category).
- Income: up to \$15,000; between \$15,000 and \$29,999; \$30,000 or more (omitted category).
- Psycho-social skill scales: academic discipline (ADISI), training commitment (TCOMI), and academic confidence (ACONI).²¹

The NSC variables included the number of FTE enrolled months in the spell (at 12 and 18 months); number of completions since random assignment (at 12 and 18 months); sum of enrollment statuses in December, May, and July (at 12 and 18 months); and number of months that the spell overlaps with the 12- or 18-month window.

The survey variables consisted of the eight outcomes that were imputed in step 3 above: number of college credentials, the number of vocational credentials, the number of licenses, receipt of a short-term college certificate, receipt of a long-term college certificate, receipt of an associate degree, receipt of a bachelor's or graduate degree, and self-assessed career progress.

We used stepwise selection procedures in the modeling. To quantify the improvement (relative to the similar procedure used in the prior, short-term report), we performed this process 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 first report. In the second iteration, three-year follow-up survey outcomes were also included. Thus, for each outcome, two models were created this time, one model from each of these two iterations. We chose the better of the two, defined as the one with the higher *R*-squared.

The chosen models from each iteration, along with their *R*-squared, are shown in Exhibit B-2 below. Note that the use of the three-year follow-up survey variables led to a substantial

²⁰ For categorical variables with three or more levels, we always left one level out of the predictor list. These omitted categories are not shown here.

²¹ See Appendix A for definitions of these scales.

improvement in the model for *degree attainment* (*R*-squared boosted from 12.6 to 37.7 percent), and a modest improvement in the model for *college certificate earned by 36 months* after randomization (*R*-squared boosted from 10.5 to 17.6 percent). Use of the survey data for the credits outcomes (Credits12, Credits24) provided little benefit.

Exhibit B-2: Predictive Power of Models for Madison College–Reported Education Outcomes

Madison College–Based Outcome	Without Three-Year Follow-up Survey Variables		Including Three-year Follow-up Survey Variables	
	Independent Variables	<i>R</i> -squared	Independent Variables	<i>R</i> -squared
Degree36	<u>Baseline survey</u> <ul style="list-style-type: none"> • Age 21 to 24 <u>NSC</u> <ul style="list-style-type: none"> • FTE enrolled months in spell at 18 months • FTE enrolled months in spell at 12 months • Number of completions at 18 months • Number of completions at 12 months • Months of overlap at 18 months • Months of overlap at 12 months 	0.126	<u>Baseline survey</u> <ul style="list-style-type: none"> • none selected <u>NSC</u> <ul style="list-style-type: none"> • FTE enrolled months in spell at 18 months • FTE enrolled months in spell at 12 months • Number of completions at 18 months • Number of completions at 12 months <u>3-year follow-up survey</u> <ul style="list-style-type: none"> • Associate degree 	0.377
Certificate36	<u>Baseline survey</u> <ul style="list-style-type: none"> • Hardship <u>NSC</u> <ul style="list-style-type: none"> • FTE enrolled months in spell at 18 months • Number of completions at 12 months • FTE enrolled months in spell at 12 months • Months of overlap at 12 months • Months of overlap at 18 months 	0.105	<u>Baseline survey</u> <ul style="list-style-type: none"> • Hardship • Hispanic, any race <u>NSC</u> <ul style="list-style-type: none"> • FTE enrolled months in spell at 18 months • Number of completions at 12 months • Right overlap • Sum of enrollment statuses at 12 months • FTE enrolled months in spell at 12 months <u>3-year follow-up survey</u> <ul style="list-style-type: none"> • Collegecredcount • College certificates (1) • College certificates (2) 	0.176

Madison College–Based Outcome	Without Three-Year Follow-up Survey Variables		Including Three-year Follow-up Survey Variables	
	Independent Variables	R- squared	Independent Variables	R- squared
Credits12	<u>Baseline survey</u> <ul style="list-style-type: none"> Hispanic, any race Black, non-Hispanic Income <\$15,000 <u>NSC</u> <ul style="list-style-type: none"> FTE enrolled months in spell at 12 months FTE enrolled months in spell at 18 months Sum of enrollment statuses at 12 months Sum of enrollment statuses at 18 months Spell length Right overlap 	0.748	<u>Baseline survey</u> <ul style="list-style-type: none"> Hispanic, any race Black, non-Hispanic Income <\$15,000 <u>NSC</u> <ul style="list-style-type: none"> FTE enrolled months in spell at 12 months FTE enrolled months in spell at 18 months Sum of enrollment statuses at 12 months Sum of enrollment statuses at 18 months Spell length Right overlap <u>3-year follow-up survey</u> <ul style="list-style-type: none"> <i>none selected</i> 	0.748
Credits24	<u>Baseline survey</u> <ul style="list-style-type: none"> Hispanic, any race Black, non-Hispanic <u>NSC</u> <ul style="list-style-type: none"> FTE enrolled months in spell at 18 months FTE enrolled months in spell at 12 months Number of completions at 18 months Right overlap Sum of enrollment statuses at 12 months Spell length Months of overlap at 12 months 	0.801	<u>Baseline survey</u> <ul style="list-style-type: none"> Hispanic, any race Black, non-Hispanic <u>NSC</u> <ul style="list-style-type: none"> FTE enrolled months in spell at 18 months Spell length FTE enrolled months in spell at 12 months Months of overlap at 12 months Number of completions at 18 months Right overlap Sum of enrollment statuses at 12 months <u>3-year follow-up survey</u> <ul style="list-style-type: none"> Associate degree 	0.803

Details on Step 5 (Statistical Matching)

For each spell at a college other than Madison College, we calculated the weighted Euclidean distance from that spell to every spell at Madison College as

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

where z_{li} and z_{lj} are the predicted values for outcome l for the potential match between donor i and unmatched spell j . There is an equal weight c to each of the outcomes. The team selected the Madison College spell j within the same research group (treatment/control) that minimized the distance measure as the matched Madison College spell for the i -th other-college spell. The same Madison College spell could be matched to more than one other-college spell.

Exhibit B-3 provides statistics about the matching. Members of the treatment group had 72 spells at colleges other than Madison College. Each of these was matched to a single one of the 325 Madison College spells experienced by members of the treatment group. Similarly, members of the control group had 71 spells at colleges other than Madison College. Each of these was matched to a single one of the 293 Madison College spells experienced by members of the control group. The quality of the matches was generally high, with matched pair correlations between 0.8 and 1.0.

Exhibit B-3: Spell Counts and Matched-Pair Correlations for Each of the Four Match Variables

Statistic	Treatment	Control
Spell Counts		
At colleges other than Madison College	72	71
At Madison College	325	293
Matched-Pair Correlations		
Predicted credits through 12 months	0.90	0.93
Predicted credits through 24 months	0.81	0.92
Predicted degree receipt through 36 months	0.97	0.99
Predicted college certificate receipt through 36 months	0.96	0.98

Source: NSC, Madison College records, the PACE Basic Information Form, the PACE Self-Administered Questionnaire, and the PACE three-year follow-up survey.

Details on Step 6 (Propagating Madison College Values)

The final step entailed copying matched data on FTE months enrolled, enrollment by quarter, credits earned, and credentials earned from Madison College 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 Madison College—reported outcomes of the Madison College—reported spell that had been exactly matched to the NSC-reported Madison College spell that had been statistically matched to the other-college spell.²² We did this separately for every outcome based on Madison College records.

²² Let ON1 be an NSC-reported spell at a college other than Madison College. Let PN1 be the NSC-reported spell at Madison College that was statistically matched to ON1. Let PP1 be the Madison College record that was exact matched to PN1. 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.²³ 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 instrument.

The survey attempted 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 21st century 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.²⁴ 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 the team also 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. There are two motivations for this two-pass approach.

²³ The full instrument is available at <http://www.career-pathways.org/career-pathways-pace-three-year-instrument/>.

²⁴ 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: 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>

- (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, C-4, and C-5 do the same for family economic well-being outcomes domains, intermediate outcome domains, and family structure outcomes, 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 the 18-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.	Three-year: I3d, I3di, I3h 18-month: A56, A56a
Exploratory Outcomes		
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
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.	C1, C2, C3, D2
Enrolled in training or education at survey follow-up	Determined based on reported dates of enrollment in education and training activities and date of interview.	Most of modules B, C, and E

Key: IPEDS (Integrated Postsecondary Education Data System). 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 \$13 per hour or above	Analyzed response to survey question for control group. Selected \$13 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 PCPP.	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 PCPP.	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 by clerks at the U.S. Census Bureau (under a cooperative arrangement) into an 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

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 Family Economic Well-Being 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 18-month follow-up. It flags any signs of financial distress in terms of troubles paying bills (rent/mortgage, gas/oil/electricity), utility disconnects (gas/electric/oil or 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 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: <ol style="list-style-type: none"> Enough of the kinds of food you want Enough but not always the kinds of food you want Sometimes not enough to eat Often not enough to eat Responses of 1 or 2 counts as not having experienced food insecurity.	M10
Personal receipt of TANF	Respondents were asked about receipt in the prior month	M1a

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Personal receipt of SNAP	Respondents were asked about receipt in the prior month	M1b
Personal receipt of Medicaid	Respondents were asked about receipt in the prior month	M1e

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).

Exhibit C-4: 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	<p>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).</p> <p>Say you need advice or 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:</p> <ul style="list-style-type: none"> – 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	<p>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.</p> <ul style="list-style-type: none"> 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

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Exploratory Outcomes		
Perceived career progress	<p>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 in Fein (2012).</p> <ul style="list-style-type: none"> • 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 8-item scale captures persistence and determination. Response categories ranged from 1 (strongly disagree) to 4 (strongly agree).	K1
Core self-evaluation	Existing scale from Judge (2009). The 12-item scale's 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	<p>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.</p> <ul style="list-style-type: none"> • Childcare arrangements • Transportation • Alcohol or drug use • An illness or health condition • Another situation 	K7

^a Judge, Locke, and Durham 1997; 1998; Judge and Bono 2001.

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

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Secondary Outcomes		
Not living with spouse/partner or children	Respondents were asked about other people who live in the household at least half the time. For this and the next three rows, we treated spouses and unmarried partners the same. Only own children and those of partner/spouse were counted for this.	L1a, L1b, L1c
Not living with spouse/partner, living with children	Ibid	L1a, L1b, L1c

Outcome	Details on Derivation of Outcome	Follow-Up Survey Question(s)
Living with spouse/partner, not living with children	Ibid	L1a, L1b, L1c
Living with spouse/partner and children	Ibid	L1a, L1b, L1c
Living with parents	Parents of either the respondent or any spouse/partner were counted.	L1e
Living with spouse	Here, unmarried partners were not counted.	L1a
Had child since random assignment or currently pregnant (women only)	Women were asked about recent births and current pregnancy	L4, L5
Number of children living with respondent	Only children for whom the respondent or spouse/partner was the legal guardian were counted	L3

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 utilize 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. 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 imputed responses for people who failed to answer the question. Additionally, we imputed blocks of responses for two groups of people: those with large blocks of missing data and those who appear, based on administrative data, to have failed to report one or more education spells.

The goals of imputation were variance and bias reduction. 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. We made decisions for all PACE sites on a global basis. Either we implemented an imputation procedure for a question in all nine PACE sites, or we left the question blank in all sites. Specifically, we imputed for seven types of missing data:

1. number of college credits;
2. credential award dates;
3. income (personal and household);
4. early certificates 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.

Types and Rates of Imputation. Exhibit C-6 below lists the seven types of imputation and shows the imputation rates for the survey respondents in the evaluation sample for PCPP. 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-6: Imputation Rates among Survey Respondents in PCPP

Type of Imputation	Job Spells (%)	School Spells (%)	Credentials (%)	People (%)
1. Number of college credits	n/a	n/a	n/a	16.9
2. Credential award dates	n/a	n/a	4.0	n/a
3. Income				
Personal (categorical)	n/a	n/a	n/a	3.7
Personal (exact)	n/a	n/a	n/a	10.1
Household (categorical)	n/a	n/a	n/a	9.2
Household (exact)	n/a	n/a	n/a	24.8
4. Early certifications and licenses	n/a	n/a	n/a	4.8
5. Skipouts	3.8	3.3	2.6	3.1
6. Spell start and/or end dates (job, school)	3.8	6.3	n/a	n/a
7. Survey data on school spells reported to NSC but not by respondent	n/a	5.2	1.3	4.9

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 our decision to impute this outcome for the 18-month follow-up survey for those 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 “Skipout” row refers to block missingness in the Integrated Training and Employment History module. The German survey upon which this module was modeled experienced a high

level of breakoff (12 percent), 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).²⁵ 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 3.1 percent of PCPP treatment and control 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-6 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 C). 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 Certificates 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.²⁶ 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.

²⁵ 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.

²⁶ 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.

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.²⁷ 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, for some imputations, respondents from HPOG-only programs, as well. 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.²⁸ 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 5-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.²⁹

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

²⁷ See Appendix A.3 for details on the cross-validated LASSO.

²⁸ A "partition" of a sample is an exhaustive and mutually exclusive collection of subsets of the sample.

²⁹ 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.³⁰

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 baseline forms (the PACE Basic Information Form (BIF) and the Self-Administered Questionnaire (SAQ)), 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);

³⁰ 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 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-6, missing data rates were considerably higher for the continuous variables than the categorical variables. This is because categorical

income is only missing 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 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 two- and three-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 nonrespondents³¹ to the 18-month survey would have reported if they had responded at that time. 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.

³¹ 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.

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 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).³² This run found donors for 815 of the 924 skipouts on the pooled dataset.

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 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-7 below. The two sets of impact estimates are quite similar with no differences in statistical significance. The imputation allowed us to use as many as 10 more cases for PCPP, about a 3 percent increase, with the exact count depending on item nonresponse.

Imputation shifted the impact of the program most on employment in a job requiring at least mid-level skills, but not enough to change the statistical significance of the impact. That is, with or without this imputation, there is evidence of a negative program impact on that outcome at the 10 percent level.

Exhibit C-7: Comparison of Selected Impact Estimates of PCPP

Outcome and Sample	Impact Estimate	Standard Error	Sample Size	p-Value
Employed at Survey Follow-up (%)				
Full sample	0.7	5.1	326	.446
Omitting skipouts	1.1	5.1	316	.415
Employed at \$13 per Hour or Above (%)				
Full sample	10.8**	5.6	321	.027
Omitting skipouts	10.6**	5.7	312	.032
Employed in a Job Requiring at Least Mid-Level Skills (%)				
Full sample	-6.8	4.4	318	.941
Omitting skipouts	-7.3	4.4	309	.950

³² This excludes 302 three-year survey respondents that reported no training or employment between randomization and the survey interview.

Outcome and Sample	Impact Estimate	Standard Error	Sample Size	p-Value
Receipt of an Exam-Based Credential (%) (Blended Three-Year and 18-Month Surveys)				
Full sample	-2.3	5.6	326	.659
Omitting skipouts	-2.3	5.7	316	.659

Source: NDNH, NSC, PACE three-year follow-up survey, and 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 summarized as follows: *** 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 have an understanding of 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 gave consideration 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 PCPP, the overall missing data rate was lower than for the rest of the pooled sample (5 percent versus 8 percent).

Exhibit C-8 below lists the models we created for each of nine types 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.

The set of variables allowed into each model varied across the nine lag types. 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.

Exhibit C-8: Date Imputation for Three-Year Impact Study (Pooled 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 spell #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

Lag Type	Modeled Variable	R-Squared (%)	Tested Variables	Selected Variables	Sample Size	Missing Data Rate (%)	Average Lag/Duration	
							Reported (months)	Imputed (months)
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: NDNH, NSC, and the three-year follow-up survey.

Note: Sample pooled across HPOG 1.0 and all nine PACE sites.³³ 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.

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 PCPP sample, there were 16 such respondents, accounting for 5 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 16 respondents to other PCPP 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’s worth of study, but less than an associate degree during the spell;
- Received an associate degree or higher during the spell; and
- 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 PCPP. 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.

³³ 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 also studied in PACE. For more: <https://www.acf.hhs.gov/ofa/programs/hpog>.

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 PCPP, 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.³⁴

The three-year follow-up survey obtained very similar response rates in the treatment (66 percent) and control (65 percent) groups at Madison College. While such similarity suggests that any differences in baseline characteristics and outcomes for respondents versus the full sample would tend to be small and similar in size for the two groups, we did observe some differences when we studied this matter further using administrative data. These differences between the respondents and the full sample are evidence of potential nonresponse bias. We developed a set of nonresponse adjustment weights that appears to remove some 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.

³⁴ See for example, Deke and Chiang (2017). For a slightly contrarian view, see Hendra and Hill (2018).

Exhibit C-9 below considers baseline equivalence, among survey respondents, in the absence of weighting. In the first three columns reflecting all participants, as reported in Section A.2,³⁵ there are four characteristics where we see statistically significant differences between the treatment and control groups. The next three columns, which report statistics for survey respondents, also show statistically significant differences for four characteristics (two of the same characteristics and two different ones). The last column, which reweights the survey respondents, also has four statistically significant imbalances—the imbalances in race and arrests at baseline remain across all columns. 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-10) show evidence of nonresponse bias as well.

³⁵ Note that the numbers in the first three columns of Exhibit C-9 reflect baseline balance for the full sample following imputation, whereas Appendix A.2 presented pre-imputation figures.

Exhibit C-9: Baseline Balance on Full Sample, Unweighted Respondent Sample, and Weighted Respondent Sample

Characteristic	All Participants			Survey Respondents, Unweighted			Survey Respondents, Weighted		
	Treatment	Control	p-Value	Treatment	Control	p-Value	Treatment	Control	p-Value
Age (%)			.234			.345			.355
20 or under	20.0	26.9		19.4	26.7		20.7	25.4	
21-24	22.8	19.3		21.8	16.2		22.8	15.2	
25-34	32.4	27.7		29.1	28.0		28.2	28.5	
35+	24.8	26.1		29.7	29.2		28.3	30.9	
Sex (%)			.171			.191			.206
Female	86.4	81.9		86.7	81.4		86.6	81.2	
Male	13.6	18.1		13.3	18.6		13.4	18.8	
Race/Ethnicity			.021			.071			.070
Hispanic, any race	12.0	5.6		10.3	5.6		9.3	5.7	
Black, non-Hispanic	19.6	21.3		20.0	22.4		19.6	22.6	
White, non-Hispanic	63.2	69.9		61.2	68.3		61.2	68.2	
Other, non-Hispanic	8.0	5.2		9.1	3.7		10.4	3.5	
Family Structure (%)			.444			.777			.832
Not living with spouse/partner and not living with children	40.8	47.0		38.2	43.5		39.5	41.7	
Not living with spouse/partner but living with children	20.8	16.1		18.2	18.0		18.3	20.7	
Living with spouse/partner and not living with children	20.0	19.3		23.0	19.9		22.0	18.2	
Living with spouse/partner and children	18.4	17.7		20.6	18.6		20.2	19.4	
Living with parents (%)	23.6	30.9	.066	23.0	28.6	.254	25.6	28.7	.524
One parent has at least some college (%)	50.0	53.0	.502	49.7	50.3	.912	50.6	48.7	.735
High School Grades (%)			.245			.122			.098
Mostly A's	6.0	5.6		6.1	8.1		5.6	5.9	
Mostly B's	46.0	39.0		49.1	37.9		49.0	37.5	
Mostly C's or below	48.0	55.4		44.9	54.0		45.4	56.7	

Characteristic	All Participants			Survey Respondents, Unweighted			Survey Respondents, Weighted		
	Treatment	Control	p-Value	Treatment	Control	p-Value	Treatment	Control	p-Value
Current Education (%)			.195			0.090			.319
Less than high school	4.8	1.2		6.1	1.2		6.2	1.6	
High school or equivalent	44.8	43.4		40.0	46.0		41.4	42.6	
Less than one year of college	24.4	25.7		26.1	20.5		25.7	25.4	
One or more years of college	20.4	22.5		21.8	22.4		21.5	22.2	
Associate degree or higher	5.6	6.8		6.1	9.3		5.2	7.5	
Received vocational or technical certificate or diploma (%)	41.2	38.2	.488	46.7	41.6	.360	40.6	42.6	.716
Career Knowledge Index (average of items)	0.43	0.39	.180	0.45	0.40	.210	0.42	0.39	.427
Psycho-Social Indices									
Academic Discipline Index	5.05	5.07	.728	5.06	5.09	.636	5.01	5.07	.451
Training Commitment Index	5.60	5.58	.716	5.61	5.57	.516	5.59	5.58	.896
Academic Self-Confidence Index	4.45	4.39	.419	4.47	4.42	.621	4.42	4.41	.891
Emotional Stability Index	5.03	4.95	.265	5.04	4.97	.455	5.00	4.97	.713
Social Support Index	3.30	3.32	.488	3.32	3.32	.995	3.31	3.30	.840
Stress Index	2.21	2.24	.562	2.14	2.24	.184	2.21	2.25	.571
Depression Index	1.52	1.57	.267	1.49	1.55	.306	1.53	1.55	.731
Income (%)			.027			.258			.172
Less than \$15,000	26.8	23.7		24.2	25.5		25.0	24.9	
\$15,000-29,999	34.0	26.1		36.4	28.0		37.2	28.5	
\$30,000+	39.2	51.4		40.0	48.5		38.9	48.5	
Mean (\$)	31,535	34,990	.180	32,452	34,446	.553	32,355	34,227	.571
Public Assistance / Hardship Past 12 Months (%)									
Received WIC or SNAP	32.4	38.6	.151	33.3	41.0	.153	34.5	45.4	.045
Received public assistance or welfare	4.4	4.0	.831	3.0	5.0	.373	3.1	4.3	.552
Reported financial hardship	35.6	33.7	.662	33.9	37.3	.532	35.6	38.4	.607

Characteristic	All Participants			Survey Respondents, Unweighted			Survey Respondents, Weighted		
	Treatment	Control	p-Value	Treatment	Control	p-Value	Treatment	Control	p-Value
Current Work Hours (%)			.880			.484			.480
0	26.8	28.9		27.9	28.6		29.5	31.5	
1-19	11.6	11.3		10.9	13.7		10.4	14.1	
20-34	32.0	32.9		29.7	33.5		30.7	32.2	
35+	29.6	26.5		31.5	24.2		29.3	22.2	
Expected Work Hours in Next Few Months (%)			.132			.303			.277
0	16.0	20.9		17.0	22.4		17.3	24.9	
1-19	18.0	11.7		17.0	11.2		16.1	11.0	
20-34	48.8	47.4		50.9	48.5		51.6	47.6	
35+	17.2	20.1		15.2	18.0		15.1	16.6	
Life Challenges Index (average in original units 1-5)	1.45	1.43	.708	1.40	1.45	.342	1.41	1.44	.621
Owns a car (%)	87.2	83.1	.202	88.5	80.1	.038	87.9	83.9	.303
Has both computer and internet at home (%)	84.4	83.1	.702	87.3	81.4	.143	88.0	82.5	.166
Ever arrested (%)	24.0	17.7	.082	24.9	16.8	.073	25.4	17.2	.070
Sample Sizes	250	249		165	161		165	161	

Key: BIF (PACE Basic Information Form). SAQ (PACE Self-Administered Questionnaire).

Source: BIF, SAQ, and response status to the PACE three-year follow-up survey. Sample restricted to those randomized for the evaluation of Bridge to Employment.

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 B.3.2 below. Significant imbalances are highlighted in red, using a threshold for statistical significance of 10 percent.

Exhibit C-10 presents evidence about the level of nonresponse bias with and without adjustment weights. The first two panels of Exhibit C-10 compare three sets of regression-adjusted impacts on earnings outcomes from NDNH records. The third and fourth panels compare these sets of impacts on educational outcomes from school records. The first set of impact estimates (column 1) is based on the full sample. The second set (column 3) excludes survey nonrespondents. Differences between the first and second set of impacts signal nonresponse bias. The third set of impacts (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.

Exhibit C-10: Comparison of Selected Estimates of the Impact of PCPP for the Unweighted and Weighted Survey Samples

Outcome (Data Source)	Full Sample		Unweighted Sample		Weighted Sample	
	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Confirmatory Outcome (NDNH)						
Quarterly earnings (average of 12th and 13th quarters after randomization) (\$)	-175	313	-618	403	-536	403
Exploratory Outcomes (NDNH)						
Q5 earnings (\$)	-322	254	-331	333	-366	337
Q9 earnings (\$)	-275	292	-528	367	-560	363
Q13 earnings (\$)	-118	340	-453	435	-379	433
Q17 earnings (\$)	-539	412	-1070	532	-962	546
Any earnings Q5 (%)	-2.9	3.8	-4.7	4.8	-5.8	4.9
Any earnings Q9 (%)	-3.4	3.7	-7.3	4.6	-8.1	4.8
Any earnings Q13 (%)	-4.7	4.8	-2.4	4.2	-3.0	4.3
Any earnings Q17 (%)	-7.3	4.6	-6.0	4.4	-6.1	4.5
Confirmatory Education Outcome (School Records)						
Received 1+ year college credential	-3.4	2.1	-5.9	3.0	-2.6	2.7
Secondary Education Outcomes (School Records)						
Number of college credits	1.4	1.7	0.7	2.2	0.4	2.2
Full-time-equivalent months enrolled at colleges (months)	1.0*	0.7	0.5	0.9	0.6	0.9
Receipt of any credential from a college (%)	3.9	3.7	-0.5	4.9	4.0	4.9
Secondary Employment Outcomes (Survey)						
Employed at survey follow-up (%)			-0.1	4.9	0.7	5.1
Employed at \$13 per hour or above (%)			9.7**	5.5	10.8**	5.6
Employed in a job requiring a least mid-level skills (%)			-9.5	4.2	-6.8	4.4
Secondary Education Outcomes (Blended Three-Year and 18-Month Surveys)						
Receipt of an exam-based certification or license (%)			-4.0	5.5	-2.3	5.6

Outcome (Data Source)	Full Sample		Unweighted Sample		Weighted Sample	
	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Other Secondary Outcomes (Survey)						
Indicators of Independence and Well-Being						
Health insurance coverage (%)			6.2**	3.5	8.3**	3.7
Receives public benefits (%)			4.3	5.2	4.7	5.3
Personal student debt (\$)			1244	1284	940	1325
Financial hardship (%)			-0.3	5.4	-0.2	5.5
Sample size (treatment + control group)	499		326		326	

Source: NDNH, NSC, PACE three-year follow-up survey, and PACE 18-month follow-up survey.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.3. The full sample columns are blank for survey-measured outcomes as they are not available for the full sample.

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 summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

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 appears to be evidence of bias in estimated impacts based on the unweighted respondent sample. However, that bias does not change any substantive conclusions. For example, the estimated impact on the average of quarterly earnings for the 12th and 13th quarters is sizeable and negative, though not statistically significant (-\$618), whereas for the full sample the estimated impact remains negative but is closer to zero (-\$175).

Given the aforementioned evidence of nonresponse bias, as well as the centrality of earnings in the logic models for how PACE programs would affect a wide variety of life outcomes measured in the survey, some survey nonresponse adjustment is required. We implemented this solution across all nine PACE sites.

The final pair of columns shows that the nonresponse weights generally bring impact estimates based only on survey respondents back into better alignment with impact estimates on the full sample. For example, while still not statistically significant, the estimated impact on quarterly earnings in the 12th and 13th quarters for the weighted survey sample is -\$536, which is closer to the full sample estimate (-\$618) than the unweighted estimate. This illustrates how the nonresponse weights removed some of the bias in the unweighted survey sample. The weighted impacts do not agree exactly with the full-sample impacts, but that would be an unreasonable goal for an adjustment procedure. Altogether, the weights reduced nonresponse bias for half of the NDNH outcomes.

For the survey-based outcomes, the fifth, sixth, and seventh panels of Exhibit C-10 compare the unweighted and weighted impact estimates. There are only minor differences in the magnitude of the estimates and no differences in statistical significance. This is consistent with the NDNH and education outcome findings, which found no differences in inference between the full sample, unweighted, and weighted impact estimates.

C.3.2 Construction of Nonresponse Adjustment Weights

Construction of weights to reduce the biases just discussed was more complex than anticipated. At first, we tried a standard propensity scoring approach,³⁶ as was used in the short-term report on PCPP (Cook et al. 2018). However, that approach was not successful in removing the biases shown in the unweighted column of Exhibit C-10. 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 tabulations by sex, tabulations by race, and tabulations by age all to agree with pre-specified totals for sex, race, and age. In this example, sex, 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.

³⁶ 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.

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.³⁷ 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 et al. (2000). It was necessary to use both of these software packages because the analyses had to be run on two servers, one that had SUDAAN installed (at Abt) and one that did not (ACF server). 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.

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 Q12.³⁸ We verified that these weights removed most of the nonresponse bias on estimates of program impacts on

³⁷ 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.

³⁸ 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.

NDNH earnings 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. (Exhibit C-10 shows the results.) These 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. (Exhibit C-10 above shows the results.)

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 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. We applied this procedure for all nine PACE sites.

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

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, argued against the latter. The PCC study offered college records to support the analysis, making it a good choice for investigating these survey outcomes.

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 school-issued credentials in all PACE sites where we used the survey rather than college records to measure educational progress.

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 loans programs but can be a useful tool for education researchers. In this report, we used the NSC to impute college outcomes for people who attended colleges other than Madison College, as discussed in Appendix Section B.2, and to prepare alternate estimates of the impacts of PCPP. Section D.1 summarizes statistics on NSC coverage. Section D.2 provides details on how raw data from NSC were recoded to make them more relevant to the evaluation of PCPP. Finally, Section D.3 presents estimates of PCPP 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 does not cover schools that are not Title IV schools, the set of schools approved for federal student loans by the U.S. Department of Education. Moreover, although 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 Type, 2012 through 2016

Type and Control of College	Year			
	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: NSC 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.³⁹ 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 NSC in October of 2018 covering enrollment through Quarter 18 for all 499 study participants (250 in the treatment group and 249 in the control group of the study).

Records from 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.⁴⁰ To translate these to person-quarter-level outcomes, a student was counted as enrolled for the quarter if they were enrolled in any of the three months, and FTE enrollment was calculated by summing their total FTE months for the quarter.

D.3 Program Impacts on NSC-Measured Outcomes

Exhibit D-2 compares a selection of estimated impacts of PCPP using both NSC records and adjusted Madison College records.⁴¹ 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 receipt of particular types of credentials), but it also meant that we were not capturing progress for about 14 percent of students.

For enrollment outcomes—including enrollment by quarter and total FTE months enrolled—both the levels and impacts are broadly consistent across the two records systems. The adjusted

³⁹ 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 et al. (2015) report that only about 80 percent of degrees from Michigan colleges were reported to the NSC in the 2008-2010 period.

⁴⁰ 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.

⁴¹ Refer back to Section B.2 of Appendix B for details on the adjustment of Madison College records.

Madison College records generally show a slightly larger impact than the NSC records, although the differences in impacts are not statistically significant.⁴²

The two systems are much less consistent about recording of completions (i.e., college certificates and degrees). For some unknown reason, NSC reports many fewer completions than we found in local college records. One possibility is the reporting of short-term college certificates. As noted in Chapter 3, most of the college certificates received by PCPP participants were short-term (such as Nursing Assistant), and may not have been included in data received from NSC. Impacts on completion are different for the two sources (0 percentage points for NSC records and 4 percentage points for Madison College Records), but the difference between these estimated impacts is not statistically significant.

⁴² For two of the outcomes, the different data sources do lead to different significance results: the Madison College records show a positive impact on enrollment in quarter 12 of 7.5 percentage points, and an impact on FTE months of college enrollment of one month, each of which are significant at the ten percent level; the impacts on the equivalent NSC-based measures are not significant.

Exhibit D-2: Comparisons of Impacts of PCPP Based on Madison College Records with Impacts Based on NSC Records

Outcome	NSC Records				Madison College 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 Quarter 4	63.5	54.6	+8.8**	4.4	71.2	57.8	+13.4***	4.2	-4.5	2.9
In Quarter 8	43.5	42.2	+1.4	4.5	47.1	41.8	+5.4	4.5	-4.0	3.1
In Quarter 12	36.3	32.9	+3.4	4.4	40.4	32.9	+7.5*	4.4	-4.1	3.4
In Quarter 16	33.8	28.1	+5.7	4.2	34.1	27.3	+6.8	4.2	-1.1	3.3
Cumulative Number of FTE Months of College Enrollment:										
Through Quarter 12	9.9	9.3	+0.5	0.7	10.4	9.4	+1.0*	0.7	-0.5	0.4
Any Completions from a College (%):										
Through Quarter 12	9.3	9.2	+0.0	2.6	24.4	20.5	+3.9	3.7	-3.9	3.4
Sample size	250	249			250	249				

Source: Numbers in this table are based on college records from the NSC or from Madison College records.

Note: Statistical significance is based on two-tailed tests.

Statistical significance levels are summarized as follows: *** at the 1 percent level; ** at the 5 percent level; * at the 10 percent level.

Appendix E: Sensitivity Analyses of Education Impacts

This report used local Madison College 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 follow-up survey data. Across the confirmatory outcome and three secondary outcomes, there are some differences in both the magnitude and direction of impact; however, none of these differences is statistically significant.

Exhibit E-1 shows how the impacts differ between the two systems. On the confirmatory outcome—*receipt of a college credential that typically takes at least a year of study*—the three-year follow-up survey data showed that the treatment had a +2 percentage point impact on the confirmatory outcome, whereas the adjusted Madison College records data showed a -3 percentage point impact. Though these impacts are in opposite directions, neither of them is statistically significant. Furthermore, the difference between these two estimates was also not statistically significant. The differences in the secondary outcomes follow a similar pattern—none of the differences between the estimates is statistically significant.

In addition, Appendix D also found broad consistency between impacts based on National Student Clearinghouse records and impacts based on adjusted Madison College records. Taken together, the conclusions are broadly consistent across all three data sources—Madison College, follow-up survey, and National Student Clearinghouse.

Exhibit E-1: Comparisons of Impacts of PCPP Based on Madison College Records with Impacts Based on the Three-Year Follow-up Survey

Outcome	Three-Year Follow-up Survey				Madison College 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 1+ year college credential (%)	10.2	8.0	+2.2	3.1	4.3	7.6	−3.4	2.1	+5.6	3.8
Secondary										
Number of college credits	19.4	22.5	−3.2	3.7	20.1	18.7	+1.4	1.7	−4.6	4.1
FTE months enrolled in college (#)	10.9	10.5	+0.4	1.2	10.4	9.4	+1.0*	0.7	−0.6	1.4
Received any college credential (%)	23.5	23.6	−0.1	4.8	24.4	20.5	+3.9	3.7	−4.0	6.1
Sample size	165	161			250	249				

Source: PACE three-year follow-up survey and adjusted Madison College records.

Note: Statistical significance is based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes, unless otherwise noted.

^a Statistical significance for Difference in Impacts is based on two-tailed test.

Statistical significance levels are summarized as follows: *** statistically significant at the 1 percent level; ** at the 5 percent level; * at the 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. It 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.⁴³

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,⁴⁴ OCSE may disclose certain information contained in the 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 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 (MOU) allowing access to NDNH data for the PACE evaluation. Among other provisions, the MOU dictates what self-reported data from study subjects may be

⁴³ 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>

⁴⁴ 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.

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.⁴⁵ (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 in order to estimate program impacts on earnings and employment.

F.2 Data and Measures

Random assignment for PCPP started in December 2011 and ended in January 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 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 19 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 499 treatment and control group members randomized as part of the PCPP evaluation, 486 study participants reported a name and SSN that OCSE deemed to be of sufficient quality for its matching purposes.⁴⁶ 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.

⁴⁵ 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.

⁴⁶ 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.

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.⁴⁷

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 24 measures for each person (employment and earnings for the four quarters before randomization, the quarter of randomization, and the 9 quarters after randomization). In addition, we formed a quarterly average earnings for Q12 and Q13 after random assignment (the confirmatory earnings outcome, established to align with the theory of change) and an annual average for Q10-Q13.

⁴⁷ Topcoding means values above a threshold are set equal to the threshold.

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

Barnow and Greenberg (2015) review findings from evaluations including both the NDNH and surveys as data sources. 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). In the evaluation of PCPP, average quarterly earnings agree rather well between the two measurement systems, but correlational analysis shows that there is some measurement noise in one or both. The correlation in person-level quarterly earnings between the two systems at Q12 is 0.63 for the treatment sample and 0.76 for the control sample.⁴⁸ Earnings from self-employment appear to explain part of the lower correlation on the treatment sample. The difference between self-employment and NDNH-reported earnings has a correlation of 0.11 for the treatment sample, compared to -0.01 for the control sample.

This section compares estimates of employment and earnings impacts based on NDNH data and survey self-reports.⁴⁹ It also presents estimates of the impact of PCPP on self-employment earnings.

The top panel in Exhibit G-1 shows the degree of agreement of impact estimates for PCPP derived from the two sources. The estimated impact based on UI records of -\$233 for average earnings in Q12 is smaller than the estimated impact of +\$137 for Q12 based on three-year follow-up survey data. However, the difference between the two estimates is not statistically significant.⁵⁰ We explored whether earnings from self-employment could explain the difference between -\$233 and +\$137 if we were to treat the difference as real, but earnings from self-employment are too small to explain all of the difference. It could be that the difference is just due to random memory errors by respondents.

Another plausible contributing cause to the discrepancy is differential undercoverage in the NDNH. Barnow and Greenberg (2015) noted that state UI tax databases do not cover federal workers; out-of-state records; most workers at small farms, at railroads, at selected nonprofit organizations (particularly churches); and casual or irregular jobs. Hiding of tip income and income from household employment (such as childcare and cleaning) are additional important

⁴⁸ 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).

⁴⁹ 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.

⁵⁰ Using the average correlation for the earnings, the standard error between the two estimated impacts is \$343, which is roughly the size of the difference between the two impact estimates.

sources of undercoverage. In some states, independent contractors are not included. NDNH remedies the undercoverage of federal workers and of out-of-state workers, but not the other causes of undercoverage. 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. However, because the NDNH-estimated impact is not statistically significant and the difference between the NDNH- and survey-estimated impacts is also not statistically significant, there seems to be no reason to be concerned about this issue.

The second panel of Exhibit G-1 shows that the levels of NDNH-based employment are slightly higher than survey-based estimates for both treatment group members (82 compared with 75 percent) and control group members (83 compared with 74 percent). Most of the difference is probably due to the time frame. The percentage with any earnings over three months is bound to be higher than the percentage employed on a particular day. However, neither measurement strategy estimates a statistically significant impact on employment.

**Exhibit G-1: Impacts of PCPP 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 (\$)	4,355	4,587	-233	319
Self-reported earnings in Q12 (\$)	4,709	4,572	+137	475
Self-reported earnings from self-employment in Q12 (\$)	19	37	-18	43
Employment				
Average percentage with employer-reported wages in Q12	81.6	83.1	-1.4	3.6
Percentage working in the week prior to survey interview	75.1	74.4	+0.7	5.1
Sample sizes				
NDNH	250	249		
Survey	165	161		

SOURCES: NDNH and 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.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

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 .⁵¹

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, Diehr, Emerson, and Chen (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 NDNH earnings. We found no values that were clearly impossible, and thus retained all reported values in the analysis.

⁵¹ 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 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. Doi:10.1177/0193841X14564154
- 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. <https://bmcmmedresmethodol.biomedcentral.com/articles/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.
- Cook, Rachel, Jill Hamadyk, Matthew Zeidenberg, Howard Rolston, and Karen Gardiner. 2018. *Madison Area Technical College Patient Care Pathway Program: Implementation and Early Impact Report*. OPRE Report 2018-48. 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/patient-care-pathway-program-pcpp>.
- Cutrona, C., and D. Russell. 1987. "The Provisions of Social Relationships and Adaptation to Stress." *Advances in Personal Relationships*, 1.
- 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., & Shapiro, D. 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., Hemelt, S. W., & Hyman, J. M. 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/sites/default/files/opre/bridge_to_employment_implementation_and_early_impact_report_final_pdf.pdf.
- Fein, D. 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/sites/default/files/opre/cp_as_a_framework_final_508b.pdf.
- 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, Section on Survey Research Methods*, 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.
- 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](https://doi.org/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.
- 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, 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.
- 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.
- 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, 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 R., and Kristin E. 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, 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>.

- 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](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>.
- 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.
- 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>.

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, Section on Survey Research Methods*, 406-411. Alexandria, VA: American Statistical Association.