

**Pathways for
Advancing
Careers and
Education (PACE)
Evaluation
Design Report**

November 2014

Prepared for:

Brendan Kelly

Office of Planning, Research
and Evaluation

Administration for Children and
Families

US Department of Health and
Human Services

Submitted by:

Abt Associates Inc.

4550 Montgomery Avenue
Suite 800 North
Bethesda, MD 20814

PACE
Pathways for Advancing
Careers and Education



This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Abt Associates, Inc. (2014). Pathways for Advancing Careers and Education Evaluation Design Report. OPRE Report # 2014-76, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Submitted to:

Brendan Kelly, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Contract Number: HHSP23320072913YC

Project Director:

Karen Gardiner
Abt Associates, Inc.
4550 Montgomery Ave.
Bethesda, MD 20814

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 <http://www.acf.hhs.gov/programs/opre/index.html>.



Table of Contents

1. Evaluation Framework.....	1
1.1 Overview of PACE.....	2
1.2 Career Pathways Framework.....	3
1.3 Career Pathways Theory of Change.....	7
1.3.1 Main (Distal) Outcomes.....	8
1.3.2 Intermediate (Proximal) Outcomes.....	8
1.3.3 Contextual Factors.....	8
1.3.4 A Framework for Evaluation.....	9
1.4 Programs Evaluated in PACE.....	9
1.4.1 Process for Selection.....	9
1.4.2 Overview of PACE Programs.....	10
1.4.3 Similarities and Differences across Programs.....	14
1.5 Study Design.....	15
1.6 Major Evaluation Substudies and Reports.....	16
1.6.1 Substudies.....	16
1.6.2 Reports.....	17
1.7 PACE and Related ACF Studies.....	18
1.8 Organization of the Report.....	19
2. Implementation Study.....	21
2.1 PACE Implementation Study Research Questions.....	21
2.1.1 What is the design and underlying logic of each program?.....	22
2.1.2 What was the program like as actually implemented? What was the degree of fidelity to the design? Did the program change during the study period and, if so, why?.....	25
2.1.3 What are the experiences of treatment group members?.....	26
2.1.4 What are treatment-control group differences in service receipt?.....	27
2.1.5 What are lessons for the field in terms of scalability and sustainability?.....	28
2.2 Data Sources for the Implementation Study.....	29
2.2.1 Interviews with Program Staff and Partner Organizations.....	31
2.2.2 Program Documents.....	32
2.2.3 Observations.....	32
2.2.4 Program Administrative Data.....	32
2.2.5 Baseline Information and first follow-up surveys of Treatment and Control Group Members.....	33

2.2.6	Instructor, Manager, and Case Manager/Advisor Surveys.....	33
2.2.7	In-Depth Interviews with Treatment and Control Group Members	34
2.2.8	Reporting Implementation Study Findings	35
3.	Impact Study for an Individual Program	41
3.1	Research Questions.....	41
3.2	Time Horizon for Impact Analyses	42
3.3	Classification of Research Hypotheses for Testing and Reporting Purposes.....	43
3.3.1	Confirmatory Hypotheses, Multiple Comparison Adjustment and Composite Outcomes.....	45
3.3.2	Secondary Hypotheses	46
3.3.3	Exploratory Hypotheses.....	50
3.4	Basic Methods for Estimating Impacts.....	51
3.4.1	Inferential Frame	51
3.4.2	The Use of Models	51
3.4.3	Covariate Selection	52
3.4.4	Standards for Statistical Significance.....	53
3.4.5	Presentation of Impact Findings.....	54
3.5	Minimum Detectable Effects (MDEs).....	55
3.5.1	MDEs for Tests of PACE Programs	55
3.6	Subgroup Analysis and Other Types of Effect Moderation.....	56
3.7	Testing Pathways in the Theory of Change.....	58
3.8	Other Issues for the Impact Study	65
3.8.1	Multiple Sites in the Year Up and I-BEST Programs.....	65
3.8.2	Provide Estimates for Treatment on Treated (TOT) Impacts	65
3.8.3	Conditionally Observed Outcomes	67
3.8.4	Missing Data	69
3.8.5	Comparison of Findings from Mediation Analysis across Programs	71
3.9	Cost-Benefit Study	71
3.9.1	General Approach.....	71
3.9.2	Benefits, Costs and Perspectives	72
3.9.3	Net Benefit and Cost Calculations	75
3.9.4	Calculation and Analysis of Individual Benefits and Costs	75
3.9.5	Open Challenges.....	77
3.10	Program-Specific Analysis Plans and Some Open Issues	77
4.	Data Collection.....	78

4.1	Baseline Data.....	79
4.1.1	Baseline Surveys	79
4.1.2	Skills Assessments.....	80
4.1.3	Data Collected in Program Applications.....	82
4.2	Follow-up Surveys	82
4.2.1	Timing and Target Populations.....	82
4.2.2	Content	83
4.2.3	Administration	84
4.3	Data from Administrative Systems	84
4.3.1	PACE Program Data	84
4.3.2	Data from College Record Systems	87
4.3.3	Wage Records from the Unemployment Insurance (UI) Reporting System.....	89
4.4	Other Possible Data Sources	90
	References	91
	Appendix A: Analysis of Mediation in PACE using Acyclic Graphs.....	98

1. Evaluation Framework

Millions of adults lack the postsecondary education and occupational training needed to obtain jobs that provide good wages, benefits and pathways to advancement. Over the past three decades, earnings for those with high school diplomas or less education fell compared to those who had more education. At the same time, the share of jobs that require college credentials continued to grow. According to the Department of Labor, by 2018, 63 percent of job openings will require workers with at least some college education.¹

There is longstanding interest among policy makers and program operators in finding ways to increase the skill levels of low-income individuals, improve their enrollment in and completion of postsecondary education, and improve their economic prospects. A number of factors, however, have limited the success of efforts to date. Many low-income adults face challenges to postsecondary education enrollment and completion, including limited basic academic skills; limited academic or training goals due to negative school experiences and lack of college role models; work and family demands on time; inability to afford school; and stress and other issues associated with poverty. At a broader level, many postsecondary education systems are not geared towards non-traditional students, including low-income adults. For example, they have weak basic skills programs; an emphasis on long-term programs and general education degrees; fragmented and, at times, deficient academic advising and student support services; complex financial aid rules; and limited financial assistance. Supports from social services and workforce systems, which might assist these students, can be limited and/or difficult to coordinate.²

The career pathways approach is gaining steady acceptance as a promising strategy to address these challenges and improve postsecondary education and training for low-income and low-skilled adults.³

¹ Carnevale, A., Smith, N., & Strohl, J. (2010).

² See reviews of evidence on traditional remedial instruction in Bailey (2009), Grubb (2001), Hughes and Scott-Clayton (2011), and Kazis and Leibowitz (2003). On fragmentation and complexity of curricula, financial aid and other community college systems, see Rosenbaum et al. (2006), Scott-Clayton (2011), and Goldrick-Rab and Sorenson (2010). On systemic problems more generally, see Alssid et al. (2005), Brock and LeBlanc (2005), Council of Economic Advisors (2009), Jenkins (2006), and Pleasants and Claggett (2010). On low completion rates among community college students, particularly nontraditional and economically disadvantaged students, see discussions in Cooper (2010), Goldrick-Rab and Sorenson (2010), Purnell and Blank (2004), and Visher et al. (2008).

³ Three federal departments are funding career pathways research. The Departments of Labor and Education, for instance, launched a one-year Career Pathways Initiative in 2010, which provided funding to nine states and two tribal entities to develop sustainable career pathways to promote linkages among system partners. The Department of Labor produced a set of technical assistance tools for state, local, and tribal policymakers to use in development and implementation of career pathways approaches. The Department of Education's Office of Vocational and Adult Education funds the Designing Instruction for Career Pathways initiative. The

Career pathways programs are intended to improve the education and earnings of low-skilled adults by providing well-articulated training steps tailored to jobs in demand locally along with guidance and other supports. Although there is some research evidence on selected components of career pathways programs, to date, there has been no rigorous research on the overall effectiveness of this approach.

The Pathways for Advancing Careers and Education (PACE) project is a major national effort to evaluate the effectiveness of nine career pathways programs. PACE is funded by the U.S. Department of Health and Human Services Office of Planning, Research and Evaluation (OPRE) within the Administration for Children and Families (ACF) and conducted by a team led by Abt Associates.⁴

This evaluation design report describes the career pathways framework, the major study components, and study data sources. This chapter begins with an overview of PACE and the career pathways framework. It then describes the program selection process, the sites, and research questions. It concludes with the study timeline and deliverables, and an outline of the remainder of the report. Subsequent chapters describe the approach to the implementation study (Chapter 2), the impact study (Chapter 3), and data sources (Chapter 4). Program-specific analysis plans will be released as supplements to this design report.

1.1 Overview of PACE

In 2007, ACF launched the Pathways for Advancing Careers and Education (PACE) study,⁵ a ten-year random assignment evaluation of nine promising career pathways interventions aimed at increasing employment and self-sufficiency among low-income, low-skilled adults and youth. PACE utilizes an experimental study design to assess the impact of these interventions on educational attainment, employment and earnings, and other outcomes. In addition to the impact study, the evaluation includes implementation and cost-benefit studies. The goal is to produce methodologically rigorous evidence on the effectiveness of selected career pathways approaches that will address issues of interest to federal, state, and local policymakers and practitioners and have significant influence on policy and practice.

ACF conceived PACE as a next-generation study of promising approaches to promoting self-sufficiency for economically disadvantaged adults but did not specify the area of study. Following an in-depth

Department of Health and Human Services funds a career pathways research portfolio that includes PACE and the Health Profession Opportunity Grant-related impact and implementation studies.

⁴ The study team includes MEF Associates, The Urban Institute, George Washington University, the American Public Human Services Association, the National Conference of State Legislatures, and the National Governors Association. Other project partners are the Open Society Foundations (OSF) Special Fund for Poverty Alleviation, the Kresge Foundation, the Joyce Foundation, the Meadows Foundation, the Hearst Foundations, the Laura and John Arnold Foundation, and the JP Morgan Foundation. Additionally, some participating programs are funded by ACF Health Profession Opportunity Grants (HPOG).

⁵ From the study's inception in 2007 until November 2014 the study was known as Innovative Strategies for Increasing Self-Sufficiency.

knowledge development process that included outreach to over 250 diverse stakeholders to learn directions of interest and review of the evaluation literature, the PACE team recommended a focus on career pathways programs.⁶ At the heart of the career pathways framework guiding the evaluation design is the idea that postsecondary training should be based on “...a series of connected education and training programs and support services that enable individuals to secure employment within a specific industry or occupational sector, and to advance to successively higher levels of education and employment within that sector. Each step...is designed explicitly to prepare the participant for the next level of employment and education.”⁷ Each step is an opportunity for pre-college or college-level students to access postsecondary education and training and, as their skills improve, move to successively higher levels of training and employment. Students who begin with skills at the college level are able to start training at a level matched to their ability.

The career pathways framework’s developers and advocates sought to identify and address shortcomings of the postsecondary education system that impede completion by adult learners. As noted above, these include ineffective approaches to remedial education, an emphasis on longer programs and general education degrees, fragmented and inadequate academic advising and student support services, complex course selection systems and financial aid rules, and insufficient financial assistance.⁸ The PACE evaluation team recruited nine programs that, in varying ways, embodied key elements of the career pathways framework.

1.2 Career Pathways Framework

The career pathways framework aims to provide principles for change in wider systems—including national, state and local postsecondary, workforce, and human services policies and programs—as well as a template for program design. PACE is investigating how the framework can be used to classify and guide evaluation of programs.

Exhibit 1.1 shows five general levels of training and employment often represented as key steps in the literature on career pathways programs and systems.⁹ The bottom two steps (I and II) represent “on ramp” and “bridge” programs designed to prepare low-skilled participants for college-level training and

⁶ PACE, 2009

⁷ Jenkins, 2006

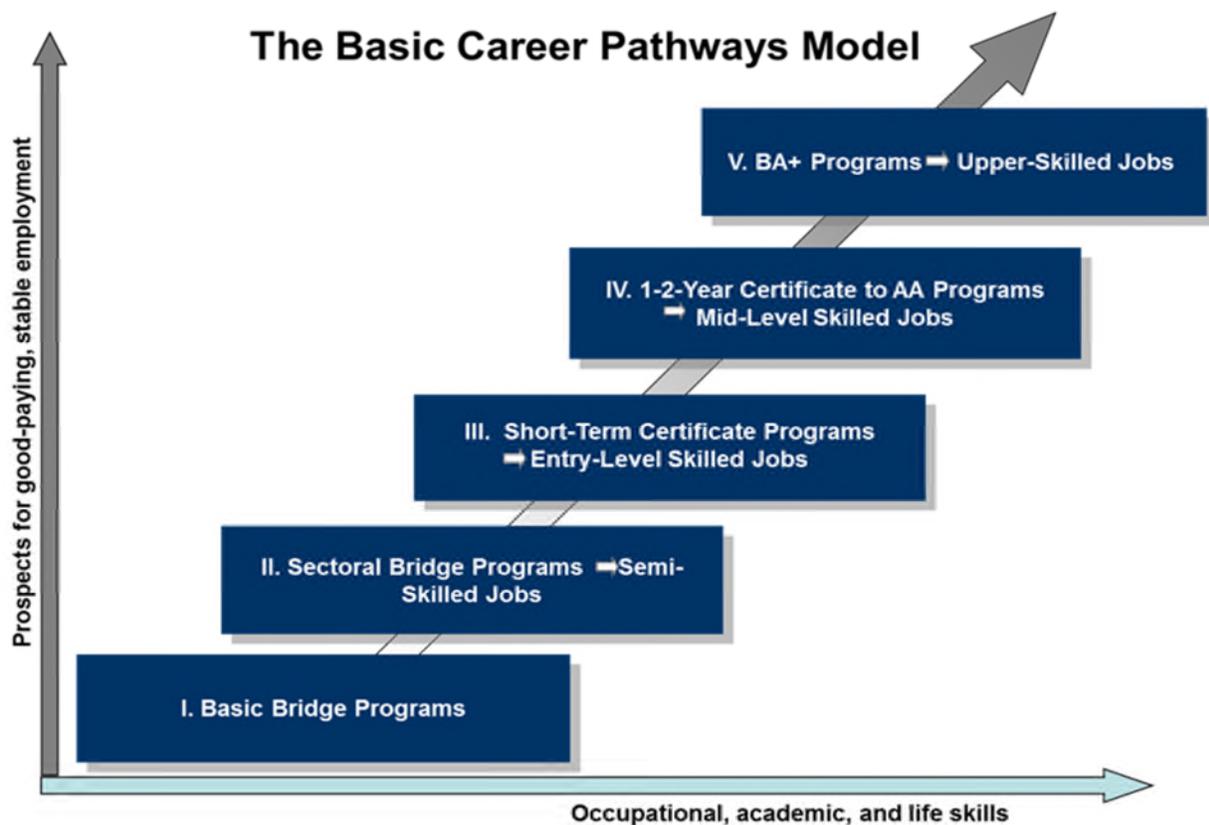
⁸ See reviews of evidence on traditional remedial instruction in Bailey (2009), Grubb (2001), Hughes and Scott-Clayton (2011), and Kazis and Leibowitz (2003). On fragmentation and complexity of curricula, financial aid and other community college systems, see Rosenbaum et al. (2006), Scott-Clayton (2011), and Goldrick-Rab and Sorenson (2010). On systemic problems more generally, see Alssid et al. (2005), Brock and LeBlanc (2005), Council of Economic Advisors (2009), Jenkins (2006), and Pleasants and Claggett (2010). On low completion rates among community college students, particularly nontraditional and economically disadvantaged students, see discussions in Cooper (2010), Goldrick-Rab and Sorenson (2010), Purnell and Blank (2004), and Visher et al. (2008).

⁹ Exhibit 1.1 is an adaptation of basic levels depicted from the Wisconsin RISE program. See Strawn (2010).

lower-skilled jobs with career progression potential. Bridge programs have been developed to serve adults with skill levels in the 9th-11th grade equivalent ranges, and usually lead to college-level training (Step III) (US Department of Education, 2011). In some cases, bridge programs serve adults with skills levels in the 6th to 8th grade equivalent ranges and serve as an “on ramp” to sectoral bridge programs (Step II). Steps III and IV provide college-level training for “middle skills” employment—jobs requiring some college but less than a bachelor’s degree (e.g., an associate’s degree or shorter certificate). Step V includes interventions promoting completion of bachelors’ degrees and more advanced credentials.

Career pathways are designed to allow entries, exits, and re-entries at each step—depending on skill levels and prior training, employment, and changing personal situations. Each step is designed to incorporate core program strategies (described below). The program strategies involve partnerships between multiple organizations, including community-based organizations, community colleges and other postsecondary training providers, human services and workforce agencies, and employers. Programs also emphasize partnerships within institutions, such as between community college departments.

Exhibit 1.1: Career Pathways Model



To engage, retain, and facilitate learning among low-skilled adults, the career pathways framework includes four categories of service strategies: (1) assessments of skills and needs; (2) promising and innovative approaches to basic skills instruction and occupational training (“core curriculum”); (3) academic and non-academic supports to promote success; and (4) approaches for connecting students with career-track employment opportunities. Within each of these categories, a variety of strategies

have emerged as emblematic, or signature, elements of promising approaches. Though there has been a trend to develop comprehensive programs inclusive of all of these strategies, the extent and ways in which programs include these strategies vary.

Assessment of student skills and needs. To identify student needs and factors that may facilitate or hinder academic success (and ultimately career advancement), the career pathways framework emphasizes assessment of a range of skills, strengths, and challenges. In the cognitive arena, assessments variously cover basic academic skills, learning styles, and learning and other disabilities. Non-cognitive domains cover an extensive array of intra- and inter-personal skills, as well as functional knowledge of how to manage well in college, at work, and in other important life tasks.

Strategies for providing basic academic skills instruction and occupational training. Career pathways programs include a growing array of promising approaches to instruction and curricular adaptations aiming to make education and training more engaging and manageable. Some of the major categories of these strategies include:

- *Well-articulated steps:* Career pathways coursework can be organized in smaller, distinct segments that can be obtained sequentially (“stacked”). Each step is designed to connect with additional training through alignment of content and organizational agreements to recognize students’ credits across modules. This approach is intended to encourage student persistence by providing quicker recognition of accomplishment (e.g., a certificate following one course or term).
- *Contextualization:* One approach to instruction is to teach academic basic skills in the context of an occupation or “real life” situation. Programs integrate content from occupational classes into basic academic skills instruction (e.g., occupation-specific materials and examples) or include basic skills instruction as part of occupational training classes. Contextualization is designed to increase motivation and understanding by increasing the relevance of what is learned.
- *Acceleration:* Programs are reorganizing curricula to enable students to earn certificates or credentials in a shorter (calendar) time period. Compressing total course or program hours into a shorter timeframe is intended to improve information retention between classes and reduce the time for outside issues to interfere with school.
- *Flexible delivery:* By offering instruction at convenient times and places and in formats that facilitate participation by adults who work and are parents, programs aim to encourage attendance and completion. Examples of flexible delivery are evening and weekend scheduling, self-paced instruction, easily accessible locations for instruction (e.g., in the community and not a central campus), and technology-supported distance learning.
- *Active learning:* Instructors use approaches that minimize traditional lecture formats and emphasize project-based learning and problem-solving tasks. This strategy also could involve group work and encouragement of classroom interaction.

Academic and non-academic supports: Academic and non-academic supports are thought to help program participants succeed in their current academic step and transition to and persist in subsequent steps. These supports aim to address gaps and deficiencies within existing support systems to assist a population that often has more extensive academic and personal challenges than traditional college students. Examples of supports are:

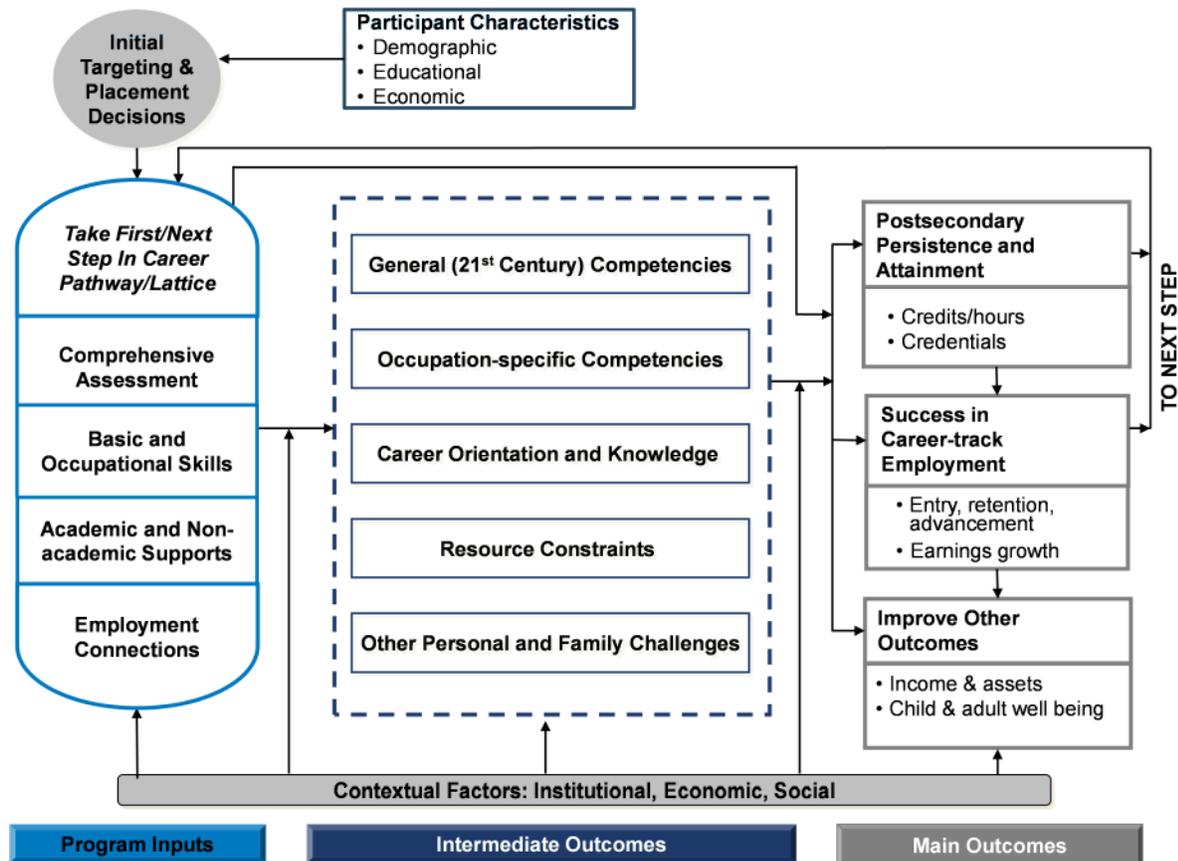
- *Personal guidance and supports:* Strategies for creating closer personal connections between program participants and staff include advising that addresses a wide range of academic and non-academic topics, such as advising on career planning and navigation, and coping with external life issues.
- *Instructional supports:* Strategies that address participants' academic needs include tutoring, study groups, and self-paced, computer-based instruction. Approaches for addressing participants' non-academic needs include supplemental workshops, courses, and support groups that emphasize skill development, effective communication, time management, and personal financial management.
- *Social supports:* The provision of social supports is designed to encourage social connections among program participants and among participants, faculty, and staff. Approaches include developing learning communities (e.g., maintaining student cohorts and consistent faculty-staff relationships), using peer and alumni mentors, and building support networks among program participants and among participants, faculty, and staff.
- *Supportive services:* This strategy involves providing services—in-house or through a referral network—to help participants cope with issues that might affect successful academic performance and school completion. Supportive services include child care, transportation assistance, substance abuse treatment, domestic violence prevention and amelioration, and mental health counseling and therapy.
- *Financial assistance:* The provision of financial assistance helps address program participants' financial needs and related stresses that can be barriers to their postsecondary participation and completion. This type of assistance includes payment for some or all academic expenses (e.g., tuition, fees), and assistance with the purchase of books, tools, or other course-related needs. In addition to direct financial assistance, other forms of support are reimbursement for supportive services (e.g., child care, transportation); provision of financial assistance with emergency needs; and award of performance-based stipends and scholarships.

Connections with career-track employment opportunities: Strategies for strengthening connections between education and work are the final element of career pathways framework. These include engaging employers and business groups as partners in designing programs or in teaching, and providing employment experiences integrated with occupational training (e.g., class projects that simulate on-the-job tasks or actual projects for local employers; visits to employers; job shadowing; internships and apprenticeships). These activities can include provision of post-training connections to employment (e.g., job development, search, and placement services).

1.3 Career Pathways Theory of Change

Based on an extensive literature review, the PACE team organized key hypotheses in career pathways into a general theory of change.¹⁰ The framework demonstrates connections between signature program elements or “inputs” and expected intermediate and longer-term outcomes. Exhibit 1.2 depicts the theory of change.¹¹ It shows how program inputs (described in the previous section) are expected to influence individuals’ career pathways outcomes. Longer-term (main) outcomes include persistence in and completion of postsecondary education, improved success in career-track employment, and improvements in other outcomes like income and assets. According to the theory, a number of short-term (intermediate) outcomes related to skills acquisition, increased knowledge, and other factors precede these main outcomes. The expected outcomes are based on a set of assumptions about the context in which the program operates. Main and intermediate outcomes are described below.

Exhibit 1.2: Theory of Change



¹⁰ Fein, 2012

¹¹ Based on Fein, 2012, with slight refinements

1.3.1 Main (Distal) Outcomes

In the career pathways framework, interventions are designed to improve employment outcomes. Main outcomes—the primary targets of change—include performance and persistence in education and training programs followed by completion and credential receipt leading to employment in high-demand occupational fields. Depending on the program or the step in the pathway, the credential might be a certificate, a one-year diploma, or a two- or four-year degree. Completion of a training step then leads to either the next step on the pathway or employment in the field of study. Participants' completion of training and attainment of credentials also leads to improved performance and advancement in jobs, bringing higher earnings and receipt of job benefits. Once employed, a participant can remain in a job or seek another, with goals of increasing earnings and advancing in the field (job security), or can return to the next training step on the pathway. An important rationale for improving low-income adults' education and earnings is to enhance their other life outcomes connected with income and self-sufficiency and, for parents, the well-being of their children. A more distal outcome is that higher incomes, benefits, and improved job opportunities may lead to improved psychological well-being of adults and enhance material aspects of daily living, increasing the quality of parenting, child care, and other resources available to children. Participants' economic outcomes also can contribute to local economic growth.

1.3.2 Intermediate (Proximal) Outcomes

In many instances, career pathways strategies seek to produce these main outcomes by improving a variety of general and occupation-specific skills, as well as by fostering career awareness and direction and addressing material and other circumstances that can impede success in school and work.

General skills—often referred to as 21st Century skills—include a variety of cognitive (literacy, numeracy, critical thinking), intrapersonal (core self-evaluation, work ethic/conscientiousness, self-regulation, meta-cognition), and interpersonal (teamwork, collaboration, leadership) skills.¹²

Other intermediate outcomes include development of career goals and knowledge (e.g., participants' ability to navigate the norms and expected behaviors of college and work settings) through advising and career navigation supports, instruction in skills for success in college and work, and exposure to expectations in different employment settings. It is also expected that programs will help participants to obtain material resources they need to persist in school and at work (e.g., transportation assistance, tuition or other financial supports). Many participants need to work, which diverts time and attention from school. Programs can help alleviate the need to work by helping participants to access student financial aid and other services.

1.3.3 Contextual Factors

Local environments likely affect the extent to which career pathways programs foster positive participant outcomes. Although programs are designed to train participants for occupations with strong

¹² National Research Council, 2012.

local demand, local forecasting can be difficult and jobs may not be available as anticipated. In addition, the ability of participants to continue to move up a career ladder by enrolling in additional courses and acquiring more credentials is affected by access to quality follow-on programs.

1.3.4 A Framework for Evaluation

As an outline of major domains and relationships in career pathways, the conceptual framework in Exhibit 1.2 provides a useful map to the inputs and outcomes that are important to measure in PACE. The implementation study will document the program inputs and context, along with PACE participants' receipt of services and how these services differ from services received by the control group in the PACE study. The impact study will measure impacts on intermediate and main outcomes (based on experimental comparisons) and explore the connections between program inputs and intermediate and main outcomes (using nonexperimental analyses of mediation). Chapters 2 and 3 of this report summarize approaches to the implementation and impact studies, respectively.

1.4 Programs Evaluated in PACE

The PACE team recruited nine programs aimed at increasing postsecondary occupational training opportunities and incorporating key elements of the career pathways framework. This section describes the program selection process and the nine programs that are in the evaluation.

1.4.1 Process for Selection

In addition to interest in participating in PACE, each program had to meet criteria in three categories: (1) their designs had to be promising and include a number of basic elements from the career pathways framework; (2) they had to be able to recruit a sufficient sample; and (3) they had to be able and willing to implement random assignment effectively. Also, at ACF's request, in addition to these criteria, three of the nine sites were to be Health Profession Opportunity Grant (HPOG) recipients.¹³

Program designs: In assessing candidate programs for the study, the PACE team explored the extent to which they incorporated key career pathways elements described above. The team considered the extent to which programs incorporated assessment of skills, interests, and service needs; innovative approaches to basic skills and occupational training; support-related services provided as part of the

¹³ Programs funded by Health Profession Opportunity Grants (HPOG) provide education and training to TANF recipients and other low-income individuals for occupations in the health care field that pay well and are expected to either experience labor shortages or be in high demand. HPOG is administered by the Office of Family Assistance (OFA) within ACF. In FY 2010, OFA made \$67 million in grant awards to 32 entities located across 23 states. These demonstration projects are intended to address two challenges: the increasing shortfall in supply of healthcare professionals in the face of expanding demand; and the increasing requirement for a postsecondary education to secure a well-paying job. Grant funds may be used for training and education as well as supportive services such as case management, child care, and transportation. For additional information, go to:
http://www.acf.hhs.gov/programs/opre/welfare_employ/evaluation_hpog/overview.html.

program and in the community; and employment-related services provided as part of the program and beyond (e.g., job search assistance, on-the-job training or other work experience). Recruitment focused on relatively well-established and widely regarded programs, while allowing also for some newer programs that appeared to be promising.

Potential sample size: Two key criteria were program size and the type of services available to the control group (and their likely take-up rate). To ensure adequate statistical precision, the PACE team sought programs that could attain a sample size of at least 1,000 individuals per program over 18-24 months, evenly assigned to treatment and control groups.¹⁴ The team sought also to choose programs where services similar to, but outside of, the program under study were not readily available in the wider community, since if treatment and control group members received similar services at similar rates, albeit from different sources, little about the overall effectiveness of career pathway program services would be learned.

Implementing random assignment: Programs selected for the study needed to be able and willing to implement high-quality random assignment procedures. They needed, for example, to work with the PACE team to set up expanded recruitment and intake procedures, implement baseline data collection and random assignment, and maintain the treatment and control conditions over time.

HPOG grantee: Abt's original PACE contract included a sample of six programs with an option for three additional programs. ACF exercised the option in order to include three HPOG-funded programs. The selection process for the HPOG programs mirrored that of the larger study.¹⁵

1.4.2 Overview of PACE Programs

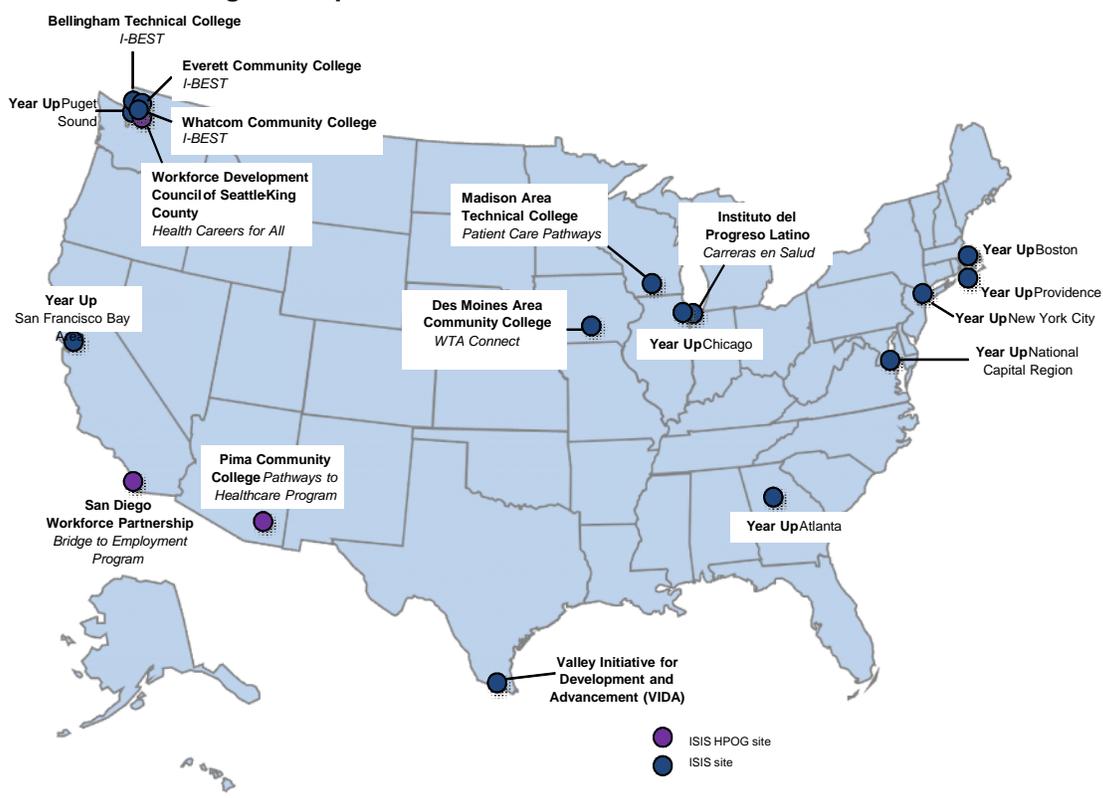
The nine programs selected all emphasize occupational fields in high demand, and most focus on completion of certificates and degrees in these fields. The career pathways in these programs include ladders that enable low-skilled adults to progress along a series of steps within an occupational area as well as to transfer credit to other areas. Although the programs' target populations and strategies to serve these populations vary across the programs, they all provide core career pathways services such as assessment, instruction, supports, and employment connections. Below are short summaries of each PACE program and a map of the PACE program locations (Exhibit 1.3).¹⁶

¹⁴ The exception is Year Up. The random assignment ratio is 2:1 treatment/control, and the total sample size is expected to slightly exceed 2,500.

¹⁵ In addition to the three programs selected, a fourth is a subgrantee to an HPOG program not included in PACE. (As described further in Section 1.7, there are two ACF-funded studies of the HPOG program. PACE and HPOG evaluation staff are collaborating to align design and data collection features of the three studies where feasible and appropriate.)

¹⁶ More detailed profiles for individual programs may be found at the OPRE website: <http://www.acf.hhs.gov/programs/opre/research/project/innovative-strategies-for-increasing-self-sufficiency>

Exhibit 1.3: PACE Program Map



Bridge to Employment in the Healthcare Industry: San Diego Workforce Partnership – San Diego, California. The San Diego Workforce Partnership, a Workforce Investment Board, received HPOG funding to operate the Bridge to Employment in the Healthcare Industry program. Through subcontracts with three community-based partners (navigators), low-income San Diego County residents receive guidance in selecting a healthcare training program appropriate to their skill level, interest, and local demand for workers. Participants receive an individual training account (ITA) to fund their course of study. They also receive case management and support services through the navigators, as well as job search and preparation assistance. Participants can receive a second ITA to further their education in a given pathway.

Carreras en Salud: Instituto del Progreso Latino – Chicago, Illinois. Instituto del Progreso Latino, a nonprofit education and employment organization, operates the Carreras en Salud (Careers in Health) program. This career pathway program in nursing occupations for low-skilled and limited English proficient Latinos moves participants from a Certified Nursing Assistant (CNA) degree to Pre-Licensed Practitioner Nurse (LPN) to LPN and, ultimately, to a Registered Nurse (RN) degree. Instituto del Progreso Latino, in partnership with National Council of La Raza (NCLR), Humboldt Park Vocational and Education Center of Wilbur Wright College, and Association House, designed Carreras en Salud to address both the academic and non-academic needs of low-skilled Latinos. The training program

provides a pre-college contextualized curriculum that moves program participants along an academic, career, and social ladder toward higher paying jobs and greater community involvement.¹⁷

Health Careers for All (HCA): Workforce Development Council of Seattle-King County – Seattle, Washington. HCA targets low-income individuals with limited basic skills, including Temporary Assistance for Needy Families (TANF) and public housing clients. Participants receive career exploration and planning services, integrated case management, wrap-around support services, and systems-navigation assistance through grant-funded navigators. Occupational training is offered within three tiers—foundational, entry level, and more advanced—providing multiple entry and exit points for participants to access employment in the healthcare field and advance in their careers over time. Instruction at the foundational level integrates introductory healthcare content with basic academic skills. Entry-level training prepares participants for jobs such as Nursing Assistant, Medical Office Assistant, and Phlebotomist. More advanced training is customized to help participants reach and complete next-step programs including Medical Assisting and Nursing. Training is paid for through Individual Training Accounts or purchased classes at local community colleges.

Integrated Basic Education and Skills Training (I-BEST) program: Washington State. Implemented in 2004, I-BEST is a multi-occupation program that concurrently provides basic skills or English as a Second Language (ESL) instruction in a range of credit-based occupational training programs along with counseling, financial assistance, and other supports. The program targets individuals who score too low on the college placement test to access their programs of choice, but high enough to benefit from short-term remediation—generally above the 10th-grade level. I-BEST programs have articulated pathways to additional credentials, including one-year diplomas and two-year Associate’s degrees. I-BEST operates in all 34 state community and technical colleges. Each emphasizes matching course offerings to local labor market conditions. Three colleges are participating in PACE: Bellingham Technical College (BTC), Everett Community College (EvCC), and Whatcom Community College (WCC). At BTC, four I-BEST programs are in PACE: Nursing Assistant Certified (NAC), Automotive Technology, Welding, and Electrical Foundations. EvCC I-BEST programs in PACE are NAC, Sustainable Office Skills, and Welding. The WCC I-BEST program in PACE focuses on office services.

Pathways to Healthcare: Pima Community College (PCC) – Tucson, Arizona. With funding from HPOG, the Pathways to Healthcare program is a collaboration between PCC and the local Workforce Investment Board that offers low-income Pima County residents training in 16 healthcare professions organized into 5 pathways. Where a student starts on a pathway depends on his/her test scores and interests. Programs include nurse assistant, medical assistant, medical records, Health IT, and EMT (basic and paramedicine) among others. Once enrolled, students have access to a variety of services including case management, remedial or developmental education if needed before beginning occupational training, a dedicated Pathways advisor, and a dedicated Pima County One Stop workforce development specialist who works with them to address supportive service needs and find employment upon

¹⁷ Instituto is an HPOG sub-grantee to Will County Workforce Investment Board. Some PACE program slots are funded by HPOG.

completion of the training program. Pathways staff support students to train to their highest earning potential by completing training and securing a job, then supporting them through more Pathways training and ultimately a well-paying career.

Patient Care Pathway Program: Madison Area Technical College (MATC) – Madison, Wisconsin. The Patient Care Pathway Program provides short-term condensed training that allows lower-skilled students to take courses for college credit and prepares them for health care diploma and degree programs. There are two tracks offered in the program that are organized according to students' skill levels—Patient Care Academy 1 (PCA 1) and Patient Care Academy 2 (PCA 2)—both of which ladder into several established academic programs in the health field.¹⁸ PCA 1 is for students interested in a one-year health diploma program or a two-year health degree program, and whose placement test scores are too low for PCA 2. PCA 2 is designed for students interested in pursuing a two-year health associate's degree. Through the program, students who come to MATC with academic skills that are too low to begin their health program are able to accelerate their remediation and work toward earning their degree. Both Patient Care Academies integrate developmental coursework with health program prerequisites and contextualize the developmental courses for the health field. Students in the Patient Care Academies take classes in a cohort. The Patient Care Pathway Program also provides proactive advising.

Workforce Training Academy (WTA) Connect: Des Moines Area Community College (DMACC) – Des Moines, Iowa. WTA Connect was developed to expand access to DMACC's Workforce Training Academy, which provides vocational certificates in targeted high-growth, high demand sectors at no cost to participants. DMACC found that the Workforce Training Academy's eligibility requirements prevented applicants with low basic skills and/or who lacked a secondary credential from being admitted. WTA Connect admits students with low basic skills and/or no secondary credential and prepares them to enroll in and complete a vocational certificate course. The program packages the vocational education with basic skills remediation, psychosocial skills development, and advising. Students who lack a secondary credential can take classes for the HiSET® test (a high school equivalency exam) concurrently. After completing the preparatory components of the program, WTA Connect participants enroll in certificate courses alongside traditional Workforce Training Academy students. Each short-term certificate offered in WTA Connect ladders into one or more specific degree or diploma programs at the college. WTA Connect completers can either continue their education at DMACC or enter employment.

Valley Initiative for Development and Advancement (VIDA) – South Texas. Serving adults on public assistance or with incomes below 200 percent of the poverty level in the Lower Rio Grande Valley, the nonprofit Valley Initiative for Development and Advancement (VIDA) aims to help students achieve an associate's degree and gain occupational training in allied health, manufacturing, technology, business, education, and other specialized trades. VIDA offers a bridge program, the College Prep Academy, to

¹⁸ The Patient Care Pathway Program also included a third track, Patient Care Certified Nursing Assistant (PCCNA) that was piloted during the latter portion of the PACE enrollment period. This track was designed to provide a certificate and subsequently ladder into PCA 1 or PCA 2. Demand for this offering was lower than expected.

build individuals' basic educational and language skills to prepare them to enroll at a local college. The program also provides extensive, proactive advising, financial assistance, and support services. Modeled after Project QUEST in San Antonio, VIDA requires a strong commitment to full-time education on the part of both the individual and the program.

Year Up – Atlanta, Bay Area, Boston, Chicago, National Capital Region, New York, Providence, Seattle.

Founded in 2000, Year Up is an intensive one-year program that provides low-income high school graduates and GED recipients ages 18 to 24 with a combination of hands-on skill development and corporate internship opportunities. The program addresses program participants' social and emotional development and provides appropriate support to place them on a path to economic self-sufficiency. The first six months of the program focus on technical and professional skill building. During the second six months, participants apply these skills in corporate internships. Participants earn an educational stipend throughout the program. They can also earn college credits for their participation (typically ranging from 18 to 30 credits). After graduation, participants continue to receive support and build their professional networks through Year Up's Alumni Association.

1.4.3 Similarities and Differences across Programs

The programs in PACE exhibit similarities and differences on key attributes. These include the organizations involved and their roles; target populations; and focal occupations, credentials, and career pathways steps.

Lead organization. The largest number (four) of programs in the study are operated by community or technical colleges. Three are operated by nonprofit organizations (Year Up, Instituto del Progreso Latino and VIDA), and two are led by Workforce Investment Boards (San Diego Workforce Partnership and the Workforce Development Council of Seattle-King County).

Target population. By design, all PACE programs target low-skilled, low-income populations, although they do so with emphases on varying characteristics. The Instituto del Progreso Latino program focuses on low-skilled Latinos with reading skills beginning at the 6th grade level. Des Moines Area Community College's program targets participants at similarly low skill levels (7th to 8th grade equivalent), but not a particular ethnic group. A number of the remaining programs target adult or young adult populations with skills closer to (though still below) college level and who have multiple other challenges to entering and succeeding in postsecondary education. For example, I-BEST and Madison Area Technical College target students with skills between 10th grade and college entry levels, and Year Up recruits disadvantaged youth ages 18–24 with a high school diploma or equivalent who meet program expectations regarding academic readiness.

Occupational focus. Five PACE programs focus exclusively on training for health-related occupations. These programs include the three primarily HPOG-funded programs (Pima Community College, San Diego Workforce Partnership, Workforce Development Council of Seattle-King County), Instituto del Progreso Latino, and Madison Area Technical College. The other four programs focus on multi-occupational areas, such as welding, electrical, health, and administrative support (Washington I-BEST and Des Moines Area Community College); health, business, education, specialized trades (VIDA); and IT, quality assurance, customer service and financial operations (Year Up).

Steps on career pathway. The PACE programs vary in both the number and level of career pathway steps supported. The Instituto del Progreso Latino program has the longest potential path, with participants being able to enter at the pre-CNA level and advance to RN. Another longer-term program, VIDA, primarily focuses on higher pathway steps with the associate’s degree as the primary target. However, VIDA also supports one-year and shorter certificates. VIDA also operates a 16-week accelerated academic bridge program for individuals testing at 10th grade or higher but who have not yet qualified for college credit courses. In contrast, the Des Moines Area Community College program, the Workforce Development Council of Seattle-King County, and the San Diego Workforce Partnership help participants attain short-term credentials. Madison Area Technical College’s programs are designed to accelerate participants’ entry into one- or two-year diploma or degree programs, while I-BEST program participants attain a short-term credential but have articulated pathways to additional diplomas, certificates, and degrees. Year Up focuses on moving program completers into entry-level career track jobs rather than credential receipt, though it has worked out agreements with local colleges to grant credits for Year Up courses completed.

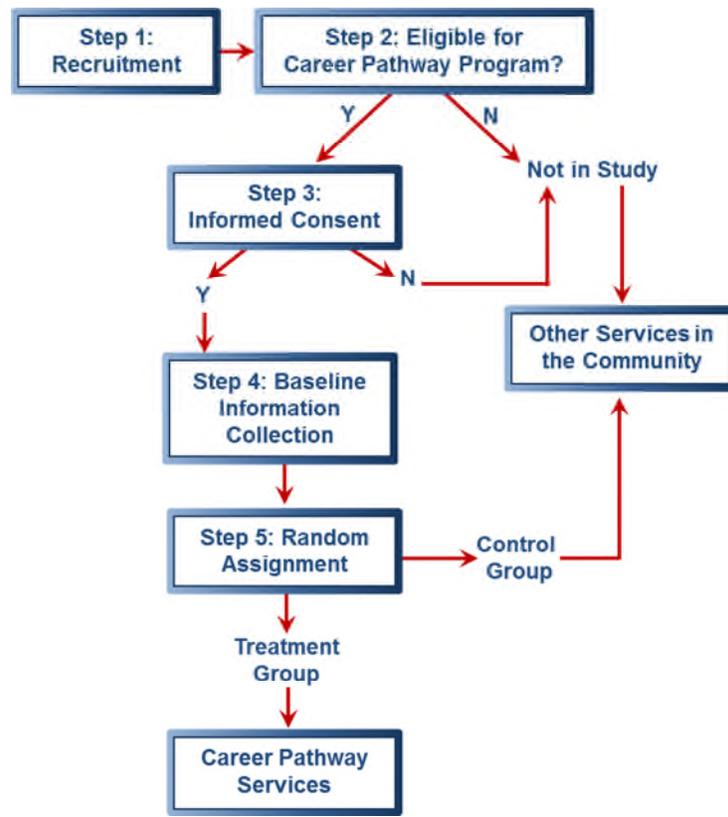
1.5 Study Design

PACE is using an experimental design to estimate program impacts on education, employment, and other outcomes. In each site, a random assignment process assigns applicants that meet program eligibility requirements to one of two groups: a treatment group that can access program services and a control group that cannot, but that can access other services in the community.¹⁹ Random assignment ensures that the characteristics of the two groups have no systematic (i.e., non-random) differences at baseline. This design makes it possible to estimate impacts without bias by comparing average outcomes for treatment and control group members over time.

While all PACE programs share a common approach for conducting random assignment, each tailors it to account for local practices for recruitment, eligibility determination, and enrollment. Exhibit 1.4 depicts the common approach.

¹⁹ The “embargo period” is the period after random assignment during which the control group remains ineligible for treatment group services. In PACE it ranges from two to three years.

Exhibit 1.4: Random Assignment Approach



As the exhibit shows, program staff recruit potential participants and determine their eligibility. As part of participant eligibility determination, staff typically administer some kind of academic skills assessment or college placement test to determine skill levels. (Chapter 4 provides information about these assessments.) Program staff then administer the informed consent form, which describes the study, the data collected, and the rights and responsibilities of the participant. After signing the consent form, participants complete two baseline data collection forms: the Basic Information Form (BIF) and Self-Administered Questionnaire (SAQ) (see Chapter 4 for a description of these forms). Applicants choosing not to sign the informed consent form are excluded from the study and from participating in the career pathways program for the duration of the study enrollment period. Program staff then conduct random assignment using a web-based tool. Individuals assigned to the treatment group are offered the program while those assigned to the control group receive information about other services in the community.

1.6 Major Evaluation Substudies and Reports

PACE includes three substudies. This section briefly summarizes the major questions and reports provided under each. Chapters 2 and 3 provide detailed approaches to the implementation and impact and cost-benefit studies, respectively.

1.6.1 Substudies

PACE includes implementation, impact, and cost-benefit substudies. Each is described below.

Implementation Study. The implementation study will describe and document the design and operations of each of the nine PACE programs. In doing so, it will provide insight into what is involved in the design or modification, as well as operation, of career pathways programs. The implementation study provides contextual information for understanding and interpreting the impact estimates and findings on implementation experiences to guide future program designs and implementations.

Impact Study. The impact study will measure differences between the treatment and control group members on main and intermediate outcomes. Main outcomes include educational outcomes (e.g., persistence in education, receipt of certificates and degrees) and employment and earnings-related outcomes. Analyses also will assess whether impacts vary by subgroup and estimate impacts for intermediate outcomes in the theory of change for each program. The PACE team will analyze and report findings for each program separately. The rationale for separate studies is that, although all nine programs utilize basic building blocks in career pathways, they vary on a substantial number of fundamental design features, including eligibility criteria and target populations; services provided; goals for participants; and timeframes for expecting goals to be achieved. The first impact reports will cover at least 15 months of follow-up after the point of random assignment. Additional reports, should ACF exercise follow-up survey options, will extend analyses of education and employment outcomes for longer time periods and cover other life outcomes (e.g., individual and family well-being) that might be influenced in the longer run. Limited cross-site analyses also may be conducted.

Cost-Benefit Study. The cost-benefit study will describe monetized benefits and costs from varying perspectives (social, participant, and taxpayer). Key areas that will be addressed include the gross and net costs of each program, and whether the benefits of providing services outweigh the costs.

1.6.2 Reports

Although all nine PACE programs fit within the career pathways framework, as noted above they differ in terms of target population, credentials emphasized, vocational fields, and other services. The evaluation team will produce program-specific reports at up to three follow-up points. The first follow-up impact report will be based on data collected starting in the 15th month after randomization and will include implementation study findings (e.g., on how programs were implemented, alterations to program designs, and patterns of service receipt and completion within the treatment group). The second follow-up report will be based on data collected starting in the 36th month after random assignment and will include program impacts and the cost-benefit study findings. The third follow-up report will be based on data collected starting in the 60th month following random assignment.^{20 21} In addition to these reports, the evaluation has previously released profile reports for each program

²⁰ Though samples will be released at 15, 36, and 60-months of follow-up, the average exposure reflected in the first, second and third follow-up reports will typically be somewhat greater: To maximize response rates, the data collection window for each sample cohort (i.e., sample randomly assigned in same month) will remain open for 6 months (and longer for some sample members that are difficult to track and interview).

²¹ Thus far ACF has funded only the first and second follow-up studies (through the initial PACE contract and a follow-on contract). Decisions on funding for the third follow-up study will be made at a later date.

describing the relation of each program to the career pathways framework and early implementation of the PACE evaluation.

1.7 PACE and Related ACF Studies

As noted earlier, there is a great deal of interest among policymakers in the potential for career pathways programs to provide occupational and other skills to help low-skilled, low-income adults attain self-sufficiency and advance to the middle class, while filling a vital need for skilled workers in the economy. Three federal departments are funding initiatives in this area. The Departments of Labor and Education, for instance, launched a one-year Career Pathways Initiative in 2010, which provided funding to nine states and two tribal entities to develop sustainable career pathways to promote linkages among system partners. The Department of Labor produced a set of technical assistance tools for state, local, and tribal policymakers to use in development and implementation of career pathways approaches.²² The Department of Education's Office of Vocational and Adult Education funds the Designing Instruction for Career Pathways initiative.²³

The Department of Health and Human Services also has a growing interest in career pathways programs, and PACE is one study within its larger career pathways research portfolio. ACF funded impact and implementation studies of the Health Profession Opportunity Grant (HPOG) program (see Box 1). Like the programs in PACE, many HPOG-funded programs embody key elements of the career pathways framework. Additionally, as noted earlier, three PACE programs are HPOG-funded and one additional program receives funding as an HPOG subgrantee.

Both the PACE and HPOG are evaluating occupational training programs serving similar populations. Thus, the studies provide an opportunity to document and assess a collection of career pathways-oriented programs supported (through evaluation funds or direct program funds) by ACF. The PACE and HPOG evaluation teams are coordinating on terminology (e.g., career pathways program elements), data collection instruments (for baseline data collection, the implementation study and the impact study),²⁴ and to the extent it is appropriate, research questions and evaluation frameworks.

²² Go to www.learnwork.workforce3one.org; information is also on the PACE project website at www.projectPACE.org.

²³ The website is <http://www.acp-sc.org>

²⁴ As described further in Chapter 2, a number of implementation study data collection tools will be utilized by PACE and HPOG National Implementation Evaluation (NIE). PACE baseline and follow-up survey data collected from study participants in the four PACE/HPOG sites will be used by HPOG Impact Study staff in their analyses.

Box 1: Health Profession Opportunity Grant Research Projects

HPOG Impact Study. The HPOG Impact Study is evaluating the effectiveness of approaches used by 20 of the HPOG grantees providing TANF recipients and other low-income individuals with opportunities for education, training and advancement within the health care field. The study will also evaluate variation in participant impact that may be attributable to different HPOG program elements and models. The study uses an experimental design in which eligible applicants are randomly assigned to a treatment group that is offered participation in HPOG and a control group that is not permitted to enroll in HPOG. In a subset of sites, eligible applicants are randomized into two treatment arms (a basic and an enhanced version of the intervention) and a control group. The HPOG Impact Study will use implementation interview guides for partnering employers, instructors, HPOG program management, and HPOG program staff. Follow-up surveys of participant and control group members will also be fielded following the 15-month anniversary of random assignment. Data collected from the HPOG participants served by these 20 grantees will also be used for the National Implementation Evaluation (see below).

HPOG National Implementation Evaluation (NIE). The NIE will describe and assess implementation, systems change, and outcomes of the 27 HPOG grantees focused on TANF recipients and other low-income individuals. The NIE will collect data about HPOG program designs and implementation, HPOG partner and program networks and indicators of systems change, employers' perceptions of HPOG programs, the composition and intensity of HPOG services received, participant characteristics and HPOG experiences, and participant outputs and outcomes. The NIE will collect data through the HPOG Performance Reporting System (a Management Information System containing descriptive information on all HPOG participants), a Grantee survey, Management and Staff survey, Employer survey, and Stakeholder/Network survey. A first follow-up survey of participants will also be fielded to a small subsample of HPOG participants not involved in the Impact Study.

Given some overlap in sites and common use of experimental designs, it is helpful to note the key differences between the PACE and HPOG impact studies. The main emphasis in PACE will be on learning in depth about the implementation and impacts of a smaller number of differing programs—evaluating each program specific to its target population, design, and hypothesized outcomes. In contrast, the HPOG Impact Study is focusing on variation in impacts and factors associated with impacts across a pooled sample of 20 grantees (and many more program sites). Commensurate with these differing aims, PACE is gathering more detailed information on implementation in each site and measuring a wider array of outcomes in follow-up surveys.

1.8 Organization of the Report

The remainder of the evaluation design report includes the following sections:

Chapter 2 describes the implementation study design. This chapter begins by outlining the major components of the implementation study: developing logic models to explain the design of each program; describing the intervention as it was implemented for each program and the extent to which it was implemented as intended; documenting the experiences of program participants; examining differences in service receipt and educational experiences between treatment and control group members; and assessing the implications for scalability and sustainability. It then describes data sources for the study and analysis methods.

Chapter 3 describes how the impact study will be conducted for individual programs. The chapter begins with the time horizon for the impact analyses and then describes hypotheses, the basic methods for estimating impacts, minimum detectable impacts, subgroup analyses, and potential analyses to understand the sources of impacts for a single program. This chapter also describes the cost-benefit study.

Chapter 4 describes the quantitative data sources used for the impact, cost-benefit and implementation studies. Key data sources discussed include: baseline data, follow-up surveys of treatment and control group members, and administrative data from programs and national databases.

2. Implementation Study

The PACE implementation study has multiple objectives. First, it describes and documents the design of the nine career pathways programs participating in PACE. Second, it documents each program's model—that is, how the program as implemented diverges from what was intended, and what changes were made, when, and why. Third, it provides contextual information for understanding and interpreting the impact estimates. Finally, implementation study findings provide information to policymakers and practitioners on program design and implementation useful in shaping policies and practices.

As noted in Chapter 1, although all programs in PACE embody key components of the career pathways framework, they differ along a number of dimensions, including lead organization, target population, occupational focus, steps on the career pathway, approaches to basic skills and occupational instruction, and employer connections.

The main research questions in the implementation study are:

1. What is the basic design and underlying logic of each PACE program? What are its elements? What is its institutional and community context?
2. To what extent was the intervention delivered as planned? What was the fidelity to the design? If the program's structure and services changed during the study period, why?
3. What were the treatment group's participation patterns and experiences with program services? With transitions beyond the core career pathways program?
4. What are the differences in type, duration, and content of services received by treatment and control group members?
5. What are the lessons from PACE programs for the field in the areas of scalability and sustainability?

Section 2.1 of this chapter describes the approach to each implementation study question. Section 2.2 provides additional detail on data sources. Section 2.3 outlines implementation study reports.

2.1 PACE Implementation Study Research Questions

The five questions above shape the design of the implementation study for the PACE evaluation. Each is discussed in turn below.

2.1.1 What is the design and underlying logic of each program?

A program's theory of change is a formulation of what the program needs to do to bring about intended changes in participant outcomes. It operationalizes the theoretical relationships between components and expected outcomes, as well as underpins quantitative analyses by documenting the assumed conditions and relationships.²⁵ The PACE Theory of Change (Exhibit 1.2) specifies the key intermediate and main outcomes that are expected to occur. How each program is structured to achieve these outcomes is one focus of the implementation study.

The implementation study will provide a logic model for each PACE program, setting models in the context of the broader theory of change described in Chapter 1. A logic model describes how a program is supposed to work.²⁶ It breaks down complex relationships into distinct variables or components. As such, it is a road map of what the program is designed to accomplish and a tool for linking what it does to what it hopes to achieve. The logic model also specifies what needs to be measured. Exhibit 2.1 presents a generic model that will be used in the implementation study. For each PACE program, the evaluation team will document the following:

Program goals. The team will document the overarching goal(s) the program seeks to achieve. For example, is it a particular type of credential, a particular kind of job or wage level, or a successful transition from one level of education and training to the next? In documenting goals, it will be important to differentiate between the objective of the program in PACE, which might be time limited (e.g., one semester, one credential) and the higher steps on a pathway that programs may hope participants someday reach (e.g., a bachelor's degree).

Underlying assumptions. The team will describe the program's underlying assumptions about the target population (including academic skill level and demographics) and its needs. This includes beliefs about the connection between the program components and the outcomes sought. More specific questions include: What is the perceived barrier to successful completion of postsecondary training? Why do staff think it is amenable to intervention? Making explicit underlying assumptions helps clarify what the program is trying to achieve and for whom.

Context. The team will document factors in the environment that are perceived to be likely to affect implementation of the program and the achievement of its goals. This includes the local community that might affect the program (e.g., local community program expansions or closures, the economy, and population demographics), organizational characteristics (e.g., characteristics of the host organization such as size, range of services offered, its service area, where the career pathways program fits within the organization), and collaborative arrangements between the organization and other programs (e.g., state or local government programs, community-based organizations). Other contextual factors include other programs that provide similar services in the area that could be available to control group members (whether sponsored by the host organization or other organizations), and other organizational

²⁵ Flick, 2006

²⁶ Logic model framework based on McGroder and Gardiner (2006).

relationships, such as referral partners or those to which programs could/do refer participants for services not provided by the program.

Program inputs. The study team will document the resources programs use to develop and carry out key activities. These include funding sources, staff (personnel, training, professional development), materials (e.g., curricula), space, and equipment. Other inputs include federal, state or local policies (e.g., surrounding postsecondary education); program funder guidelines (e.g., Washington State Board of Community and Technical College I-BEST policies and oversight of college plans); service and/or referral partners; and data collection systems.

Program components. The study team will describe each of the four overarching career pathways components included in the design of each program. Specifically, the team will document the design and delivery of:

- *Assessments of academic skills and other abilities and need*, specifically which assessments are administered, what information is collected on each, when they are administered (once or at multiple points in time), and how the information is used and by whom (e.g., instructors, case managers, placement staff).
- *Approaches to education and training*, for basic academic skills, occupational skills, employability skills, and other training (e.g., college success, study skills). The team will also document strategies for delivering each type of training (e.g., integration and/or contextualization, active learning approaches).
- *Academic and non-academic supports to promote success in the program*, including advising and counseling, instructional supports (e.g., tutoring, study groups), social supports (e.g., student cohorts, peer/alumni mentoring), supportive services (e.g., child care, transportation, provided either in house or via referral) and financial assistance (e.g., assistance with financial aid applications, direct financial support). This also includes support for transitions to further credentials along the designated pathway, either immediately following completion of a pathway step or after a period of work.²⁷
- *Approaches for connecting participants with career-track employment opportunities*, such as the role that employers play in program design and instruction and connecting participants with employment; employment experiences available during training (e.g., internships, work study); and linkages to employment after training (including job placement services).

²⁷ A number of factors can facilitate or impede transitions, including whether instructors in the career pathways program are communicating with higher-level instructors about what students need in order to succeed in future classes, whether guidance and supports continue beyond the core program or cease, whether cohorts continue and the ease of scheduling and registering for classes (Wachen et al., 2012).

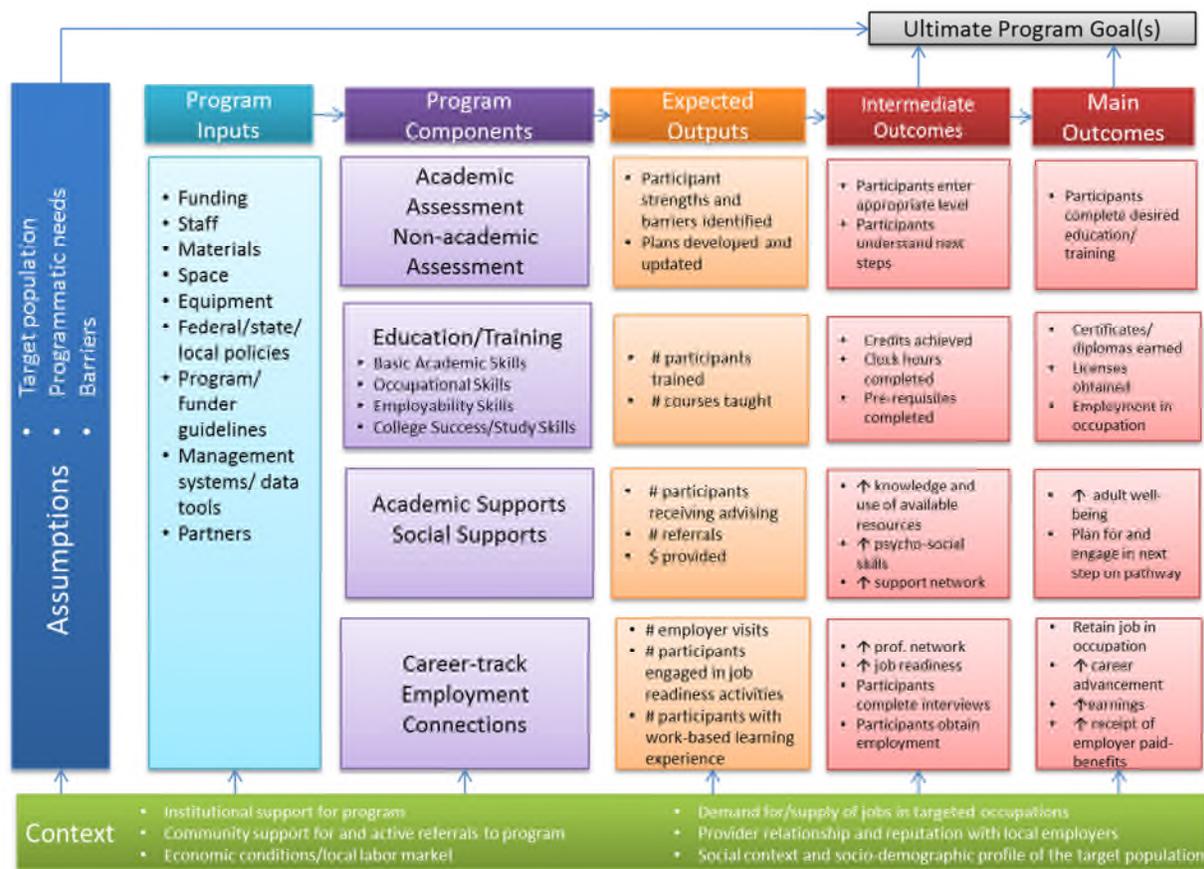
Within this four-component framework, the study team will document the occupations, credentials, and basic career pathway levels addressed by the program.

Outputs. The team will also document outputs, that is, what the program will produce immediately with respect to participation and service delivery. Examples of outputs related to assessment include numbers of assessments conducted, participant strengths and barriers identified, and educational/support plans developed. Examples of outputs related to education and training include the number of participants trained and the number of courses offered. For academic and social supports, outputs include the number of participants receiving advising, the number of referrals made for services, and financial assistance (amount) provided. Finally, outputs related to career track employment connections include the number of employer site visits, the number of participants engaged in job readiness activities, and the number of participants with work-based learning experience.

Intermediate (proximal) and main (distal) outcomes. Finally, the team will describe the outcomes expected for program participants, following participation and over the longer term. More proximal outcomes related to academic and non-academic assessment include correct placement in an education/training course and participants' understanding of next steps on the pathway. Education and training-related proximal outcomes include credits achieved, clock hours completed, and pre-requisites for further training completed. More intermediate outcomes related to academic and social supports include increased knowledge and use of available resources, improved psycho-social skills, and support networks. Finally, intermediate employment connections outcomes include improved professional networks, job readiness, completion of interviews, and employment. Main outcomes include completion of desired education and training; certificates and/or diplomas received, licenses obtained, and employment in target occupation; increased individual well-being; and job retention, career advancement, improved earnings, and receipt of job-related benefits.

A logic model is a useful tool to describe the program as designed. A program's underlying logic might be explicit (e.g., described in a detailed plan) or implicit in its structure and activities. In detailing each program's logic, the team will draw initially on program documents (e.g., HPOG applications for grants, I-BEST program applications to the state board for approval, annual reports, and marketing materials) and interviews with key program staff. Following the first round of site visits (see Section 2.2), site teams will develop a model for each program. These documents will inform areas for further exploration during the second site visit.

Exhibit 2.1: Logic Model



2.1.2 What was the program like as actually implemented? What was the degree of fidelity to the design? Did the program change during the study period and, if so, why?

Each program’s logic model will summarize how the program is intended to operate and will suggest areas to explore in greater detail. Inevitably, program administrators and staff make adjustments to their initial design as realities on the ground raise unanticipated challenges and opportunities or require more detailed thinking about certain operations. Documenting the program model components, including their rationale and any adjustments made, is valuable both for understanding the actual program responsible for measured impacts and for insights into lessons for operation and replication. Actual implementation is not just an issue for new programs or those that added new components for PACE to strengthen their career pathways models. It is also important for well-established programs that are scaling up to accommodate PACE sample size requirements or are responding to changes in their external environments.

Through staff interviews, staff surveys, program data analysis, and observations, the study team will document key career pathways program components as implemented, providing quantitative measures where possible (e.g., number of classes or advising sessions, duration of classes and sessions). Additionally, as part of an in-depth description of the program, the team will assess the emphasis the program places on each component. Finally, the team will describe and assess program management

and organization, staffing and professional development, funding, and partnerships with other service providers and employers.

What career pathways elements are implemented in each program? Were they modified?

The study team will document the key career pathways program components as implemented. Exhibit 2.4 at the conclusion of this chapter outlines the key components and subcomponents examined by the implementation study. As the exhibit shows, the team will document: 1) for key services, the nature and content of the career pathway components provided, as well as how long they are typically offered and utilized; 2) for the overall design, individual components, and management/staffing, what aspects, if any, changed and why; and 3) the emphasis and strength given to each of the program components provided within the broader career pathways framework.

For example, in describing the role of assessments in each program, the team will document assessment tools used, academic and non-academic areas assessed, when and how assessments are conducted, and how results are used in providing services. Any of these aspects could change—for example, a shift from limited to more consistent assessment or assigning a staff member to follow up with each participant based on assessment results—and the reasons might illuminate lessons of wider interest. Some programs may make only cursory use of assessment, while others put assessment and re-assessment of an array of skills and needs at the center of their models.

How are programs managed and staffed?

Program management and staffing arrangements vary greatly and play a critical role in translating abstract logic models into functioning programs. Key dimensions include: management and staffing functions and qualifications, professional development, and day-to-day organization and management. Questions to be investigated by the research team include: how do staff see their roles in the program, the nature of their work, their relationships with participants, and their understanding and interpretation of program goals? What educational backgrounds and experiences do managers and frontline staff bring?

2.1.3 What are the experiences of treatment group members?

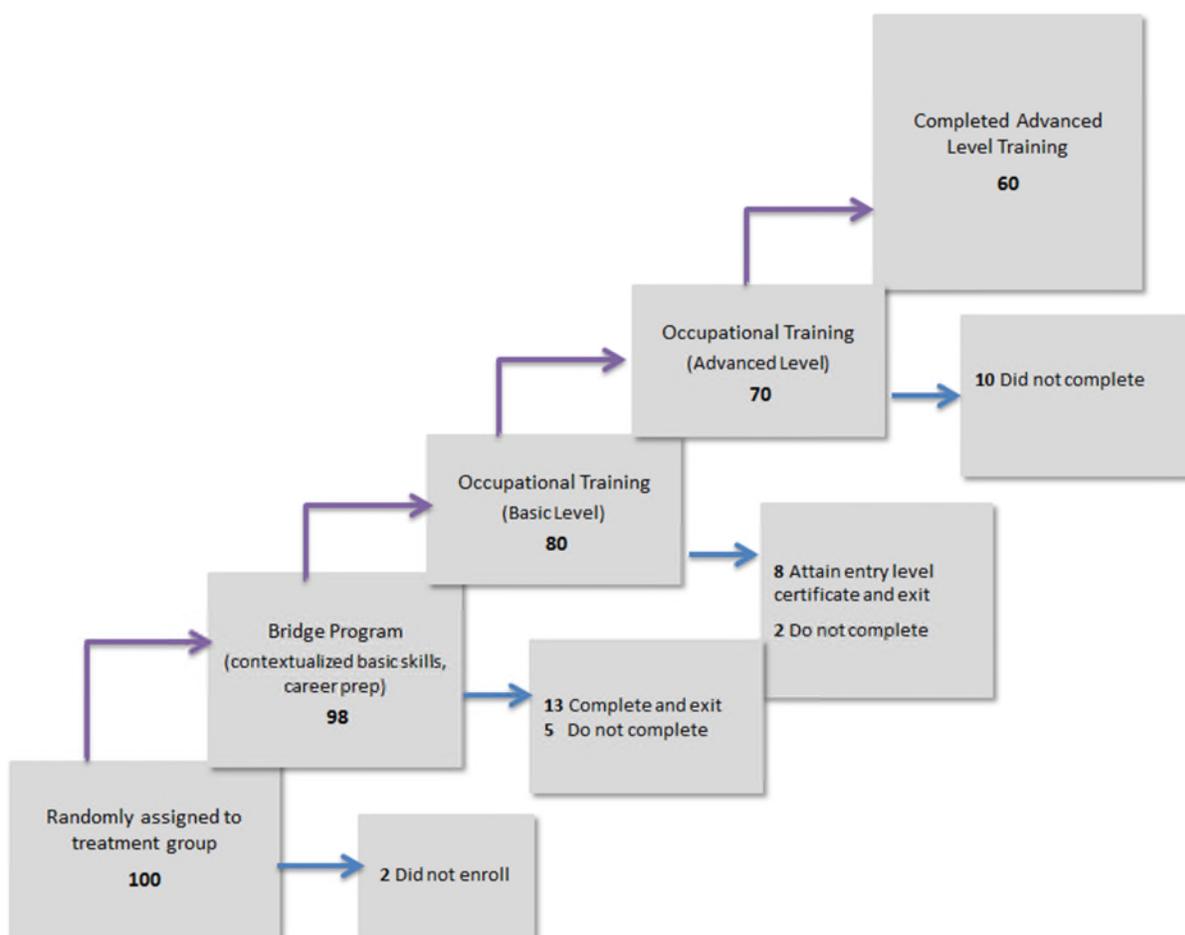
A key requirement for generating program impacts is treatment group members' exposure to different aspects of each program. In practice, some program participants will receive all or most services, while others will not because they fail to enroll in the program, drop out partway through, or participate at a lower-than-expected level. Such analyses will cover services such as assessment, basic and occupational skills training, support services, and employment services, as well as any unique, program-specific components (for example, weekly group sessions). The first follow-up survey will also collect information on service receipt needed to document the experiences of treatment group members.

For the analysis of patterns of services receipt, the team will include a schematic view of how participants flow through major steps and services offered by each program. Exhibit 2.2 illustrates this analysis for a program involving a fairly extensive set of career pathway steps. The figure shows the flow for 100 typical participants through key steps in the career pathway as well as outcomes for those who leave the program at different points. For simplicity, the illustration does not show new or re-entries at different steps, though analyses of programs encouraging re-entries will represent such events. Though

not shown in the figure, for each step the analysis also will document—to the degree that administrative data allow—outcomes such as credential receipt. Finally, the analysis will examine how entry and exit rates and related outcomes (e.g., credential completion) vary for participants within different subgroups.

In addition to documenting program flows through key steps and associated characteristics, analyses will summarize key aspects of participation in specific program components. Illustrative measures include number of months individuals participate, number of hours of instruction they receive, receipt of support services, number of advising sessions, receipt of financial aid assistance, receipt of employment services, and hours and credits earned.

Exhibit 2.2: Participant Flow for 100 Typical Students through an Illustrative Career Pathway Program



2.1.4 What are treatment-control group differences in service receipt?

One of the most important tasks in the implementation study is to document the “dosages” underlying experimental impacts. These analyses will draw heavily on data collected in the first follow-up survey, which provides comparable measures for treatment and control group members. These statistics will be illuminated by rich qualitative data for smaller samples from the two groups that participate in in-depth interviews. Analyses will focus on differences in both the levels and kinds of services received. Results

might indicate, for example, that engagement in postsecondary education and training was similar, but the treatment group experienced, on average, more highly contextualized and less strictly lecture-based instruction.

Analyses of differential receipt will cover areas such as the following:

- Levels and type of education and training services received, such as the number and types of program(s) attended (basic education, ESL, college credit, occupational training, career/life skill programs), type of organization providing training, and occupational area and credentials targeted.
- Length and schedule of programs attended, including the timing and intensity, dates and hours attended, and whether the individual attended full time or part time.
- Type of instructional practices characterizing education and training received, including characteristics of classroom approaches (e.g., active participation, contextualization).
- Participation in other skills activities, such as instruction in study skills, problem-solving, time management.
- Support services received, including academic advising, financial aid advising, tutoring, career counseling, job search/placement and help arranging supports (e.g., child care, transportation, health care).
- Financial assistance received, by source (family or friends, grants, loans, direct assistance from program).
- Employment assistance received, including career-relevant work opportunities.

The study team will examine also factors that facilitate or hinder participation, such as service needs (e.g., child care, transportation), financial issues, and work/life/school balance. The main sources of data for the above analyses are the first follow-up survey and in-depth participant interviews.

2.1.5 What are lessons for the field in terms of scalability and sustainability?

Because most PACE programs had to expand the numbers they served to participate in the study, their experiences afford insights into requirements for successfully scaling up programs. As an extension of the investigation of this key topic, the implementation team also will document and assess PACE programs' approaches to securing financial and institutional supports needed to sustain their efforts beyond the end of PACE. Such approaches may involve efforts to maintain—and possibly replicate—stand-alone approaches to program operations, as well as to adapt key features in ways that can be readily incorporated in wider systems.

Analyses in this area will draw on interviews with the range of stakeholders involved in developing and operating PACE programs, including program managers and staff, funding agencies, and other stakeholders such as community agencies. The team will marshal this qualitative information to help interpret key outcomes such as numbers of participants recruited, funds raised, services secured, and success in sustaining services. Major questions on scalability and sustainability include:

- For scalability, how did programs seek to increase scale by expanding recruitment, enlarging operations at existing sites, and expanding partnerships? This investigation will explore a number of factors identified in the literature on scalability, such as support and buy-in from key stakeholders; approaches to planning, management, and staffing; success in securing additional resources; and local factors such as demand for services and competing programs.²⁸
- On sustainability, do programs plan to maintain or increase services, and what are their strategies for doing so? What are their approaches to securing funding and maintaining operations in the face of possible fluctuation in student interest, national and state policies and programs, local occupational demand, and staff turnover? The investigation also will focus on success in developing strong community partnerships or coalitions, the implementation of new organizational practices and policies, efforts to strengthen integration between the specific intervention and the host organization’s mission and operating routines, and the role of program advocates or champions.²⁹

2.2 Data Sources for the Implementation Study

The implementation study will draw on a variety of qualitative and quantitative data sources. Qualitative data sources include: semi-structured interviews with program managers, staff and partners, and in-depth interviews with treatment and control group members; a review of program documents; and observation of program activities. Quantitative data sources include on-line data on demographic and other characteristics of treatment and control group members at the time of random assignment (baseline); a survey of treatment and control group members starting 15 months past random assignment; surveys of instructors, managers, and case managers/advisors; PACE program administrative data; and college student record systems. This section describes key measures and data collection methods for each data source. (See Exhibit 2.3 for a summary).

²⁸ Harris, 2010; Granger, 2011; Quint et. al., 2011.

²⁹ Scheirer and Dearing, 2011

Exhibit 2.3: Data Sources for Implementation Study Research Topics

Implementation Study Area	Qualitative Data				Quantitative Data				
	Interviews with Program Staff and Partners	Treatment and Control Group Interviews	Program Documents	Observations	Baseline Data	First follow-up survey	Manager, Instructor and Case Manager/Advisor Surveys	National Administrative Data (college records)	Program Administrative Data
Describing Program Logic									
Program inputs	✓		✓				✓		
Intervention/activities	✓		✓	✓			✓	✓	
Outputs	✓		✓				✓	✓	
Immediate and subsequent outcomes	✓		✓				✓	✓	
Contextual factors	✓		✓		✓		✓	✓	
Describing Programs as Implemented									
Career pathways components	✓		✓	✓			✓		✓
Management and staffing	✓		✓				✓		
Treatment-Control Difference in Use of and Experiences in Services									
Types of education and training services		✓				✓		✓	
Length and schedule of programs		✓				✓			
Instructional practices received		✓				✓			
Participation in other skills classes		✓				✓			
Supports received		✓				✓			
Financial assistance received		✓				✓			
Employment assistance received		✓				✓			
Assessment of services received		✓				✓			

Time involved		✓				✓			
Factors that facilitate or hinder participation		✓				✓			
Documenting Experiences of Treatment Group Members									
Types and level of services received		✓				✓			✓
Flow through program and progression through activities/classes		✓				✓			✓
Understanding Scalability and Sustainability									
Issues in scaling up	✓								✓
Ability to sustain programs	✓		✓						

2.2.1 Interviews with Program Staff and Partner Organizations

The study team will collect information about the PACE program services in interviews with administrators and staff (including instructors and advisors). These interviews will collect information on a wide range of implementation study topics, including: program and organizational structure; key partners; target group; career pathways program components (including nature and content of training and support services); linkages with employers; programmatic priorities; funding resources; sustainability of the program after the study; and the economic and programmatic context. The aim is to provide a rich description of factors (programmatic, institutional, and economic) that facilitate or inhibit the successful implementation and operation of the program. In addition to information on PACE programs, interviews will explore and help to document the existence of other programs and services in the local area available to members of the control group (as well as the treatment group post-PACE program).

The team will use semi-structured guides to conduct discussions. These guides, tailored to each type of respondent, will identify key topics and provide open-ended questions and probes on key topics. The semi-structured format provides flexibility to explore interviewees’ perspectives and surface new issues and responses, while ensuring that research questions central to the overall PACE study and each site’s design are covered.

The team will conduct two rounds of visits to most programs: the first visit during the initial six months after sites begin random assignment and a second visit towards the end of the random assignment period. During the visits, PACE staff will interview program managers; staff involved in evaluation activities (e.g., recruitment, intake, random assignment); instructors; advisors; case managers; and staff at any partner agencies with an important role in in service delivery. The goal of the first visit is to document the program’s logic and key components, and to assess implementation of evaluation procedures. The principal focus of the second visit will be to document any modifications to operations

or the provision of services, as well as implementation challenges. The team will also focus more on approaches to, and lessons on, scaling up and sustainability in the second round of visits.

The team will synthesize site visit notes from both rounds of site visits into one document. The notes will be organized by topic and respondent type. The text will be loaded into NVivo to identify and summarize themes in findings.

2.2.2 Program Documents

The team will obtain and review key documents from each program pertaining to its design, operating plan, funding, and context. Examples of documents include policy manuals, staff training materials, syllabi, curricula, marketing and recruitment materials, progress reports, aggregate statistical reports, and other relevant documents to help inform understanding of how the program was implemented and the circumstances of the institutions involved. Published labor market and demographic information will also be used to document current and projected occupational demand and the potential supply of workers to meet this demand.

2.2.3 Observations

During the second round of site visits, the team will observe classes and other group activities (e.g., study groups, orientations). If possible, team members will observe individual advising and counseling sessions, if these are part of the program. Observations will be structured to support systematic identification and coding of important characteristics and events associated with each type of activity.³⁰ Study teams will conduct observations during the second site visit.

Observers will use a structured protocol to record key activities. For example, observations of classes will record the proportion of time instructors spend lecturing and working one-on-one with students, the processes instructors use to integrate basic skills and occupational skills (if applicable), the amount of time students spend working on their own and in small groups, and the level of engagement of students in the class/activity. This information will augment the description of the nature and content of instruction and other services provided, as described above.

2.2.4 Program Administrative Data

The team will use administrative data from each program to document treatment group members' participation in activities, the duration and intensity of the services, course/program progress and completion, point of program exit and reason for exit, and any return to the program. Chapter 4 of this report describes this data source in greater detail. Because the team will rely on existing management information systems, the data available will vary by program. Possible items include: program and/or course enrollment, hours attended, program completion, credential receipt, credit receipt, enrollment in additional education, receipt of support services (counseling, child care, transportation, financial

³⁰ Marshall and Rossman, 2006.

assistance), level of support services (e.g., number of counseling sessions, amount of financial assistance), and reason for program exit.

2.2.5 Baseline Information and First Follow-up Surveys of Treatment and Control Group Members

The PACE Basic Information Form (BIF), Self-Administered Questionnaire (SAQ), and first follow-up survey will be important sources of implementation study data. The BIF and SAQ contain information on an extensive set of characteristics of the sample at the point of random assignment (see Chapter 4 for additional information). The BIF generally covers more objective characteristics such as demographics, educational background, and current employment and income, whereas the SAQ focuses on psycho-social factors and potential barriers to work and school. As described in Chapter 4, the first follow-up survey will provide detailed measures of service receipt, as well as a range of educational, employment, and psycho-social outcomes to be analyzed in the impact study. The chief advantages over measures of service utilization from program administrative records, is that the survey will provide measures that are comparable for treatment and control group members and consistent across PACE programs.

2.2.6 Instructor, Manager, and Case Manager/Advisor Surveys

A self-completed online survey of program instructors, managers, and advisors/case managers will supplement the information collected from staff interviews by providing quantifiable data on different aspects of program implementation and operations. From the surveys, the team will develop descriptive statistics to describe the nature of the services provided by program staff and the instruction provided, as well as a range of other factors related to management and staffing. The same instruments will be used in the HPOG National Implementation Evaluation (NIE), which will allow for comparisons to these programs as well.

The case manager/advisor survey collects information on the type and intensity of services staff members provide to students. This includes the type of services provided (e.g., academic and/or non-academic advising, assistance with support services, employment assistance); the number of students they work with; the frequency, length, and mode of interactions with students; and how closely student progress and completion are monitored. The survey also will collect information on staff background and qualifications, staff development activities, service philosophy, staff morale, and perceived effectiveness of program. The manager survey will cover similar questions written to gain the perspective on the same issues from those in supervisory roles. The sample will include the universe of supervisors and case managers/advisors in each program.

The instructor survey will collect data on the nature of instruction provided in the PACE program. Aspects include: class size; the extent to which basic skills are integrated with training instruction; the use of and time spent on different instructional modes (e.g., lectures, small groups, one-on-one, computer-assisted, experiential learning); instructors' role in providing academic and non-academic advising and other supports; strategies for helping students with academic problems; closeness of monitoring of student progress; and materials and resources used in class. Like the case manager/advisor survey, this survey will also collect information on staff background and qualifications, staff development activities, service philosophy, staff morale, and perceived effectiveness of the program. The team will attempt to include all teachers affiliated with the PACE program, although is

some cases where participants can select from a very large number of training programs the focus will be on those most commonly used.

2.2.7 In-Depth Interviews with Treatment and Control Group Members

Another key implementation study data source will be in-depth interviews with small samples of treatment and control group members. This information will support a more detailed understanding of each group's experiences.

The design involves conducting two rounds of interviews with up to 15 research subjects in most PACE programs. The first interview will be conducted in-person during the first 4–6 months after random assignment—the period treatment group members normally still would be receiving services—and the second will occur one year later. As detailed below, sampling will be random within three strata: treatment group members that sustained participation in the program, those with a lesser degree of engagement, and control group members.

The discussion guides will focus on participants' backgrounds and motives for applying to the PACE program. Background characteristics help to understand the prior experiences and unique strengths and challenges shaping subsequent experiences. Examples of background characteristics include family histories, role models, prior education and employment experiences, parenting and other family responsibilities, time constraints, aspirations, and housing and neighborhood context.

For treatment group members who persist in their programs, interviewers will probe factors associated with persistence, including personal/family factors and aspects of programs that may have been especially helpful (e.g., engaging curriculum and teaching styles, support services, supportive community of peers). Interviewers also will explore plans for employment and future training. With regard to treatment group members who do not persist, the interviews will focus on reasons for dropping out (e.g., inability to juggle school and work or family responsibilities, didn't like the program, faced financial constraints) and plans for pursuing postsecondary education and training in the future. Interviews with control group members will focus on options for occupational training and supports and employment experiences. Did control group members find alternative training programs and, if so, how did they choose them and what were their experiences? What perceptions about school, work-related factors, and other considerations influenced decisions to forego training? What kinds of plans do they have for school in the future?

Although both rounds of interviews will focus on education and employment experiences and plans, the first interview will focus more on issues related to participation in the PACE program or other services, and the second will focus on longer-term progress in education and training, employment and career advancement, and financial stability. Both interviews will be approximately one hour in duration. Between the two waves there will be a short telephone check-in to help maintain contact and update school, work, and other important aspects of personal and living situations.

Interviews will be recorded and transcribed. The text will be loaded into NVivo for analysis. At a minimum, three types of coding will be used:

- Attributes—characteristics of study members (e.g., where they grew up and went to school, living situation and neighborhood).
- Descriptive—including key words and categories such as “goals,” “classes,” “costs,” “supports,” and “persist,” and subcategories, such as educational goals and work goals, support from family/friends and support from the program.
- Patterns—by key word/category within treatment and control group member cases and between these types of cases.

2.2.8 Reporting Implementation Study Findings

The team will provide the following reports based on data and analyses described in this chapter:

Program-specific profiles following the first round of site visits. The profiles will be organized to describe key aspects of each program’s logic model, including: (1) program goals, target population, and structure; (2) major intervention components (e.g., the specific assessment, curriculum, support, and employment strategies included); and (3) other services available in the community.

Implementation sections of program-specific follow-up reports. Each section will include a description of the program, documentation of participant experiences in the program, and implementation experiences and lessons. Each will integrate information from both rounds of site visits; supervisor, case manager/advisor and instructor surveys; program administrative data for treatment group members; and in-depth interviews with participants. Each section will include:

- A logic model detailing the processes through which changes in outcomes are theoretically expected to occur.
- A detailed program description, focusing on signature career pathways components. The descriptions will describe the intervention as designed and implemented. The descriptions will include analytical tables that systematically organize the key findings for each program under each main research area.
- Program timelines or chronologies of the key stages of program development and implementation.
- Documentation of participants’ experiences in the program including participant flow through the program, the extent to which treatment group members receive core program components and reach key milestones, and participant perspectives on the program based on the in-depth interviews.
- Implementation experiences and lessons, including those for scaling up and sustainability.

Exhibit 2.4: Principal Career Pathways Program Components

Component, Subcomponent	What to Look For	Implemented (Y/N)? Description of Services (including frequency)	Modifications (timing)	Assessment of the Strength of Component and Unique Aspects to Highlight
Assessments				
Academic and non-academic	<p>Academic skills/interests</p> <ul style="list-style-type: none"> ✓ Basic academic skills ✓ Learning styles/disabilities ✓ Career aptitude/interest <p>Non-academic skills and needs</p> <ul style="list-style-type: none"> ✓ Psycho-social skills ✓ College knowledge ✓ Job readiness skills ✓ Coping skills (personal and family challenges) ✓ Service and support needs. 			
Instruction and Curriculum				
Contextualization	<ul style="list-style-type: none"> ✓ Evidence of integrating applied content into basic skills instruction? ✓ Infusing basic skills into occupational content? 			

Component, Subcomponent	What to Look For	Implemented (Y/N)? Description of Services (including frequency)	Modifications (timing)	Assessment of the Strength of Component and Unique Aspects to Highlight
Acceleration	<ul style="list-style-type: none"> ✓ Reorganization of instruction and curriculum to allow completion in a shorter (calendar) time period ✓ Mainstreaming unprepared students in college-level classes (with additional supports) 			
Flexible delivery	<ul style="list-style-type: none"> ✓ Evening and/or weekend scheduling ✓ Self-paced instruction ✓ Convenient training locations (e.g., in community) ✓ Technology-supported distance learning 			
Active learning	<ul style="list-style-type: none"> ✓ Project-based learning and problem-solving tasks ✓ Involvement of more work in groups ✓ Foster more classroom interaction 			
Supports				
Personal guidance and supports	<ul style="list-style-type: none"> ✓ Intensive or proactive advising or case management ✓ Faculty and/or peer mentors ✓ Group sessions 			

Component, Subcomponent	What to Look For	Implemented (Y/N)? Description of Services (including frequency)	Modifications (timing)	Assessment of the Strength of Component and Unique Aspects to Highlight
Instructional supports	<ul style="list-style-type: none"> ✓ Supplemental academic support such as tutoring ✓ Ad hoc sessions on particular topics ✓ Study groups ✓ Self-paced computer instruction ✓ Curricula addressing psycho-social skills, college knowledge, study skills, job readiness skills, and financial literacy 			
Social supports	<ul style="list-style-type: none"> ✓ Learning communities ✓ Peer/alumni mentors ✓ Teaching social support network building skills 			
Supportive services	<p>Provided in house or via referral</p> <ul style="list-style-type: none"> ✓ Child care ✓ Transportation ✓ Substance abuse counseling/therapy ✓ Mental health counseling/therapy ✓ Domestic violence services 			

Component, Subcomponent	What to Look For	Implemented (Y/N)? Description of Services (including frequency)	Modifications (timing)	Assessment of the Strength of Component and Unique Aspects to Highlight
Financial assistance	<ul style="list-style-type: none"> ✓ Assistance completing financial aid applications and/or identifying sources of support ✓ Direct financial support for education and certification (e.g., tuition, books, test fees) ✓ Direct financial support for other purposes (e.g., reimbursement for child care, transportation or other supports) 			
Connections with Employers				
Role in program design and instruction	<ul style="list-style-type: none"> ✓ Employers involved in designing program/curriculum ✓ Employers providing instruction ✓ Employers hosting visits/activities at work sites 			
Employment experience during training	<ul style="list-style-type: none"> ✓ Cooperative education ✓ Class projects involving simulations of key occupational tasks or real projects for local employers ✓ Internships ✓ Work study ✓ Visits to local employers ✓ Job shadowing 			

Component, Subcomponent	What to Look For	Implemented (Y/N)? Description of Services (including frequency)	Modifications (timing)	Assessment of the Strength of Component and Unique Aspects to Highlight
Employment after training	<ul style="list-style-type: none"> ✓ Transitional or subsidized employment ✓ Apprenticeships ✓ Unsubsidized employment ✓ Job development services 			
Meta Strategies				
Packaging to promote "bounded choice"	<ul style="list-style-type: none"> ✓ Bundling courses in sequences with needed supports ✓ Help participants navigate course offerings and program requirements and admissions and financial aid systems ✓ Help participants obtain other public benefits and social services 			
Creating a continuous improvement ethos	<ul style="list-style-type: none"> ✓ Data systems connecting information from comprehensive assessments, college records, financial aid and other services and benefits ✓ Close monitoring of local economic outlooks, adjustments to training programs based on shifts in demand and technology 			
Moving towards scalability and sustainability				

3. Impact Study for an Individual Program

The PACE team will conduct a separate impact analysis for each PACE program. This chapter presents the general approach to these analyses with a special focus on statistical procedures.³¹ In addition, it discusses the timing for reports. Differences in target populations and logic models across programs are such that a number of decisions will need to be made on a program-specific basis. In this chapter, we identify key points where this more detailed specification is needed, but do not provide program-specific analysis plans. Such plans will be developed prior to beginning analysis for each program.

The chapter begins with an overview of research questions (Section 3.1) and the timing of follow-up surveys and impact reports (Section 3.2). It then identifies several broad classes of hypotheses and associated reporting considerations (Section 3.3). Discussions of basic impact estimation (Section 3.4) and statistical power (Section 3.5) follow. Sections 3.6 and 3.7 provide overviews of strategies for exploring sources of variation in impact and causal pathways, respectively, and Section 3.8 discusses several other analytic issues (e.g., multiple sites in two programs, conditionally observed outcomes, missing data). Section 3.9 provides the framework for the cost-benefit analysis. Section 3.10 discusses plans for writing detailed program-specific analysis plans and some open analysis issues.

3.1 Research Questions

The ultimate research questions concern the degree to which PACE programs positively affect the main outcomes specified in the career pathways theory of change (Exhibit 1.2):

- What is the impact of each program on key indicators of progress in career pathways-relevant training, such as persistence in education and the achievement of certificates and degrees?
- What are the impacts of each program on entry to career-track employment, and earnings?
- What are the impacts of each program on individual and family well-being?

The first round of follow-up and its associated reports (capturing at least 15 months of post-randomization experiences) will focus on early education and training outcomes, with provisional analyses of employment-related outcomes. As discussed in Section 3.2, additional follow-up will be needed to begin to get definitive answers to all three questions.

Inasmuch as programs use multiple strategies and target multiple intermediate outcomes to serve multiple subgroups, the impact study for each program also will focus on questions such as:

³¹ Analysis of the second follow-up (which will be conducted under a separate contract) may include a limited exploratory analyses across PACE sites. See Section 3.8.5.

- Do impacts vary by subgroups or other moderators of impact, and, if so, which characteristics are associated with larger or smaller effects?
- Are there impacts on program services and/or intermediate outcomes that seem to be acting as pathways to observed impacts on attainment of certificates and degrees, or earnings?

The program services of interest will be those that are not systematically offered to all program participants. The team will also study the systematically offered services as part of the implementation study, but we will not attempt to study their impact on student outcomes.³²

3.2 Time Horizon for Impact Analyses

The career pathways theory of change distinguishes several basic categories of hypotheses that apply to every PACE program. Each program aims to promote completion of one or more postsecondary training steps leading to employment and improved earnings and advancement potential. It does this by providing a set of components (assessment, skills and vocational training, supports, employment connections) addressing intermediate outcomes thought to be related to successful completion of training. These intermediate outcomes—organized into six general domains—are expected to show improvement (e.g., more favorable levels for treatment than control group members) and thereby lead to increased performance and persistence in training; improved performance and advancement in jobs; and improved adult, child and family well-being.

Since PACE programs vary in length and in the kinds of credentials that they target (e.g., one-year certificates versus two-year associate’s degrees), they also vary in the expected timing of impacts. For the sake of consistency and future meta-analytic projects, however, the PACE plan provides measures and analyses at the same follow-up intervals across sites. Using the same core measurement protocol³³ for all nine programs provides substantial evaluation cost efficiencies. Common measurement also provides an improved basis for likely future analyses involving a larger number of programs.³⁴

The first follow-up period captures at least the initial 15 months after random assignment. This time point allows the majority of participants to have completed a substantial portion of their intended training but is short enough to allow them to accurately report on their training experiences. Interviewing participants soon after initial training experiences is important in ensuring that their

³² Readers interested in the effects of systemic program features on student outcomes will want to look for results from the HPOG Impact Study, as discussed in Section 1.7.

³³ As is discussed in more detail in Chapter 4, this common protocol does not include administrative data. The team has not imposed a common protocol for reporting of program data and cannot impose one on states for reporting of enrollment and degree awards.

³⁴ As explained in Chapter 1, the HPOG Impact Evaluation is planning such analyses, and at least three PACE programs (those receiving HPOG grants) will be included.

memories of these experiences are still fresh. Employment and earnings during periods of substantial training involvement often will be depressed as participants shift time into school in expectation of higher future earnings. For this reason, short-term impacts of education and training programs on employment and earnings are often negative and not necessarily indicative of longer-term results.³⁵ As discussed below, first follow-up reports will provide employment and earnings estimates, treated as secondary, rather than confirmatory, analyses for all but one of the sites.

Consistent with early specifications from ACF, subsequent rounds of follow-up and analysis at 36 and 60 months will be needed to test hypotheses for middle- and longer-term impacts, respectively. The emphasis in these cycles would shift to receipt of longer-term credentials (e.g., associate's degree), employment and earnings, and (at 60 months) other life outcomes that plausibly might be influenced by better working conditions and income (e.g., life satisfaction and mental health, child well-being).

Across sites, analyses for each follow-up period will address the same domains and use data from a single survey questionnaire and the same administrative databases (principally wage records in the National Directory of New Hires and college enrollment records in the National Student Clearinghouse). Given variation in programs' target populations and logic models, specific outcomes constructed from survey data will be tailored to provide optimal tests for each program. The following section identifies a few key areas where tailoring will occur, particularly in measures of educational progress and career-track employment, and Section 3.10 notes that upcoming analysis plans will provide details for specific sites.

3.3 Classification of Research Hypotheses for Testing and Reporting Purposes

As described in Chapter 1, the career pathways framework encompasses a wide range of program inputs and outcomes. Given the large number of impacts which are measurable and of interest and will therefore be the subjects of formal hypothesis tests for program effects, it is critical to structure the analysis in such a manner as to avoid over-interpretation of the large number of false positives that would arise absent a plan to address a multiplicity of tests.³⁶

Emerging approaches to this “multiple comparison problem” typically distinguish two categories of analyses.³⁷ The first category involves tests of “confirmatory” hypotheses— analyses of the most critical

³⁵ This is clearly observed in Roder and Elliott (2011) where negative impacts on earnings were observed for the first four quarters following randomization among students randomized to the 12-month intervention.

³⁶ For example, if a program is totally ineffective but 100 tests are run on various outcomes and subgroups, it is expected that $100p$ of these will falsely indicate evidence of a program effect, where p is the significance level used in the testing. If the outcomes are independent of each other, the actual number of false positive findings will be close to $100p$, but if the outcomes are strongly correlated with each other, it is possible for all or none of the significant results to be false positive findings.

³⁷ Schochet, 2008.

effects programs seek to produce and for which findings carry the most consequence. Given that these are the results on which consequential decisions are likely to be made, there is increasing recognition that significance levels should be adjusted to account for multiple comparisons so as to avoid using societal resources on ineffective programs. When sample sizes are modest—as in PACE—such adjustments can take a substantial toll on statistical power, making it harder to detect impacts that do occur. Accordingly, evaluators typically specify a very small number of confirmatory hypotheses.

As multiple comparison adjustment is relatively recent in social policy research, there are no universally accepted standards about when adjustment is necessary. Some organizations have adopted the standard that adjustment should occur whenever there is more than one confirmatory hypothesis. However, others, such as the What Works Clearinghouse (WWC), sanction multiple confirmatory hypotheses without adjustment provided that there is only one confirmatory hypothesis per domain.³⁸ A rationale for this is that different domains are of interest to different audiences, or that they will be used for different purposes, so that there is no danger that decisions will be made on the basis of significant findings in any domain. The strength of protection against misusing societal resources then depends on broadness with which domains are defined. For its career pathways studies, including PACE, ACF adopted the WWC standard.

The second category covers “exploratory” analyses, a broad category including impacts for a wide range of outcomes and subgroups where findings are less fundamental to the basic question of whether programs were successful. Views of the need for multiplicity adjustments in such analyses vary, and there is generally less of a premium on setting stringent statistical tolerance levels. In part, varying perspectives on exploratory analyses reflect the broad mix of research questions that may be treated as non-confirmatory. For PACE analyses, it will be useful to distinguish two types of non-confirmatory analyses: tests of a limited number of “secondary hypotheses” connected to key assumptions in program logic models, and a larger number of “exploratory analyses” in which expected signs on experimental estimates are theoretically ambiguous or nonexperimental methods are used to analyze statistical mediation. We discuss our approach to analysis and reporting for each of these three categories in a separate section below.

One feature of both confirmatory and secondary hypotheses is that, with the exception of a very limited set of subgroup analyses (see Section 3.6), they will all be one-sided with no ability to distinguish harmful effects from null effects, while all exploratory hypotheses will be two-sided.³⁹ If the theory linking a program to a specific outcome is not sufficiently developed to clearly identify the expected

³⁸ See Section B of Appendix G of Version 3.0 (undated) of the What Works Clearinghouse Procedures and Standards Handbook, currently available at the web site of the Institute of Educational Sciences of the U.S. Department of Education.
http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

³⁹ See Section 3.4 for more discussion of the choice between one-sided and two-sided hypotheses.

direction of an effect of the program on that outcome, then any hypothesis about the impact of the program on that outcome will be treated as exploratory.

This classification of hypotheses will be reflected in various planned reporting. One-page executive summaries will only include results from confirmatory analyses. Longer executive summaries may also include the results from secondary analyses—especially where there is a pattern of effects that are consistent with each other and with established positive estimates for the confirmatory hypotheses—but will highlight findings from confirmatory analyses as key evidence of whether the program “works.” The full impact report for a site will assign each category of finding to a separate chapter or section, with clear discussion of limitations and appropriate weight readers should assign to findings.

The following three subsections provide more detail on each class of hypothesis.

3.3.1 Confirmatory Hypotheses, Multiple Comparison Adjustment and Composite Outcomes

Hypotheses in this category concern impacts on “primary” outcomes, a subset of the main outcomes articulated in the theory of change for the follow-up point. To keep the risk of false positive findings close to the stated p-values, avoid the need for adjustment (given WWC standards) and attendant loss of power, and simplify interpretation, ACF decided that analyses should limit confirmatory tests to no more than two per program per report, and no more than one per major outcome domain.

For analyses of the first follow-up period (between 1-2 years after random assignment) we will focus on a single confirmatory hypothesis in the education domain. The exact outcome will vary across programs, depending on their theories of change, but will be a simple and well-recognized outcome such as cumulative credits earned or degree/certificate award. In addition, where the local theory of change suggests the strong possibility of early positive earnings effects, a second confirmatory hypothesis about earnings in the fifth quarter following randomization will be tested.

At the second follow-up period (about three years post-random assignment), there will be two confirmatory hypotheses—one in the education domain and one in the employment and earnings domain. As at first follow-up, the primary outcome in the education domain at second follow-up will depend on each program’s theory of change. It again is likely to vary across programs and may differ also from the outcome declared confirmatory for the program at first follow-up. The primary outcome for all programs in the domain of employment and earnings will be earnings in the twelfth quarter following randomization. This is the latest quarter that we think will be available at about the same time as data from the second follow-up survey will become available. The motivation of selecting a single quarter as late as possible was to maximize the proportion of sample members who might be past an initial period of expected negative earnings effects due to intentional reduction of labor force participation during training.

Confirmatory analyses will not adjust for multiple comparisons across the nine PACE programs or across points of time for the same program. The aim is not to generalize to career pathways programs in general or beyond the time periods covered; rather, analyses will focus on each individual program and the longest available follow-up period available. Following the WWC in this respect, the team will also

not adjust for multiple comparisons across the two domains. Since there will be only one primary outcome per domain, there will not be any multiple comparison adjustments in these reports.

The PACE team will designate confirmatory outcomes at 60 months as part of the development of program-specific analysis plans. It is likely that they will address hypotheses concerning cumulative earnings and educational attainment.

3.3.2 Secondary Hypotheses

Secondary analyses will consist of tests of a small set of subgroup differences in effects on primary outcomes and tests of overall program impacts on a wider range of outcomes. The latter include: more varied measures of progress in the domains of education and employment/earnings; “intermediate outcomes” in the career pathways theory of change; and additional main outcomes related to primary outcomes but less central to testing confirmatory hypotheses. As mentioned earlier, with the exception of subgroup analyses, all secondary hypotheses will be one-sided. All will be named in the analysis plan for the program before any program effects are estimated on first follow-up data. The reports will show both those with and those without significant program impacts. By way of illustration, secondary hypotheses might concern impacts on:

- *More varied measures of primary outcomes.* The simple measures selected to be primary will not capture all potential dimensions of impact in key domains. The research team is developing measures of progress in education and progress in obtaining career track employment, as discussed below.
- *The primary outcome in subgroups of high programmatic interest.* One example of this is VIDA, where the program randomized individuals not enrolled in a college program and ongoing (enrolled) students, so it would be of programmatic interest to compare effects on the two groups. As another example, consider groups defined by baseline academic assessments. Finding an optimal lower bound on these assessments that could be used as a criterion in the admissions process would be an exploratory analysis, but the program has already defined thresholds on these assessments for who needs to take developmental classes. That threshold could be used to define subgroups for secondary analysis. Again, to discipline the investigation, subgroup analyses in the secondary hypothesis category will be limited to a small number of characteristics.
- *Intermediate outcomes in the theory of change for the program.* Psycho-social factors—for example, a sense of belonging at school—are one likely candidate for this category of analyses in the first follow-up reports.
- *Primary outcomes at off-target points of time.* A program with a long duration, for example, might not be expected to have strong positive impacts on employment as early as 15 months (in fact, they may well be negative), but since policy makers are often interested in getting answers as early as possible about the impact of a program on the long-term goals, the team will likely include employment and earnings as secondary outcomes in the first follow-up reports.

Where appropriate and feasible, the team will use factor analytic methods to reduce the number of non-primary outcome dimensions to fewer indices and their associated hypotheses. Indices will be either pre-specified based on theory or derived empirically using a version of the research sample that omits the treatment/control status of study subjects.

Similar to the analytic procedures planned for confirmatory hypotheses, the team will rely on limiting the number of analyses to control for false positive findings. However for secondary hypotheses, the limitations will be less strict in that the team will allow a careful, but more generous, pre-specification of outcomes (as outlined above) and a moderate number of tests to constrain the risk of false positive findings. These findings will be presented in their own report chapters separate from confirmatory analyses. The introductory section of the chapter on secondary hypotheses will clearly articulate the risk of false positive findings posed by running uncorrected one-sided hypothesis tests for all of them. It will then be left to the reader to decide whether the collection of findings in the section constitutes reasonable evidence that the program is having effects on important peripheral outcomes. If a coherent pattern of significant impacts on secondary outcomes emerges, the team will argue that this will increase the trust the reader should place in the findings. Results of tests of secondary hypotheses will fall into one of two categories, either worthy of being the basis for judging the success of the program when they are part of a consistent pattern that includes positive confirmatory findings, or useful only for refining a future research agenda by suggesting ways to change current programs for future testing.

Whereas primary outcomes for confirmatory hypotheses will be simple, intuitive measure, it is also of value to explore effects on more complex measures tailored to the career pathways theory of change. Two central domains in this theory of change pose especially challenging definitional issues. The first follow-up questionnaire was designed to support the development of more sensitive measures of progress in these domains than the current standard measures. The issues and possible directions, to be worked out in program-specific analysis plans, are summarized below.

Conceptualizing “progress in education and training.” The career pathways theory of change presented in Chapter 1 used the shorthand of “persistence and attainment” leading to occupationally relevant certificates and degrees to characterize a focal domain whose operational specification involves substantial complexity. In brief, great diversity in programs and credentials makes it difficult to arrive at parsimonious, widely applicable definitions for persistence and attainment. At a minimum, impact analysis requires measures that are valid and have the same meaning for all treatment and control group members and over varying durations of research follow-up. It also might be convenient to have consistent measures across PACE programs. However, plans to analyze and report on each program separately do not require consistency; in fact, they acknowledge differences in program designs and target populations, implying that different credentials and markers of progress towards them will be appropriate.

For some PACE programs, target credentials with fairly wide applicability to treatment and control groups provide suitable outcomes over the mid to longer range. For example, the I-BEST, Madison Area Technical College and VIDA programs place substantial emphasis on helping participants to achieve associate’s degrees—albeit with greater emphasis on intermediate certificates in the first two. Some explicitly target multiple levels in relatively highly differentiated career pathways in a particular

occupation, such as health (e.g., Carreras En Salud at Instituto, Pathways to Health Care at Pima Community College). Other programs aim primarily to boost receipt of shorter-term college certificates and occupational credentials (e.g., WTA Connect at DMAAC) or college persistence more generally (e.g., Year Up). Still others aim at a diverse array of credentials, including training at colleges and potentially more widely at proprietary and other non-college institutions (Bridge to Healthcare at San Diego Workforce Partnership, Health Careers for All at Workforce Development Council). Though control group members not entering these programs on average share similar educational aspirations at the outset, they potentially may pursue an even wider range of types and levels of education and training.

Measures of educational persistence and attainment must reflect outcomes likely to have occurred in successful programs at successive follow-up stages. The brief sketch above suggests that appropriate short-term measures of educational progress also will vary considerably across programs. For example, at the first follow-up, persistence in college and accumulation of regular college credits are likely to be the most valid indicators for VIDA, I-BEST, and MATC, while elsewhere more general measures of hours of school/training, completion of short-term occupational credentials, or enrolling in any regular college class may be most appropriate.

Expected diversity in types of programs and rates of progress against a variety of credentials within both treatment and comparison groups suggest the potential benefits of indices summarizing instructional hours and credentials. Such indices would convert varying kinds of hours (e.g., college credit/non-credit, non-college) and credentials (e.g., short and long-term) into a comparable metric.

Facing a similar problem in measuring productivity in higher education, a National Academy of Science panel recently issued a general recommendation along these lines—based on an index of “adjusted hours” combining credits and estimated increments (in hour units) from credential receipt.⁴⁰ NAS proposed to investigate methods for translating credentials into hour units based on estimates of their incremental utility (that is, beyond the value represented by credits alone). Recognizing the substantial challenges in defining and estimating incremental utility, the panel acknowledged that considerable work was needed to create valid performance measures. For example, the panel felt that basing weights on economic utility raised large conceptual and empirical (data and analysis requirements) challenges.⁴¹ It concluded that more technical work was needed to devise and test measures suitable for high stakes applications like performance measurement.

In PACE, the requirements are somewhat broader given the importance of counting instruction not resulting in regular college credit (at colleges as well as other kinds of institutions) and a wide range of types of credentials (e.g., professional certifications and occupational licenses). It currently seems most

⁴⁰ Sullivan et al., 2012.

⁴¹ Utility here would be lifetime benefits from the training, education, or credential. Benefits would include both material benefits (e.g., earnings) and immaterial benefits (e.g., satisfaction with the nature and contributions of one’s work).

feasible to establish, in each site, a limited number of discrete measures based on measures of hours, persistence (allowing for breaks in enrollment consistent with the flexible entry/exit notion in career pathways), and credentials most relevant to each program and target population. At the same time, it also seems worth investigating further the potential for joint measures for individual sites.

This work will be guided by the information collected in the site visits and other aspects of the implementation study. The team will select the primary outcome measure for each program that will be most sensitive to its goals for participants at the first follow-up. Looking at program-specific target populations and interventions, the team will determine whether a weighted composite measure of substantial educational progress or, alternatively, a small number of discrete outcomes (such as credits, non-credit instructional hours, and credential attainment) would be most appropriate. The PACE first follow-up survey provides substantial detail on these outcomes.

Measuring “career-track employment.” Accession to, and progress in, career-track employment are primary objectives in career pathways programs. The concept of career-track employment does not have a widely accepted operational definition. It is generally understood to denote employment in fields requiring at least some postsecondary training and providing opportunities for growth in skills, responsibilities, earnings and satisfaction.

The PACE survey will obtain information on occupation sufficient to code to standard U.S. Census categories. There is a great deal of secondary data and analysis on skills, education, and wages associated with these categories and their international counterparts. Notably, the U.S. Bureau of Labor Statistics O*NET databases link occupational codes to skill and education requirements, and there has been extensive estimation of the degree to which specific occupations require skills needed in a knowledge economy, the extent of postsecondary education possessed by average incumbents, and average wages.

It thus should be possible to create one or more indicators of the degree to which PACE survey respondents are working in “career-track employment.” The indicators likely would be operationalized as joint outcomes summarizing both employment and whether employment was career-track based on one or more occupation-based criteria (e.g., requisite training, skills, and typical wage growth).

Joint measures of career pathways progress summarizing education/training and employment dimensions. Given that progress in career-oriented education and employment are both primary hypotheses, it may be desirable to construct a more general measure of progress in career pathways summarizing the two. Such an approach has two potential virtues. First, it helps to address the central idea in career pathways of closely linked progress in education and employment. Because the time paths to impacts on the two fronts can vary across individuals and programs, a joint measure appropriately credits impacts occurring on either front. Second, where appropriate, combining indicators within a single domain is a recommended approach to avoiding loss of power due to multiplicity.⁴² The simplest

⁴² See, for example, Schochet, 2008.

such outcome would be a measure summarizing whether there has been a positive outcome in neither, either, or both domains.

There is also the possibility for interesting analyses of progress in career pathways based on two new items developed for the first follow-up survey:

I am going to read you two statements. Please tell me whether you would say you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with the following statements:

		Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree	REF	DK
a.	I am making progress towards my long-range employment goals.	<input type="checkbox"/>					
b.	I see myself on a career path.	<input type="checkbox"/>					

As these items use a common scale, they readily can be combined across respondents, programs, and time periods (including periods respondents are and are not in school). The strength of this approach is that individual satisfaction is, in itself, an important outcome—and individual respondents may be in the best position to gauge progress in meaningful terms. The most obvious validity concern is that the questions elicit subjective assessments for which underlying frames of reference are unknown (and may vary greatly across individuals and groups). It thus is important to investigate their psychometric properties through analysis of correlations with more objective measures of progress in education and employment at the first follow-up. If results are promising, including these measures in analyses of secondary hypotheses could be informative.

3.3.3 Exploratory Hypotheses

All hypotheses of program effects on outcomes not treated as confirmatory or secondary hypotheses will fall into this classification, as will most hypotheses concerning differential program effects on subgroups (Section 3.6), all hypotheses concerning mediation (Section 3.7) of program effects via program components and/or intermediate outcomes, and all hypotheses concerning program effects on conditionally observed outcomes (Section 3.8). The number of exploratory hypotheses is likely to be large (e.g., several hundred hypotheses, including subgroup effects, per program).

With this many hypotheses being tested, a considerable number of false positive findings are to be expected, but the team will not make any adjustments to control the number of false positive findings other than switching from one-sided tests to two-sided tests. Instead, report readers will be warned that many of the findings may reflect chance variation rather than true impacts, and that all the findings in this chapter of the report should be used primarily as a guide for shaping future research agendas. These findings can guide future research by distinguishing “dry wells” – approaches that do not show promise – from those with potential impacts. Further, for subgroup analyses, even exploratory findings may be useful for program administrators who must restrict admissions due to excess demand. In such cases, even guidance that may include some false positives may be better than rules that would otherwise be used to restrict admission.

3.4 Basic Methods for Estimating Impacts

This section describes the basic impact estimation methods, including the inferential frame, the use of models, the selection of covariates, and standards of statistical evidence. It also discusses how the reports will present findings.

3.4.1 Inferential Frame

The main PACE findings will represent the average effect of granting access to the career pathways program for the whole population determined to be eligible. Since analyses will include all individuals who were randomly assigned, regardless of participation in the program, these analyses will constitute estimated effects of the intention to treat (ITT).⁴³

3.4.2 The Use of Models

The team will estimate impacts using a model that “regression adjusts” the difference between average outcomes for treatment and control group members by controlling for exogenous characteristics measured at baseline. Controlling for baseline covariates reduces distortions caused by random differences in the characteristics of treatment and control group members and thereby improves the precision of impact estimates. It also helps to reduce the risk of bias resulting from attrition (i.e., individuals failing to complete follow-up surveys).⁴⁴ As is discussed in more detail below, there is no need to correctly specify the shape of relationships between covariates and outcomes to obtain these benefits. Valid results are obtained even with misspecified models.

Experimental impacts for binary⁴⁵ and continuous outcomes will be estimated by the coefficient on a 0/1 treatment variable in a linear regression model such as the following:

$$Y_i = X_i\beta + \delta T_i + e_i, \quad (3.1)$$

⁴³ The team will also prepare estimates the impacts of the treatment on the treated (TOT)—that is, participants who at least initially appear for service, or receive a specified minimum service dosage. Stronger assumptions must be made to estimate these. The estimating approach is given in Section 3.8.2.

⁴⁴ If attrition is varied across treatment and control groups, as is often the case given that the provision of services creates opportunities for study researchers to stay in touch with members of the treatment group, and if the baseline covariates are related to both the propensity to attrit and to the outcome of interest, then controlling on those covariates will reduce nonresponse bias in the same way that covariate adjustment reduces selection bias in an observational study. The team will also take other steps to reduce this risk, as discussed in Section 3.8.4.

⁴⁵ The team considered using logistic or tobit regression for binary outcomes, but communication of results is much more challenging when these models are used.

where Y_i is the outcome, X_i is a row vector of baseline covariates, β is the vector of parameters indicating the contribution of each covariate to the outcome, T_i is a 0/1 dummy variable indicating treatment group membership, δ is the effect of treatment, and e_i is an error term that is independently and identically normally distributed across study members.⁴⁶ The team may analyze experimental impacts for ordered categorical outcomes in the same manner, or may break them up into one or multiple binary outcomes and then analyze them in that form.

3.4.3 Covariate Selection

As described in Chapter 4, data collected at baseline provide a rich set of candidates both for inclusion of covariates and for subgroup analysis described below in Section 3.6. Baseline survey data include: demographic information, such as race and gender; educational history of the individual and his/her parents; and scores on a number of psycho-social scales, such as academic self-confidence and discipline, social support, personal and family challenges and depression (each score constructed from 7 to 12 responses). Administrative data include prior earnings from the National Directory of New Hires and scores on the program's academic assessments (if available).

There is a long-standing tradition of adjusting for covariates in social experiments, and a recent paper lends formal support to this tradition.⁴⁷ However, this paper also reiterates one point stressed by critics of covariate adjustment: purposeful covariate selection can alter estimates of program effects and thus it is vital that the selection of covariates be made in a manner blind to treatment effects. Also, it is important to avoid strong colinearity among covariates.⁴⁸ In PACE, there are too many available covariates to consider using all of them and many of them will be subject to strong colinearity, so the team will be careful in selecting them.

One standard approach to “blind” selection of covariates is to select them in advance based on their expected correlations with key study outcomes and each other. With this approach, it is impossible to steer effects in directions desired by the analyst. However, this approach would likely affect power negatively, since it is difficult to tell in advance which of the potential covariates will be the best predictors of the outcomes. An approach that is more powerful but still rigorous is to fit separate models for outcomes on the treatment and control samples using both personal judgment and formal

⁴⁶ The errors for binary outcomes cannot be normally distributed, but research such as Judkins and Porter (2013) indicate that this ordinary least squares (OLS) model will work well for non-normal outcomes. This research demonstrates that for unclustered data, it is not necessary to use robust standard errors in testing or to use alternate robust testing methods unless sample sizes are very small and the binary outcome is very rare. As long as five events are expected in both the program and control groups, standard OLS inferences are valid.

⁴⁷ Lin, 2013.

⁴⁸ If some of the covariates are strongly correlated with each other or linear combinations of each other, a critical matrix that must be inverted to obtain program effect estimates can become “ill-conditioned.” Ill-conditioned matrices are difficult to invert and some parameter estimates can become very unstable.

covariate selection algorithms such as “stepwise” selection and then to take the union of the variables that are selected for the two separate models.⁴⁹ The team will use this latter approach as it likely will yield the greatest power, but since it is implemented prior to estimating impacts, this approach prevents shopping for favorable covariate decisions. The team will do this for all the primary and secondary outcomes and then take the union of all covariates selected for any of the outcomes. Within the union, the team will check for any instance of strong collinearity. If such pairs of variables are discovered, the team will only keep one variable from each collinear pair and then use this single set of covariates for all substantive models in the report. Having this consistent set of covariates will simplify programming, quality control, and documentation. To the extent possible, the team will seek to utilize a consistent set of covariates for all programs, although the availability of different administrative data⁵⁰ across programs will likely result in the covariates not being identical in all of them.

3.4.4 Standards for Statistical Significance

There are several common standards for judging statistical significance. The choice among them depends on evaluators’ tolerance for false positive findings (Type I errors) and false negative findings (Type II errors). In the PACE impact reports, tests will be considered statistically significant in the text and highlighted in tables if the p-value is less than or equal to 0.10. Tests with smaller p-values will be separately flagged and discussed as providing various degrees of strength of evidence. More specifically, one star for 0.10 will designate “suggestive” evidence, two for 0.05 will designate “moderate” evidence, and three for 0.01 will designate “very strong” evidence.

A p-value threshold of 0.10 for associated tests to be declared as statistically significant is a less stringent standard than the threshold of 0.05 used in most pharmaceutical and medical device trials and in many social experiments. In those experiments, decisions are being made to approve a drug or device for universal usage and public funding or perhaps to continue an expensive national program. In that context, a low risk of false positive findings is desirable in order to avoid massive and decade-long societal expenditures on ineffective products, and, therefore, the 0.05 threshold is appropriate. In contrast, when deciding whether further development and testing of a promising innovation is warranted, the less conservative 0.10 standard of evidence of impact is appropriate since the societal burden of continuing to research such programs is small. Most PACE programs fall into the latter category, although at least two (Year Up and VIDA) may be closer to the former.

⁴⁹ This is similar to the approach suggested by Tsiatis et al. (2008). There are a number of variations on this basic schema that have been considered such as only using the control sample to select covariates or using separate analysis teams to do the work on the treatment and control samples.

⁵⁰ Most administrative data will not be eligible to use as covariates since they will only be observed on the treatment sample, but baseline assessments of college readiness and basic academic skills are an important exception. When these are given consistently to both groups prior to randomization, the team will consider using them.

As mentioned earlier, one-sided tests will be used for confirmatory and, for the most part, secondary hypotheses while two-sided tests will be used for exploratory hypotheses. A one-tailed test is more sensitive to beneficial program effects but is appropriate only when it is not important to be able to distinguish harmful program effects from no effect. The team believes all of the PACE confirmatory and secondary hypotheses (except for the limited number of subgroup analyses in the latter category) are of this kind.⁵¹

In exploratory analyses, the team will apply two-tailed tests. These involve mainly differential impacts by subgroups where we expect to have no a priori grounds for judging which subgroups will experience stronger effects. While it might be possible to argue that some of the other exploratory hypotheses could be appropriately set up as one-sided, it will reduce confusion if the same type of test is used for all the exploratory analyses.

3.4.5 Presentation of Impact Findings

In some journals, it is common to report estimates of both β and δ , even though only the latter are of substantive interest. However, in the field of evaluation of federal programs it is more common to relegate findings for the full model to technical appendices and just present estimates of δ in the main text, perhaps after rescaling as effect sizes.⁵² However, estimates of δ will be more useful if they are presented in the context of the level of the outcome under treatment and control conditions, and thus we will include both in tables like Exhibit 3.1.

Exhibit 3.1: Mockup of Effect Reporting

Outcome	Program Group	Control Group	Effect of Program	Standard Error	p-value
Attained an Associate's degree. within 60 months after random assignment	XX.X	YY.Y	-/+ZZ.Z	WW.W	.VVV

As noted earlier, the reports will highlight those program effects that are judged to be significant (p -value less than 0.10) by placing a star next to the point estimate. In addition, devices will be used to draw more attention to those effect estimates with smaller p-values. Those estimates of program effects with p-values less than 0.05 will be flagged with two stars and those with p-values less than 0.01 with three stars. Highlighting three levels of statistical significance will allow readers with different

⁵¹ Furthermore, in the near term, it seems unlikely that any of the programs could lead to lower attainment of postsecondary education than otherwise would be the case. In the long term, the team expects all programs to raise earnings even if there may be some counter-effects on earnings while the participants are in training.

⁵² This is typically done by dividing the estimated program effect by the pooled uncorrected standard deviation of the outcome within the treatment and control arms. This is known as Cohen's d . For a discussion of this statistic and related approaches to standardizing effect measurement, see Rosenthal (1984).

tolerances for Type I and Type II errors to quickly scan tables for program effects that meet their own evidentiary standards. The reports will also provide standard errors so as to facilitate better understanding of null findings.

Although most of the column headings are self-explanatory, a clarification in what will be entered in the program and control group columns is needed. The program group column will show the mean observed outcome in the treatment group, such as the percent who obtained a credential or the mean number of college credits earned. The control group column, however, will not show the actual mean response among control group members. Showing the actual mean for the control group in this column would be confusing because the difference in actual means between the treatment and control groups will not be equal to the estimated treatment effect. To avoid this potential confusion, the control group column will include regression-adjusted projections of what would have been observed in the treatment group if there had been no intervention. With this convention, if a reader calculates the difference between the two columns, they will obtain the estimated program effect. The reports will provide a note to impact tables explaining this procedure.

3.5 Minimum Detectable Effects (MDEs)

It is important to calculate minimum detectable effects (MDEs) before beginning an evaluation to ensure that the design provides samples large enough to detect impacts that matter to policy makers and practitioners. The MDE is the smallest true effect that a study will be able to detect at specified levels of power and statistical significance. Power refers to the probability of detecting a statistically significant impact of given size when it exists (i.e., avoiding Type II error) and typically is set to 80 percent. The statistical significance level in a hypothesis test equals the probability of rejecting the hypothesis of no impact when it is correct and there really is no impact (i.e., making a Type I error). As discussed earlier (in Section 3.4.4), the standard for statistical significance in PACE will be .10, though tables will report whether results meet more stringent levels (0.01, 0.05) as well. This section discusses the MDEs pertaining to PACE samples.

3.5.1 MDEs for Tests of PACE Programs

Exhibit 3.2 presents MDEs for three sample sizes—one for sample size targeted in most PACE programs, one for a program falling short in recruitment, and one for the larger sample recruited for Year Up.

Exhibit 3.2: Minimum Detectable Effects for Confirmatory Hypotheses

Statistic	Percent with Substantial Educational Progress (Confirmatory at First Follow-up)	Average Quarterly Earnings (Confirmatory at Second and Third Follow-ups)
MDE for Sample Sizes with:		
500 T: 500 C (most programs)	6.2	\$324
300 T: 300 C (recruitment shortfall)	8.0	\$418

1695 T: 848 C (Year Up)	4.1	\$207
Control Group Mean	50.0	\$2,863

Note: MDEs based on 80% power with a 10% significance level in a one-tailed test, assuming baseline variables explain 15% of the variance in the binary outcome and 30% of the variance in earnings. We have set the variance for credential attainment conservatively at the highest possible level for a dichotomous outcome, 25% (based on the assumption that 50% of the control group experiences the outcome). The variance estimate for earnings in both sample size categories comes from special tabulations of survey data from the second follow-up year of a small random assignment test of Year Up.⁵³

The exhibit shows that PACE will be able to detect impacts on the percentage making substantial educational progress (e.g., receiving specified credentials) as small as 6.2 percentage points for most programs and 4.1 percentage points for the Year Up program. The corresponding MDEs for quarterly earnings are \$324 and \$207, respectively.⁵⁴ Assuming a recruitment shortfall of 40 percent of the general target of 1,000, the MDE for substantial educational progress will be 8.0 percentage points, and the MDE for quarterly earnings will be \$418.

These MDE estimates assume data will be available for 100 percent of sample members, as will be the case when administrative data are used to measure outcomes. MDEs for outcomes based solely on survey data will be larger—typically about 12 percent higher (in relative terms) than those shown in Exhibit 3.2.

3.6 Subgroup Analysis and Other Types of Effect Moderation

An earlier section discussed the important category of hypotheses addressing potential differences in impacts for research subjects with different personal and family characteristics at the start of the study (i.e., the point of random assignment). This section outlines several aspects in specifying and estimating subgroup differences.

Hypothesized moderators can be specified in different ways. Consider possible specifications of father’s educational attainment, for example. One specification would be as a binary variable indicating whether the father had a college degree. Another might be to code father’s educational attainment on the basis

⁵³ These earnings variance estimates (derived from standard deviations of \$12,748 and \$10,160 in annual earnings for treatment and control groups, respectively) were the only variance estimates available for populations actually served in PACE sites. Though based on a small survey sample (120 treatment, 44 control), they are very close to estimates P/PV provided PACE for participants in its sectoral demonstration project, which involved a wider age range than the youth (18–24) targeted in Year Up. The projected variance reductions due to use of baseline variables are from Nisar, Klerman, and Juras (2013).

⁵⁴ As discussed earlier in this chapter (Section 3.3.1), the team will define measures of progress in training as applicable to each site’s target population at specified follow-up intervals. The team anticipates that credit hours and credential receipt will figure prominently in the definitions.

of years of school completed—say on a scale of 1 to 19—and treat this scale as if it were interval-valued. A third option would be to define an indicator variable for each of several levels of father’s educational attainment, such as high school graduate, associate’s degree holder, and bachelor’s degree holder or higher, leaving one level out as the comparison category (in this example, less than high school graduation). Expanding on equation (3.1), let X_i^d denote the value of a specific variable or cluster of indicator variables for the levels of a categorical variable from X_i , the row vector of baseline covariates. The moderating effect of X_i^d on treatment will be estimated using the model:

$$Y_i = X_i\beta + \delta T_i + T_i X_i^d \gamma_d + e_i, \tag{3.2}$$

where γ_d represents the moderating effect of X_i^d . A standard two-sided t-test will be used for testing whether $\gamma_d = 0$ if X_i^d is binary or continuous. If X_i^d is categorical with three or more levels (such as a three-level racial classification), then a standard F-test will be used for testing whether all the components of $\gamma_d = 0$.

For continuous moderators (e.g., baseline skills assessment scores), analyses will test for nonlinearities in the interaction of treatment with the moderator. This might happen, for example if the program is most effective for the students in the middle range of the entrance exam but is largely ineffective for students at both low and high levels. Tests of nonlinearities will use categorical specifications for moderators, corresponding to the quintiles of the moderator and then fit models like:

$$Y_i = X_i\beta + \delta T_i + T_i X_i^d \gamma_d + \sum_{Q=1}^5 T_i X_i^{Qd} \gamma_{Qd} + e_i \tag{3.3}$$

In this model, γ_d is the overall linear interaction of the moderator with the treatment and γ_{Qd} is how the interaction differs in the Q-th quintile. An F-test will then be run for $\gamma_{Qd} = 0$ for all five quintiles. Given a finding of nonlinearities, differences will be presented by quintiles. Compared with continuous forms, presenting results for quintiles makes it easier for analysts and readers to identify subgroups for which programs may be relatively more or less promising.

Absent nonlinearity (that is, on failing to reject the hypothesis that $\gamma_{Qd} = 0$ in all components), presentation will center on linear effects (e.g., students at the zero level of X_i^d experience a program effect of $\hat{\delta}$ and that effect changes by $\hat{\gamma}_d$ for every unit increase in X_i^d). Note that $\hat{\delta}$ and $\hat{\gamma}_d$ are sensitive to the location and scale of X_i^d . If there is a natural scale for X_i^d , like age in years, then analysis will use this scale. Otherwise, variables will be transformed to have a mean of zero and standard deviation of one—then, $\hat{\delta}$ is more readily interpreted as the effect of the program at the average values of X_i^d , and $\hat{\gamma}_d$ shows the intensification of the effect for a one-standard deviation change in X_i^d .

3.7 Testing Pathways in the Theory of Change

Experimental contrasts in PACE were designed to estimate each program’s overall impacts. Nonexperimental methods are necessary in tracing the causal pathways through which any overall impacts may have occurred. This section describes plans for exploratory analyses of the degree to which experimental impacts are mediated by key program service components and intermediate outcomes in program logic models. It briefly describes the planned approach and then explain its underlying concepts, the assumptions on which the approach is based, and the analytic opportunities and challenges.

Analyses of mediation will be limited to programs with statistically significant impacts on primary outcomes. These analyses will estimate a variety of pathways from measured services to more proximate and distal outcomes as summarized in the general career pathways theory of change. Hypothesized mediators—represented as systems of equations—will be specified to capture testable features of the program’s logic model. For two reasons, it will not be possible to test the full range of potentially influential program features and mediators. First, the analysis leverages variation in exposure to services at the individual level, and effects cannot be estimated for program features to which all treatment group members are exposed. For example, all Year Up participants take the same courses and are subject to the same rules concerning professional behavior. Second, notwithstanding rich measures of intervening outcomes in PACE follow-up surveys (e.g., psycho-social factors, sense of career development, material resources), some domains cannot be measured well through self-reports. Most notably, valid measures of developing skills require task-based assessment.⁵⁵

The first step will be to delineate, for each PACE program, hypothesized mediators that can and cannot be measured through either survey or program administrative data.⁵⁶ Measureable mediators will be highlighted within full logic models to identify subsets of pathways and pathways segments that can be observed. The resulting set of potentially testable hypotheses for each program is likely to be substantial, given the relatively comprehensive nature of PACE programs and rich available data.

The first follow-up survey measures an extensive array of specific features of instruction, supports, and employment experiences recognized as “signature career pathways strategies” and to which participants may experience varying degrees of exposure. It establishes for instruction, for example, the extent of participation by program type and provider, as well as appraisals of instructional approaches (e.g., hands-on learning, contextualization). The wide range of supports measured includes varied forms of personal assistance (e.g., academic guidance, tutoring, help securing supports), support services,

⁵⁵ Chapter 4 discusses assessment tools that could be used to remedy these gaps. Decisions on whether to collect these additional data will be made in the near future.

⁵⁶ Although program administrative data mostly provide measures of differential exposure to intervention features only for treatment group members, it may be possible to determine that virtually no control group members were subject to some features and thus impute zero exposure to all control group members.

financial assistance, and instruction in general skills needed to succeed at school and work. The survey obtains measures also for exposure to work-based learning through internships and project-based assignments.

Two steps will be taken to identify the most parsimonious possible sets of pathways. First, as in the main experimental impact analyses, where possible multiple indicators will be summarized as indices for high-level constructs corresponding to principal domains in the career pathways theory of change. Second, experimental findings will serve to refine hypothesized pathways. Analyses generally will not include services or intermediate outcomes for which there are no statistically significant treatment-control differences.⁵⁷

Although the equations the PACE team will use for analysis of mediation methods are complex, the basic concepts and the assumptions required to justify them are not. The section below describes the concepts the team will employ, and the assumptions on which they depend, using examples from PACE programs. Subsections use a programmatic example that moves from the simplest version of the analysis method to the most complex method.

One Mediator in a Linear System⁵⁸

Many PACE programs offer intensive and frequent academic advising intended to improve nontraditional students' completion of credentials by imparting a variety of skills including individual and group study skills, strengthening academic self-confidence, and time management.⁵⁹ If a PACE program offers intensive academic advising and has a positive effect on completion of college credentials, two questions arise: did the advising contribute to the positive outcome, and if so, to what extent did it contribute? Two equations characterize the approach to answering these questions.

Let (T_i, M_i, Y_i) represent treatment status (treatment versus control group member), the value of the mediator (e.g., how much career counseling an individual received), and the main outcome (e.g., completion of a degree). The indirect effect of treatment on Y mediated by M , the contribution of the

⁵⁷ The analysis will consider two kinds of exceptions to this approach. First, where subgroup differences in impacts occur for a mediator for which impacts are not statistically significant overall, the mediator may be retained, along with corresponding interaction terms. Second, mediators may be retained where theory or prior evidence indicates that likely impacts may have been “suppressed” by impacts on a countervailing factor that also can be included in the analysis. For example, a program that increases perceived social support by creating a strong “learning community” might reduce utilization of counseling for help with self-esteem issues.

⁵⁸ For ease of explanation, for the most part in this section discusses linear models, but in the actual analysis the team will use more general approaches that do not require an assumption of linearity. This is discussed briefly at the end of this section and in greater detail in Appendix A.

⁵⁹ The text begins with an example of a program component for ease of explanation, but analytically could have begun with an intermediate outcome.

mediator to the outcome, is often estimated as the product of the effect of T on M with the effect of M on Y . The standard system of equations from Baron and Kenny (1986) is:

$$\begin{aligned}M_i &= \alpha_1 + aT_i + e_{1i} \\ Y_i &= \alpha_2 + dT_i + bM_i + e_{2i}\end{aligned}\tag{3.4}$$

where it is assumed that the errors are normally and independently distributed across individuals with mean zero and constant variance, and that e_{1i} and e_{2i} are furthermore independent of each other. Here, a represents the intervention's effect on M and b further indicates how increases in M —whether treatment-induced (through T) or by other channels—increase Y . The core idea is that if the team can estimate how much a unit increase in M increases Y , as well as how much T increased M , then multiplying one by the other will provide the average effect across all treatment group members. The direct influence of treatment on this outcome (e.g., of T on Y) is reflected in d . Under the stringent conditions regarding error terms noted above, the two equations in 3.4 can be estimated with ordinary least squares and the indirect effect of treatment on Y mediated by M estimated as $\hat{a}\hat{b}$. Also, \hat{d} is taken to be the direct effect of T (i.e., the effect of T on Y that is not mediated by M).

If M (or Y) is an unconditional outcome, i.e., where its having a value does not depend on some post-random assignment status and thus it holds for all treatment and control group members, the effect of T on M (or Y) can be estimated experimentally.⁶⁰ However, since, unlike T , M is not randomly assigned, its effect on Y cannot be estimated experimentally. This lack of random assignment makes it is easy to imagine many ways in which the assumptions of this simple model will be violated. If those who receive more advising are more disadvantaged academically than other students, then the two errors will be negatively correlated. In this illustration, prior academic disadvantage is a common cause of both the mediator (intensity of advising) and the main outcome (completion). Another way to think about what it means for the errors to be independent is that there are no common causes of the two variables. Since in this illustration academic disadvantage is a common cause of both the mediator and the main outcome, the two error terms are not independent and, therefore, a critical assumption of this model is violated. Dependence of the two errors is commonly referred to as either confounding or selection bias.⁶¹

⁶⁰ Potential mediators may not always be measurable for the full sample, for example, if having a value depends on being in postsecondary education and training and not all sample members are (see Section 3.8.3 for a fuller discussion). In such a case, the team will drop sample members who are not trainees, but this has the consequence that estimates of impact on both M and Y will be nonexperimental and thus require stronger assumptions.

⁶¹ Note that the model contains no interaction term for T and M , so there is an implicit assumption that the career counseling affects completion the same whether other components of the intervention are in place or not. Theoretically, interaction terms could be added but findings would be difficult to present and sample sizes likely would be inadequate.

In this example, adding a baseline variable that fully captures academic disadvantage insofar as it causes both higher levels of advising usage and lower completion rates overcomes the problem of selection bias. Thus the team adds a vector of p baseline covariates, X_i to the model:

$$\begin{aligned}M_i &= \alpha_1 + X_i\beta_1 + aT_i + e_{1i} \\Y_i &= \alpha_2 + X_i\beta_2 + dT_i + bM_i + e_{2i}\end{aligned}\tag{3.5}$$

More generally, the intent is to try to include **all** exogenous common causes of the mediator and the main outcome. Adding these covariates into both equations makes it more plausible that the errors have constant means of zero, constant variance, and that they are independent of each other. However, this is still a very strong assumption and is essentially the same one made in a nonexperimental evaluation with a comparison group when it is assumed that there are no omitted covariates that influence both the outcome and selection into treatment; if this is violated the estimates will contain bias.⁶²

Parallel Multiple Mediators in a Linear System

As noted earlier, career pathways program designs are based on the assumption that successfully improving main and intermediate outcomes requires multiple program components that do not operate independently (i.e., assessment, training, services, employment connections). If this assumption is true, then the analysis of a single mediator in isolation is not likely to yield an accurate estimate of its contribution to the main outcome. To continue the example, in addition to more intensive advising, a number of PACE programs provide various forms of financial support to reduce participants' need to work while they are in the program and thus allow them to focus on education and training. In this case, one can imagine how the availability of greater financial support could cause treatment group members both to attend more advising sessions and to complete a credential. This would create a post-random assignment confound similar to the pre-random assignment confound that required adding covariates with the single mediator. So in addition to the need to add covariates, in a multi-dimensional intervention such as a career pathways program, the team cannot be satisfied with a model with a single mediator.⁶³

⁶² It is notable that although there are a rich set of baseline characteristics, they were not chosen to predict which individuals would need particular services.

⁶³ Fitting a series of single mediator models is sometimes done, but for the reason given above, the team does not adopt it. See, for example, page 511 of Bullock and Ha (2011). "This practice [of considering mediators one at a time] makes biased inferences about mediation even more likely. The researcher, who already faces the spectre of bias due to the omission of variables over which she has no control, compounds the problem by intentionally omitting variables that are likely to be important confounds."

Let $(T_i, X_i, M_{1i}, \dots, M_{qi}, Y_i)$ represent a richer vector of student-level data with p baseline covariates and q parallel mediators. In this type of model, all the mediators are exclusively program components.⁶⁴ This assumption makes the assumption of parallel mediation more plausible. Preacher and Hayes (2008) indicate how to fit a system of $q+1$ equations for this vector:

$$\begin{aligned} M_{1i} &= \alpha_1 + X_i\beta_1 + a_1T_i + e_{1i} \\ &\vdots \\ M_{qi} &= \alpha_q + X_i\beta_q + a_qT_i + e_{qi} \\ Y_i &= \alpha_{q+1} + X_i\beta_{q+1} + dT_i + b_1M_{1i} + \dots + b_qM_{qi} + e_{q+1,i} \end{aligned} \quad (3.6)$$

In this model, all the errors must have zero means and constant variance (although each can have a different constant variance), all errors must be independent across individuals, all errors must be independent across equations,⁶⁵ and the linear specification of the M s' influences on Y must be correct. If these conditions are met, then the indirect effect of any particular mediator M_{ri} is estimated to be $\hat{a}_r\hat{b}_r$. They can be compared to each other to see which is largest. Formal testing can also be done on contrasts of the indirect effects.

Serial and Parallel Multiple Mediators in a Linear System

As illustrated in Exhibit 1.2, the components of PACE programs are expected to lead to improved main outcomes through specific intermediate outcomes. Thus, in the example, more intensive advising is not expected to affect completion directly, but rather through an intermediate outcome such as greater academic self-confidence. Similarly, financial supports in and of themselves do not directly affect completion but may have a positive effect on completion through greater time for study or reduced financial stress. This suggests the need to consider parallel mediators as well as serial ones.

This pattern of serial and parallel multiple mediation can be formalized as follows. Let

$(T_i, X_i, D_{1i}, \dots, D_{qi}, W_{1i}, \dots, W_{ri}, Y_i)$ represent a richer vector of individual-level data with p baseline covariates (X), q parallel program components (D) and r parallel intermediate outcomes (W). The program components cause the intermediate outcomes, so the team estimates serial as well as parallel mediation. To utilize this data, the team needs to fit a system of $q+r+1$ equations that involves:

⁶⁴ All the mediators could have been exclusively intermediate outcomes.

⁶⁵ For this independence assumption to be met, there must be no omitted common causes of any of the mediators and the outcome. These common causes could be either baseline characteristics or post-randomization phenomena. This is not quite as strong an assumption as saying that there must be no omitted mediators, but it is close to being this demanding since, by definition, any mediator is a cause of the main outcome. So the only mediators that could be omitted without violating model assumptions are those that are independent of all the included mediators.

- Regressions of each program component on treatment status and covariates,
- Regressions of each intermediate outcome on treatment status, covariates, and all of the program components, and
- A regression of the main outcome on all the other variables.

Fitting such a system would provide estimates of the indirect effects of specific services on specific intermediate outcomes, the indirect effects of specific intermediate outcomes on the main outcome, and doubly indirect effects⁶⁶ of specific services through specific intermediate outcomes on the main outcome. Stringent assumptions must be made about the errors as in the simpler models, but they become more plausible with this level of complexity.

Serial and Parallel Multiple Mediators with Simultaneous Moderation and Moderated Mediation in a Linear System

Effects of treatment are likely to vary across subgroups defined in terms of baseline characteristics (as discussed in section 3.6). For example, training through a career pathways program may be more helpful to women than to men. Furthermore, the various indirect effects are also likely to vary across subgroups, for example, those with a high baseline score on academic self-confidence may be unlikely to benefit much from program components aimed at that intermediate outcome. Following Baron and Kenny (1986), the first of these added complications is referred to as moderation and the second is referred to as moderated mediation. If different subgroups both benefit differentially from the basic intervention and seek out services at different rates, it can be easy to confuse moderation and mediation. Thus models are needed that simultaneously allow moderation, mediation, and moderated mediation.

To meet these demands, the models fitted are likely to contain one or two dozen baseline covariates, less than five program components, a handful of intermediate outcomes, and a single main outcome. To keep the presentation simple, an illustrative equation for a single binary baseline covariate, two program components, two intermediate outcomes, and a single main outcome has the form:

$$\begin{aligned}
 D_{1i} &= \alpha_1 + X_i\beta_1 + a_1T_i + X_iT_ic_1 + e_{1i} \\
 D_{2i} &= \alpha_2 + X_i\beta_2 + a_2T_i + X_iT_ic_2 + e_{2i} \\
 W_{1i} &= \alpha_3 + X_i\beta_3 + a_3T_i + X_iT_ic_3 + D_{1i}f_{11} + D_{2i}f_{12} + X_iD_{1i}g_{11} + X_iD_{2i}g_{12} + e_{3i} \\
 W_{2i} &= \alpha_4 + X_i\beta_4 + a_4T_i + X_iT_ic_4 + D_{1i}f_{21} + D_{2i}f_{22} + X_iD_{1i}g_{21} + X_iD_{2i}g_{22} + e_{4i} \\
 Y_i &= \alpha_5 + X_i\beta_5 + a_5T_i + X_iT_ic_5 + D_{1i}f_{31} + D_{2i}f_{32} + X_iD_{1i}g_{31} + X_iD_{2i}g_{32} + \\
 &\quad + W_{1i}h_1 + W_{2i}h_2 + X_iW_{1i}k_1 + X_iW_{2i}k_2 + e_{5,i}
 \end{aligned} \tag{3.7}$$

⁶⁶ “Doubly indirect” means the relationship between A and D when A causes B causes C causes D.

In this model, the α coefficients give the direct effect of T (the local PACE program) for those with baseline covariate $X=0$, the c coefficients give the direct effect of T for those with $X=1$, the f coefficients give the direct effects of program components (on both intermediate outcomes and the main outcome) for those with $X=0$, and the g coefficients give the direct effects of program components for those with $X=1$. The indirect effect of T via D_1 on intermediate outcome W_1 for those with $X=0$ is then estimated as $\hat{a}_1\hat{f}_{11}$ while the indirect effect of the PACE program via D_1 on intermediate outcome W_1 for those with $X=1$ is then estimated as $\hat{a}_1\hat{g}_{11}$. Similar terms are used to estimate the indirect effects of T via D_1 on intermediate outcome W_2 and via program component D_2 on intermediate outcomes W_1 and W_2 .

The indirect effects of treatment on the main outcome in this model are slightly more complex. The indirect effect of treatment on the main outcome Y via program component D_1 is estimated as $\hat{a}_1(\hat{f}_{31} + \hat{f}_{11}\hat{h}_1 + \hat{f}_{21}\hat{h}_2)$ for those with $X=0$ and by $\hat{c}_1(\hat{g}_{31} + \hat{g}_{11}\hat{k}_1 + \hat{g}_{21}\hat{k}_2)$ for those with $X=1$. Similarly, the indirect effect of treatment on main outcome Y via program component D_2 is then estimated as $\hat{a}_2(\hat{f}_{32} + \hat{f}_{12}\hat{h}_1 + \hat{f}_{22}\hat{h}_2)$ for those with $X=0$ and by $\hat{c}_2(\hat{g}_{32} + \hat{g}_{12}\hat{k}_1 + \hat{g}_{22}\hat{k}_2)$ for those with $X=1$.

Examining the indirect effect of T on Y via D_1 more closely indicates that it is composed of doubly indirect effects via the intermediate outcomes W_1 and W_2 estimated by the terms $\hat{a}_1\hat{f}_{11}\hat{h}_1$ and $\hat{a}_1\hat{f}_{21}\hat{h}_2$ as well as a residual indirect effect $\hat{a}_1\hat{f}_{31}$ that is not mediated by either of the intermediate outcomes.

The indirect effects of T via intermediate outcomes are also slightly more complex. The indirect of treatment on the main outcome Y via intermediate outcome W_1 is estimated as $\hat{h}_1(\hat{a}_3 + \hat{a}_1\hat{f}_{11} + \hat{a}_2\hat{f}_{21})$ for those with $X=0$ and as $\hat{k}_1(\hat{c}_3 + \hat{c}_1\hat{g}_{11} + \hat{c}_2\hat{g}_{21})$ for those with $X=1$. Similarly, the indirect of treatment on the main outcome Y via intermediate outcome W_2 is estimated as $\hat{h}_2(\hat{a}_4 + \hat{a}_1\hat{f}_{12} + \hat{a}_2\hat{f}_{22})$ for those with $X=0$ and as $\hat{k}_2(\hat{c}_4 + \hat{c}_1\hat{g}_{12} + \hat{c}_2\hat{g}_{22})$ for those with $X=1$. Examining the indirect effect of T on Y via W_1 more closely, we see that it is composed of doubly indirect effects via the program components D_1 and D_2 estimated by the terms $\hat{a}_1\hat{f}_{11}\hat{h}_1$ and $\hat{a}_2\hat{f}_{21}\hat{h}_2$ as well as a residual indirect effect $\hat{a}_3\hat{h}_1$ that is not mediated by either of the program components.

These estimates will be unbiased if the five error terms in (3.7) are mutually independent, have constant variance, have zero means, and are independent across individuals.

Generalizations for Nonlinear Relationships

As mentioned above, although the approach described above is useful in illuminating our approach, it will not work in all cases, for example, where a potential mediator is binary and not a continuous variable with a normal distribution. To address this will require adopting a statistical framework based

on recently developed methods involving generalized structural equation modeling (Pearl, 2001, 2012a, 2012b; Imai, Keele and Yamamoto, 2010; Imai, Keele, and Tingley, 2010). Estimation involves fitting models (nonlinear or linear) to each of a set of mediators and outcomes in specialized path diagrams known as “directed acyclic graphs.”⁶⁷

3.8 Other Issues for the Impact Study

Finally, this subsection briefly identifies and discusses approaches to several other technical issues arising in the impact analysis.

3.8.1 Multiple Sites in the Year Up and I-BEST Programs

The Year Up program will be conducting the experiment across eight sites, with treatment samples in the 100–300 range per site. Similarly, the I-BEST program is conducting the experiment across three technical and community colleges within Washington State. The team plans to pool both the eight Year Up sites and the three I-BEST sites for the program-specific reports. Nonetheless, for both Year Up and even more so for I-BEST, there are likely to be important differences in local implementation and context whose effects on program impacts would be valuable to understand. Year Up staff in particular are very interested in program-specific impact estimates, and the team plans to provide (but not publish) such impact estimates with due caution about small sample sizes and associated uncertainty. However, for neither program is the number of sites sufficient to support statistical analysis of local factors associated with observed variance in program impacts. Expressed very briefly, in social programs (as opposed to the delivery of pharmaceuticals or medical devices), there is typically much greater variation in implementation, and it is usually impossible to separate the effects of, for example, different personnel from the effects of manipulable intervention features. With just three or even eight sites, the team cannot separate the mediating effects of manipulable intervention features from the mediating effects of local staffing.

The team can, however, use the variability of estimated effects across the sites to get better inferences about the range of effects that might be achieved if either program were to expand into new sites, and will do this in addition to making an inference about the average effect of each program in its existing sites. The inference about the effect of either program in a new site will have a wider error band than the effect within the existing offices. The team will estimate the standard errors for the wider error bands using multi-level modeling techniques. The analysis with narrower error bands will be more appropriate for decision making about funding of new cohorts in the existing sites while the analysis with the wider error bands will be more appropriate for decision making about funding for new sites.

3.8.2 Provide Estimates for Treatment on Treated (TOT) Impacts

The discussion thus focused on “intent to treat,” or ITT impacts of the intervention. An ITT impact shows the average effect of *access* to training, even in instances where the treatment group includes

⁶⁷ See Appendix A for an illustrative acyclic graph and a detailed description of the method.

individuals who did not show up for training after being admitted. Typically, intervention designers are very interested in the average effect of *receiving* training—the effect of “treatment on the treated,” or TOT. Estimates of such effects involves making assumptions about the impact of the program on the “no-shows,” people assigned to treatment who fail to participate in the program being evaluated. TOT estimates are typically larger than the effects of merely being given access to training.⁶⁸ ITT estimates and TOT estimates have different audiences to some extent. Typically, the ITT estimate is of most interest to the policy analyst since programs cannot generally force participation by persons and often cannot quickly refill unclaimed classroom seats. In contrast, TOT is the most relevant estimate to individuals considering entrance into the program since they assume they can control their own participation after acceptance into the program. However, depending on the costs of no-shows and the sensitivity of the analyst to program costs, the TOT may be more interesting to the policy analyst as well.⁶⁹ In addition, TOT may be useful to future meta-analysts, as well as being helpful to those readers who might compare PACE’s findings to those from observational studies of similar interventions for which only TOT estimates were published.

The conventional approach to estimating the TOT effect is to rescale the ITT estimate—i.e., the overall treatment-control group difference in outcomes for the entire experimental sample—to reflect just those cases that received program services. This methodology was developed by Bloom (1984) and assumes that program group members who do not participate experience no impact and that there is no “crossover” in which control group members receive the same services as those in the treatment group. Computationally, the TOT estimator can be computed as the ITT impact estimate divided by $1-RN$, where RN is the PACE training nonparticipation rate in the treatment group (i.e., the “no-show” rate).⁷⁰ Division by this factor, which is always less than 1, increases the absolute value of the estimate of impact (and its standard error); i.e., the measured effect of participating compared to not participating (TOT) is larger on average than the measured effect of being granted access to the training through random assignment—in situations where some who are granted access do not participate. The same statistical test for significantly positive (or negative) impact applies to the TOT estimate as to the ITT estimate, because no additional hypothesis is tested in preparing TOT estimates.

⁶⁸ The only way that measured effects get smaller with adjustment for no-shows is if the ITT estimates of program effects are negative. In that case, adjusting for no-shows will result in estimated effects that are still negative but are now larger in absolute value.

⁶⁹ Consider two programs that would each cost \$300 million per year to implement nationally that both produce an average benefit to admitted persons of an extra \$1000 in quarterly earnings at the end of 8 quarters, but that program B can graduate twice as many trainees because half the persons admitted to program B never show up and the program refills their seats at no additional cost while all those assigned to program A show up. Then the TOT for program B is \$2000, twice as large as the \$1000 for program A, while both produce the same number of graduates per year. In practice, there are usually substantial costs for no-shows, both in terms of empty seats and in terms of the screening process for program admission.

⁷⁰ See Bloom (2006) for a discussion of this formula (which he characterizes as division by the impact of random assignment on intervention receipt).

This approach could also be expanded to adjust for contemporaneous crossover defined as participation in the treatment by members of the control group. Angrist, Imbens, and Rubin (1996) generalized Bloom's approach to handle contemporaneous crossover as well as no-shows.⁷¹ However, this approach does not handle lagged crossover.⁷² Procedures are in place with all PACE programs that will strongly discourage contemporaneous crossover, but in at least one site some lagged crossover is likely.⁷³ Just as adjusting for no-shows typically increases the size of measured effects, adjusting for contemporaneous or lagged crossover will also typically increase the size of measured effects. Because crossover could not occur if the experiment were not being conducted, there are strong reasons for adjusting for it. If contemporaneous crossover occurs to a sizeable extent despite the plans to minimize it, the team will consider the Angrist, Imbens, and Rubin approach for adjusting for it, as well as other approaches such as imputing counterfactual outcomes for crossover cases based on experiences of similar members of the control sample who did not crossover. With respect to lagged crossover, Bell and Bradley (2013) recently published a new methodology, which the team will consider applying if the need arises.

3.8.3 Conditionally Observed Outcomes

Many outcomes will only be observed for a subset of the study sample.⁷⁴ For example, details of course-taking will only be observed among those who attend school, and details of job benefits will only be observed among those who obtain jobs. For many of these variables, we will be able to assign a logical value to those not asked the question, such as assigning zero earnings to those who fail to obtain jobs. There may be a few outcomes where this approach is not very suitable. One example might be job satisfaction. Should not having a job be ranked below the lowest level of job satisfaction among those who are employed?

Because these subsets are endogenous to treatment (e.g., who gets a job or goes to school can be affected by the treatment), ignoring the conditionality of outcomes when evaluating social experiments can give rise to biased estimates and is thus not good practice. For example, it seems plausible that some who obtain jobs because of training provided by a career pathways program may have lower-rung jobs with less desirable conditions of employment than is typical among people who can obtain employment without the training. If so, estimates of program effects on conditional outcomes that ignore conditionality will be more negative than warranted. Although the team is unaware of any strong evidence for this proposition in the field of job training, this sort of phenomenon is very apparent in

⁷¹ Often referred to in the literature as the AIR approach.

⁷² Appendix B of Abt Associates et al. (2006) contains a procedure for correction for lagged crossover. The team has not studied whether it would make sense to try to adapt it for PACE.

⁷³ The VIDA program has only a two-year embargo on program services for control group members.

⁷⁴ In fact, all survey outcomes will only be measured among study participants who are still alive. To date, the team is aware of one death in the sample. However, mortality is expected to be low in this population, thus the team will ignore deceased study participants when studying survey outcomes. For the most important long-term outcomes of employment and earnings, matching to the NDNH will correctly identify deceased participants as having zero earnings.

other areas of education and training such as college entrance exam scores, for which it has been demonstrated that average scores are inversely related to participation rates across states.⁷⁵

The team will use two approaches to conduct analyses of conditional outcomes. One approach is to create composite outcomes based on the conditional outcomes, such that the composite outcomes are unconditional. For example, to estimate effects on having employment in a target occupation, the team can create two outcome categories—employed in target occupation versus either employed in an off-target occupation or unemployed—and then test for significance.⁷⁶ The advantages of this approach are that, first, it answers an important question, “Does having access to the treatment increase the proportion of individuals employed in the target occupation?” and, second, the estimate is fully experimental, applying to all treatment and control group members and thus not subject to bias.

The primary disadvantage of this approach is that it forces somewhat arbitrary decisions in order to create the composites (or at least decisions which could reasonably be made in a variety of ways). For example, if the team is interested in estimating effects on hourly wage rates, the needed categories are not as obvious as in the example of having employment in a target occupation. Thus, the team could use the following scheme for composite outcomes—either not employed or earning up to X dollars per hour, and earning over X dollars per hour—but more than one threshold could be justified, and different thresholds could produce different results. A second shortcoming with this approach is that it does not answer the question of whether the career pathways program improves wage rates for those who would have worked even without having access to the treatment. Rather than making an arbitrary decision in this circumstance, or perhaps in addition to it, it may be desirable to estimate the impacts for the group of participants who would have been employed with or without the intervention.

This group of participants is one of the “principal strata” defined by Frangakis and Rubin (2002) in their approach to the analysis of experimental outcomes that are observed conditionally on a post-randomization event. Translating their arguments to this setting, it does not make sense to discuss the effect of treatment on the hourly wage rates or job satisfaction of those persons whose employment status depends on their treatment randomization status. Rather, the focus should be on those who would work either way. Since, however, this group is not directly observed, additional assumptions must be made in order to complete the desired analysis. There are various approaches to doing this. None of them are very satisfying. One approach (McConnell, Stuart, and Devaney, 2008) imputes which of the treatment cases would have been employed even if they had been assigned to the control group. If this can be done well, then hourly wage rates or job satisfaction levels can be directly compared between employed persons in the control group and those members of the treatment sample who are imputed to be in the always-employed group. The question of whether it can be done well depends on the quality and relevance of the baseline data collected.

⁷⁵ See, for example, Clark, Rothstein, and Schanzenbach, 2008.

⁷⁶ The team could also create three categories by splitting the latter two, but a significant effect on the category of employed without health benefits would not be readily interpretable as positive or negative.

The decision of whether to attempt this second form of estimation is important for PACE given that the team plans to include many questions in the extended follow-up interviews (starting at 15, 36 and 60 months, if authorized) about the nature of training undertaken and employment obtained. This approach necessitates high quality models of who would have undertaken training or been employed at those times whether or not they were selected for the treatment groups.

3.8.4 Missing Data

The team anticipates encountering a complex array of missing data. There will be individuals lost to survey follow-up, individuals who refuse some questionnaire items or supply “don’t know” responses (at baseline and at each of the three planned follow-ups), and individuals with missing administrative data of various sorts. For the second follow-up reports, there will also be research subjects who will have been successfully interviewed after missing the first follow-up survey. As suggested by Puma et al. (2009), the strategy in this area will involve some mix of imputation, weighting, addition of missing value indicators to covariate lists, and case dropping. The considerations of which method to use for which variables under which conditions depends to some extent on whether the missing variables are baseline covariates, program inputs, intermediate outcomes, or main outcomes. The simplest procedure is always to simply drop cases from an analysis that have missing values on any of the variables involved in the analysis. However, the procedure can be costly in terms of lost information if the number of variables involved in an analysis is large. Bias is also a concern with simple case dropping. The other alternatives all salvage more partial information, thereby reducing variances and often also reducing nonresponse bias. Generally speaking, strategies that salvage more partial information appropriately are more complex and costly to implement.

One strategy that is often used as a compromise in experimental longitudinal studies such as PACE is to use weighting to compensate for unit nonresponse at the most current follow-up, adding missing value indicators for scattered missingness on baseline covariates used in outcome models, and case dropping for scattered missingness on outcomes. As argued by Puma et al. (2009), missing value indicators for covariates in randomized experiments are attractive because control on covariates is not required to obtain unbiased program effects. Instead, restricting the analysis to the set of cases with nonmissing covariates can increase bias. By adding the missing value indicators, no cases are dropped on account of missing covariates. This is a reasonable approach but the team aims to do better.

The main lever for developing better solutions than the default suggestion is the administrative data that are discussed in Chapter 4. This chapter has primarily used examples that rely on research subject-supplied information. However, the PACE team is collecting a rich set of administrative data. Many of these data will only be available on the treatment sample and are thus not helpful for estimating impacts. In addition, the team will have data on employment and quarterly earnings from the National Directory of New Hires (NDNH), college attendance and graduation data from the National Student Clearinghouse (NSC) and, possibly, more detailed college data from local college systems and/or state

data systems for some programs. Based on unpublished work by Judkins⁷⁷ on an evaluation of a college access program funded by federal grants to local partnerships, the team anticipates that matching to NSC data will indicate that students lost to follow-up are also less likely to be attending college. Along the same lines, it seems likely that those students who fail to find employment may be more difficult to track and interview since they may lack stable housing. When the outcomes of interest (e.g., college completion and employment) affect the likelihood of response, this is a form of nonignorable nonresponse, something that cannot be corrected using baseline data alone. Moreover, to the extent that the local career pathways program is successful at boosting college completion and subsequent employment and earnings, the biasing effect of the nonignorable missingness will be different across the study arms and tend toward inappropriate diminishment of estimated program effects.

However, when information is available on the outcomes from sources not subject to nonresponse (although they may be subject to other errors such as coverage errors due to under-the-table employment or attendance at a school that does not cooperate with the NSC), the relationships between causally prior variables (baseline variables, program inputs, and intermediate outcomes) on one side and primary outcomes on the other side can be “reverse-exploited” to make good inferences about the causally prior variables from the primary outcomes. Along such lines, Judkins developed imputation procedures that he applied to a combination of survey and administrative data to impute entire missing questionnaires in a way that was consistent with both the available survey data and the available administrative data for the evaluation of a college access program.⁷⁸ He reports that the impact of this imputation on college attendance levels was remarkably strong.⁷⁹

While the team does not have access to the imputation software previously used for this purpose by Judkins, SAS version 9.3 has a new imputation option under PROC MI that contains many of the same features (the “FCS” option). The SAS software is limited in the number of variables that can be simultaneously imputed in such a way that preserves relationships among the variables. So although whole-questionnaire imputation with it is impossible, the plan is to use PROC MI to run a custom imputation on the core set of variables for a report. This will include the unified set of covariates identified with the methodology discussed in Section 3.4.3, primary and secondary outcomes for the report, core moderators, and relevant administrative data.⁸⁰ This imputation will be restricted to cases with partially completed interviews. For cases without surveys at all we will use weighting nonresponse adjustments, including the available post-randomization administrative data in the construction of the

⁷⁷ Work was conducted during his time at Westat.

⁷⁸ Judkins, et al., 2007; Krenzke and Judkins, 2008.

⁷⁹ This research was never published due to project cancellation.

⁸⁰ In order to use the NDNH data to improve the quality of nonresponse adjustment weighting and item nonresponse imputation for the survey data, it will be necessary to include the survey data in a “pass-through” file that is submitted to the Office of Child Support Enforcement (OCSE), the federal office that houses and governs the use of the NDNH. OCSE will return the de-identified file to ACF with the substantive variables present on the pass-through file as well as earnings data.

weights if it turns out that weights based on baseline variables by themselves do not yield unbiased estimates of outcomes based on administrative data. This combination of weighting and multiple imputation will support all confirmatory and secondary analyses. When conducting exploratory analyses, the team will continue to use the nonresponse-adjusted weights but will simply drop cases missing any variables not imputed to support core analyses (such as mediators and tertiary outcomes).

3.8.5 Comparison of Findings from Mediation Analysis across Programs

After completion of the second follow-up reports, the team suggests a special study that provides a cross-site analysis of findings from mediational analyses for individual sites described earlier in this chapter. The underlying aim of the site-level mediational analyses is to identify program inputs and causal pathways operating to produce desired impacts for participants with particular characteristics. Although the aim in program-specific analyses is to discover operative mechanisms consistent with each program's logic model, the broader underlying goal is to identify mechanisms with some generality. Inasmuch as the aim in program-level analyses will be to specify concepts, measures, and statistical models consistent with wider theory, the degree to which certain results occur for multiple programs will help to establish their credibility more firmly. Thus, exploratory cross-site analyses in PACE will focus on the extent to which moderating effects of baseline study participant characteristics and mediating effects of certain program inputs and intermediate outcomes are consistent across programs.

In its simplest form, these analyses will involve testing the equivalence of estimated parameters from similar mediational models in two or more sites. This work necessarily will follow the development of analysis plans for individual sites. The analysis plan will identify specific sites and hypothesized mediational pathways for which cross-site tests would be informative. Because the plan will be based on site-specific analyses, it will be prepared after plans for the latter are drafted.

3.9 Cost-Benefit Study

The cost-benefit analysis is designed to provide a full accounting of the various consequences of the PACE interventions from multiple points of view. The team will lay the groundwork for conducting this assessment for each program, as well as for different subgroups of the target population within individual programs. Although the team will collect cost data during the PACE study period, the full cost-benefit study cannot appear sooner than the third follow-up reports.

3.9.1 General Approach

The primary objective of the cost-benefit analysis is to measure all the tangible impacts of the interventions and to assign appropriate dollar values to these impacts (per unit of impact) from the perspectives of participants; federal, state and local governments; and the rest of society—that is, individuals others than participants and non-government institutions (notably and foundations). The sum of dollar values across these perspectives constitutes the net value to society as a whole. Positive dollar values are considered benefits and negative values are costs. The analysis involves estimating *net* benefits and *net* costs—that is, benefits and costs with the PACE interventions relative to what would have happened without the interventions. Indeed, the estimation techniques are nearly identical to the methods described above for the impact analysis. Although the costs are incurred in the short term,

benefits may accrue over the full lifetimes of participants. However, we will only examine projected benefits over a 10-year timespan.

3.9.2 Benefits, Costs and Perspectives

Exhibit 3.3 summarizes the analytical framework for the PACE cost-benefit analysis. The columns of the exhibit correspond to the perspectives for valuing benefits and costs of the interventions, and the rows list the different benefit and cost components included in the analysis. The cells of the exhibit indicate whether a particular impact is expected to be a benefit (+) or a cost (-) for that perspective, or whether it is expected to have no effect (0). The final column at the right, representing all of society, shows whether each item is a benefit to society (+), a cost to society (-), or a transfer between two groups (perspectives) within society (0) that produces no net benefit or cost for society as a whole. A “bottom line,” the net value shown at the foot of each column, is the sum of all individual benefits costs in that column.

Each “+” and “-” in the table amounts to a hypothesis that the evaluation will test. Many of the expected benefits are the same effects on trainees that the impact analysis is measuring, such as earnings increases that benefit both trainees and society (the latter is based on earnings, as described below). There is no corresponding cost to society because the expense of earnings and fringe benefits incurred by employers is offset by the value of the output (the marginal product) produced by the workers who receive the compensation. In addition, the expected increase in earnings will result in taxes that could be a loss to trainees⁸¹ and a gain to governments (no net effect on society). Other expected effects on trainees are transfers that benefit federal, state and local governments, including lower TANF and Medicaid⁸² payments, if the interventions are successful in moving recipients off welfare and/or preventing trainees from needing to use welfare.

Most of the costs in the exhibit constitute outlays by TANF, the Workforce Investment Fund (WIA Titles I and II), Pell Grants, community colleges (which have multiple funding streams), and foundations that have contributed funds for PACE program services, including the Open Society Foundations, the Joyce Foundation, and the Kresge Foundation. Most of these costs are associated with services available to both treatment and control group members, with a higher expected level of resource use expected for treatment group members constituting a net cost to specific federal programs such as TANF and WIA, the federal government, state and local government, and society as a whole. However, some the costs are for PACE-specific supports provided only to treatment group members. Some of the government

⁸¹ The net effect on income taxes paid by trainees is uncertain, particularly because of the Earned Income Tax Credit (EITC), which pertains to the federal income tax and most state income taxes. Some increases in earnings would reduce income taxes, because they will increase trainee eligibility for this credit. However, other increases would increase taxes by reducing the credit or eliminating eligibility for it.

⁸² In this design report, Medicaid is identified as the primary source of government health insurance for low-income individuals. As the Affordable Care Act is implemented further, the team will revise to measure a broader range of government subsidies.

costs and all of the costs paid for by foundations that support PACE programs fall into this category. In addition, the expected increase in employment should add expenses to participants' budgets (e.g., transportation) that constitute social costs as well.

Exhibit 3.3: Expected Benefits and Costs of PACE Interventions, by Accounting Perspective

Component of Analysis	Participants	Government		Rest of Society	Society
		Federal	State/Local		
Earnings and fringe benefits	+	0	0	+	+
TANF payments	-	+	+	0	0
TANF administrative costs	0	+		+	+
Medicaid payments	-	+	+	0	0
Medicaid administrative costs	0	+	+	0	+
Payroll taxes	-	+	+	-	0
Income and sales taxes	-	+	+	-	0
UI benefits	+	-	-	0	0
UI administrative costs	0	-	-	0	-
Public housing assistance	-	+	+	0	0
Public housing administrative costs	0	+	+	0	+
SNAP	-	+	0	0	0
SNAP administrative costs	0	+	+	0	+
TANF-funded employment and training costs	0	-	-	0	-
WIA-funded employment and training costs	0	-	0	0	-
Pell Grants and other federal and state financial assistance	0	-	-	0	-
Community college tuition and fees	-	0	-	0	-
Community college costs not covered by tuition	0	0	-	0	-

Component of Analysis	Participants	Government		Rest of Society	Society
		Federal	State/Local		
Other state and local funds	0	0	-	0	-
Proprietary school tuition and fees	+			-	+
Foundation costs	0	0	0	-	-
Value of improved health and enhanced self-sufficiency	+	0	0	+	+
Work-related expenses (e.g., transportation, child care, clothing)	-	0	0	0	-
Value of improved child well-being	+	0	0	+	+
Net Benefits (+) minus net Costs (-)	?	?	?	?	?

The cost components shown in Exhibit 3.3 can be displayed differently, as shown in Exhibit 3.4. This exhibit shows program costs by type of expense and by funding source (non-program costs, such as out-of-pocket costs borne by participants, are not shown in this exhibit). For example, the program costs borne by the federal government are incurred principally by the ACF, the Department of Labor, and the Department of Education, and these costs are in some program categories (such as pre-college instruction) and not others (college instruction, the costs of which are borne by community colleges).

Exhibit 3.4: Program Costs, by Category and Source

Cost Component	Government					Foundation
	ACF	DOL (WIA)	ED (Pell Grants)	Community Colleges	State and Local	
Pre-college instruction	-	-	-	-	-	-
College instruction	0	0	0	-	0	0
Other training and services	0	-	-	-	-	-
Support services	-	-	-	-	-	-

3.9.3 Net Benefit and Cost Calculations

The team will fill some of the cells of the table with dollar-denominated impact estimates, such as earnings, TANF payments, and tuition payments to community colleges and proprietary schools. The team will discount these impacts, which will cover a total analysis period of 10 years following random assignment, to reflect their value in the base year (the year during which the largest share of costs is incurred, probably 2013). At the third follow-up, the team will calculate both costs and benefits and extrapolate impacts beyond the end of the observed data in order to cover the full 10-year period. Later, longer-term follow-up will make this extrapolation step unnecessary for earnings impacts.⁸³

Other cells involve impact estimates that (a) are not denominated in dollars and thus need to be valued using a dollar amount per impact unit; or (b) are denominated in dollars, but need to be converted to the right dollar amounts. For example, the team will value the impact on months of TANF receipt using an estimate of TANF administrative costs per month, expressed in base year dollars, and will value the impact on community college costs that are not covered by tuition using the estimated cost of such services per credit hour, again in base year dollars. The team will estimate fringe benefits and various types of taxes based on earnings impacts, which are denominated in dollars; in these cases, the team will multiply the impacts by appropriate percentages.⁸⁴

The team will describe impacts that cannot be measured in—or convert to—dollars so that policy makers can assess the extent to which these effects add to or offset the reported cost-benefit results. Such measures include potentially improved emotional well-being of participants, changes in health status, the value of leisure time lost to added work effort, and the well-being of participants' children. Some of these impacts, such as health status, can be measured. While others cannot, measurable impacts on outcomes such as household income and children's school attendance are linked to the values that cannot be directly measured.

Finally, the team will examine all assumptions used in the cost-benefit computations in a sensitivity analysis to determine how robust the results are to alternative values of crucial parameters. Examples of these parameters are the decay rate of impacts (used in extrapolating measured impacts beyond the time period covered by the data), the social discount rate (used to express all impact in base year dollars), and amount of displacement associated with employment impacts (which the team will assume to be zero).

3.9.4 Calculation and Analysis of Individual Benefits and Costs

Using the approach described above, the team will calculate the dollar value of each benefit and cost component in the analysis for each sample member, including control group members as well as treatment group members, and then calculate the dollar value of each benefit and cost component as a

⁸³ The second follow-up reports may include cost/benefit analyses, depending on whether this option is funded.

⁸⁴ In the cases of fringe benefits and payroll taxes, the percentages are fixed across all earnings. In the case of income taxes, however, the percentage applies only earnings over a given threshold.

treatment-control difference—that is, as an impact—using the same estimation techniques described above. For the cost-benefit analysis, the team will discount the earnings and other dollar-denominated outcomes (to reflect their value in 2013) prior to impact estimation.

The team will base other estimates on sample members' earnings and then estimate the value of fringe benefits for each sample member based on her discounted earnings and Bureau of Labor Statistics data on benefits as a proportion of total compensation. Then the team can estimate the impact using the same regression model (except for the dependent variable) as for earnings. In a similar manner, the team will estimate each sample member's taxes based on her earnings, data on effective federal income tax rates (by income quintile) from the Congressional Budget Office,⁸⁵ payroll tax rates, and income and sales tax rates in the state where the sample member lives.

The team will base the dollar value of some components of the analysis on survey data, but will cover the same follow-up period as the earnings data. For example, the dollar value of Medicaid costs will be based on survey-reported Medicaid coverage and estimates of the Medicaid costs per covered individual in 2013.

For the cost components of the analysis, the team will measure participation in community college classes and activities for each sample member, and apply unit costs to these classes and activities based on the college attended. In this way, the team will calculate the *gross* cost of the services received and the *gross* value of other outcomes (which could be a benefit or a cost) for each sample member for each perspective.⁸⁶ The team will estimate the *net* costs and outcome values as differences in gross costs and values between the treatment and control groups, using the same statistical techniques as described in Section 3.4 for the impact analysis. This will permit the estimation of benefits and costs for individual programs and for subgroups of sample members in the same way as in the impact analysis.

PACE includes a variety of different approaches to career pathways, involving different amounts of pre-college preparation and college-level coursework (corresponding to different occupational credentials). The team will conduct the PACE cost-benefit analysis for each program in the same way as the impact analysis. The analysis also will be able to disaggregate results between large subgroups, such as sample members with lower and higher levels of academic skills at baseline. Again, it is possible that the net return on investment in one group (lower entry skills) will be lower than in the other—either because the investment is larger or the impact smaller.

⁸⁵ The effective federal tax rate is negative in the lowest income quintile due to the effect of the Earned Income Tax Credit.

⁸⁶ In these calculations, it is assumed that the unit costs of program services, such as the cost of taking a specific type of community college class, do not vary across individual sample members. Similarly, it is assumed that dollar values assigned to outcomes—such as the administrative cost of SNAP—do not vary across sample members receiving SNAP for the same period of time.

3.9.5 Open Challenges

The analysis is grappling with several issues. First, the costs covered by the analysis involve multiple federal, state, local, and private funding streams, and these streams differ across sites. The differences are especially great between HPOG sites, Year Up sites, and the remaining sites. Second, the team will have to treat statistically insignificant impacts on earnings and other outcomes, particularly in sites with relatively small samples. The current plan calls for all impact estimates, regardless of whether they are significant, but also for appropriate use of sensitivity tests when confidence in measured effects is low. Third, in most interventions, earnings gains to participants are at least partially offset by losses to others—people that were displaced from the jobs given the availability of better trained workers from the intervention programs. The present strategy is to assume no displacements, but to use sensitivity tests to determine the importance of that assumption.

3.10 Analysis Plan

The team will create a detailed analysis plan that will include specifications for each PACE program. The analysis plan will mostly discuss the first follow-up report, but will also set the analytic goals, including identification of all confirmatory hypotheses for follow-up reports at 36 and 60 months to avoid any influence on these objectives and methods from initial results. The analysis plan will be a supplement to this design report and as such will refer back to it wherever possible rather than repeating the information on the standard features of the analysis methods. The basic elements of the analysis plan will be:

- Identification of primary and secondary outcomes.
- General plan for outcomes and subgroups to be involved in exploratory but still experimental analyses.
- Identification of a small set of potential mediators for the primary education outcome at 15 months.
- Timing of program reports.

The analysis plan also will present site-specific approaches to two outcome measurement challenges outlined at a general level in Section 3.3.2: defining and operationalizing measures of progress in career pathways and entry to career-track employment.

4. Data Collection

This chapter describes data sources for quantitative analyses in the PACE implementation and impact studies.⁸⁷ For each source, it summarizes major topics measured; the rationale for collecting data by that means; and procedures for data collection, processing, and storage.

The PACE data collection strategy measures key domains and constructs in the career pathways theory of change. As summarized in Chapter 1, this theory identifies five principal domains: (1) initial participant characteristics (demographic, educational, economic, psycho-social); (2) program inputs (assessment, instruction, supports, employment services); (3) intermediate outcomes (general cognitive skills, occupation-specific competencies, psycho-social factors, career knowledge and orientation, household resources and constraints, personal and family challenges); (4) targeted main outcomes (postsecondary attainment, employment and earnings); and (5) other longer-term outcomes (income and assets, adult and child well-being, local economic growth). Exhibit 4.1 provides a broad summary of the key data sources that PACE will use to measure each of these domains.⁸⁸

Exhibit 4.1: Overview of PACE Quantitative Data Sources by Measurement Domain

Domain	Core Data Sources* (Potential Data Sources)
Baseline Characteristics	
Basic demographic and household characteristics	Basic Information Form (BIF), program records
Initial values for intermediate outcomes	Self-Administered Questionnaire (SAQ), program records
Education & employment history	BIF, program records, UI wage records
Program Inputs	
Education and training services	First follow-up, second follow-up surveys; program records (treatment group, possible for control group in some sites)
Other services	First follow-up, second follow-up surveys; program records (treatment group)
Intermediate Outcomes	
General cognitive skills	Program records (treatment group only, sites that post-test)
Psycho-social factors	First follow-up, second follow-up surveys

⁸⁷ This chapter does not cover program cost data for the cost-benefit study (discussed in Chapter 3) or qualitative data (discussed in Chapter 2).

⁸⁸ Note that three domains have either no measurement plans or very limited measurement plans: general cognitive skills, occupation-specific competencies and local economic growth.

Domain	Core Data Sources* (Potential Data Sources)
Career awareness and knowledge	First follow-up, second follow-up surveys
Resource constraints	All survey waves
Other personal and family challenges	All survey waves
Primary Outcomes	
Educational attainment	All survey waves, college records
Employment and earnings	All survey waves, UI wage records
Other adult, child, and family outcomes	Second follow-up survey Third follow-up survey <i>Child Outcomes Substudy</i>

*Data collected for both treatment and control group members unless noted otherwise.

The chapter starts by describing three major categories of planned data collection: baseline data, follow-up surveys, and administrative data. It then briefly describes potential data collection on parenting and child well-being outcomes.

4.1 Baseline Data

Measures of participants' characteristics at baseline (just prior to random assignment) have a number of important uses in the implementation and impact studies. These uses (described in Chapters 2 and 3) include: describing study populations, regression-adjusting impacts to increase the precision of estimates by controlling for chance differences arising at random assignment, adjusting for survey nonresponse bias, and estimating impacts for subgroups. Key categories of baseline data include information sites collect using two PACE-designed surveys, standardized skills assessments, and program records (e.g., applications administered prior to random assignment).

4.1.1 Baseline Surveys

To obtain baseline measures for a variety of constructs in the theory of change, two PACE forms are administered to all eligible volunteers at intake, just prior to random assignment. Together, these forms capture a broad range of background characteristics including initial values on the intermediate and main outcomes that the interventions seek to influence.

The first, the PACE Basic Information Form (BIF), obtains mostly objective information. Participants either complete the form on paper (with data subsequently entered by site staff into the web-based Random Assignment and Baseline Information Tool [RABIT]) or one-on-one with program staff, who enter information directly into RABIT. Key characteristics measured include:

- Identification and information for contacts used to locate participants for follow-up surveys;
- Basic demographic items, including age, sex, marital status, race-ethnicity, nativity, and household composition;

- English-speaking ability (self-assessed);
- Educational attainment and aspirations, college attendance by members of family of origin;
- Recent employment and earnings; and
- Household income by source.

The Self-Administered Questionnaire (SAQ) focuses on a variety of attitudes, beliefs, and psychological dispositions, as well as more sensitive personal characteristics. Participants complete the SAQ on their own, returning completed forms in a sealed envelope to program staff for mailing to Abt Associates. Major SAQ constructs include:

- Psycho-social scales for academic self-confidence, discipline, and emotional stability;
- Social support;
- Career orientation and knowledge;
- Stress, depression, and other family challenges;
- Financial and other (e.g., time) resources for school, material hardship; and
- Ever arrested/felony conviction.

4.1.2 Skills Assessments

Assessed basic reading, writing and math skills are important determinants of eligibility and placement decisions in most PACE programs. At intake, seven of the nine PACE programs administer a standardized academic skills assessment to applicants.⁸⁹ VIDA non-systematically collects assessment information from some individuals who are already qualified for college-level work, and Year Up does not use a uniform assessment tool across its eight sites.

PACE will collect summary data on baseline scores for eligible study participants (treatment and control) for eight programs. These measures will be used primarily to describe the study population and analyze how program impacts vary with characteristics of participants.

After random assignment, most programs provide academic remediation to some or all treatment group members, with levels varying across and within programs, and there often is retesting at one or more junctures in the program. For example, Madison Area Technical College administers the COMPASS at intake and retests students who complete Patient Care Academy 1 at the end of the semester. PACE will collect the results of these tests to measure changes in skill levels. Since programs generally retest only treatment group members who still are enrolled at specified testing intervals, such data change statistics will be part of the implementation study.

⁸⁹ In more limited cases, the programs obtain scores from assessments recently administered elsewhere.

Exhibit 4.2 demonstrates the substantial cross-program variation in baseline assessment tools used by PACE programs. Such variation causes difficulties in creating comparable measures of basic skills. Sites variously are using placement tests and skills assessments to gauge math, reading, and writing skills. Placement tests (Accuplacer, COMPASS) are geared towards predicting readiness for college-level math and English courses, whereas basic skills tests (CASAS, TABE) are designed to measure students' progress (though sometimes also used in placement decisions). Five PACE programs administer placement tests (or use scores from tests administered by others), and five administer basic skills tests. Three administer both, and four administer only one or the other. Exhibit 4.2 details the assessments offered for each program.

Exhibit 4.2: Academic Skills Assessments Tools Used in PACE Sites

Program	Placement		Basic Skills	
	Accuplacer ¹	COMPASS ²	CASAS ³	TABE ⁴
Bellingham Technical College	✓		✓	
Des Moines Area Community College			✓	
Everett Community College	✓		✓	
Instituto del Progreso Latino		✓		✓
Madison Area Technical College		✓		
Pima Community College	✓	✓		✓
San Diego Workforce Partnership			✓	✓
VIDA ⁵	✓	✓		✓
Whatcom Community College	✓		✓	
Workforce Development Council of Seattle-King County		✓	✓	
Year Up ⁶				

1. Accuplacer is a suite of college level placement tests offered by the College Board (<http://professionals.collegeboard.com/higher-ed/placement/accuplacer>).
2. COMPASS is a suite of college level placement tests offered by ACT (<http://www.act.org/compass/>).
3. CASAS (Comprehensive Adult Student Assessment System) is a suite of basic skills assessments that grew out of California's adult skills testing program in the 1980s (<http://www.casas.org/about-casas>).
4. TABE (Test of Adult Basic Education) is a suite of adult assessments offered by CTB/McGraw-Hill (http://www.ctb.com/ctb.com/control/productSGViewAction?p=products&subjectGradeType=G&subjectGradeCode=GRADE_ADULT).
5. VIDA also uses the Texas Higher Education Assessment (THEA), the Formal Reading Inventory (FRI) and the Wide Range Achievement Test (WRAT). VIDA accepts ACT scores and previous college enrollment in lieu of these assessments.
6. Various assessments in different sites.

For descriptive purposes it would be useful to be able to document average initial skill levels at each program.

Although the dominant form of assessment addresses academic skills, there has been an increase in the development and availability of tools for assessing a wide range of non-academic skills and needs.⁹⁰ Illustrative domains include academic self-confidence and discipline, motivation, critical thinking and problem-solving skills, emotional stability, communication skills, and social supports. The study includes measures of several of these psychosocial factors in the PACE SAQ.

4.1.3 Data Collected in Program Applications

As part of their application processes, all PACE programs collect background information on high school and previous college experiences and financial circumstances, although they differ in the degree to which this information is stored electronically. For PACE programs located within a community college system (Des Moines Area Community College, I-BEST programs, Madison Area Technical College, and Pima Community College), course-level information on previous enrollment and completion, including transfer credits, is collected. Though the other PACE programs collect data on educational background, the information is less detailed and self-reported. Other PACE programs (Instituto, VIDA, Workforce Development Council of Seattle-King County, San Diego Workforce Partnership, and Year Up) collect information on previous college enrollment to varying degrees, much of which focuses on credential and degree attainment.

4.2 Follow-up Surveys

The only way to measure many key career pathways service inputs and other constructs in the theory of change is through follow-up surveys. In PACE such surveys also have the benefit of being uniformly administered to both treatment and control group members, thus allowing comparisons across these groups. This section discusses the timing and target populations, questionnaire content, and fielding approach for each expected wave.

4.2.1 Timing and Target Populations

The team plans to survey treatment and control group members at specified intervals after random assignment. The plan is to conduct three follow-up survey waves, with interviews as soon as possible after participants reach the 15, 36, and 60 month post-random assignment milestones.⁹¹ These follow-up intervals correspond, respectively, to the short, middle, and longer run with respect to hypothesized impacts. The first follow-up interviews will take approximately 55 minutes, and will use a mixed-mode approach, initially attempting telephone interviews with in-person follow-up as needed. The target response rate is 80 percent.

⁹⁰ See tools summarized in Wilson-Ahlstrom et al. (2011), Saxon et al. (2008), and Levine-Brown et al. (2008), for example.

⁹¹ Data collection and analysis for the second follow-up reports will be conducted under a successor contract. A third contract would be required for the third follow-up study.

The first wave, starting at 15 months after random assignment, will focus on short-term impacts on service receipt, intermediate outcomes, and postsecondary educational attainment, as well as employment, income, debt, and income support program participation. This wave will cover all programs. The team lengthened the time to this first follow-up from 12 months (in the original study proposal) to 15 months given improved knowledge of the populations and programs PACE is studying. This time point captures the equivalent of a full academic, or program, year for each intervention with an additional three-month interval allowing for transitions to employment and continuing schooling. As such, it maps well to the academic and program calendars of a number of PACE interventions. Examples include the one-year Year Up program, programs where participants are attending schools operating on a traditional academic calendar (e.g., I-BEST, Madison Area Technical College, VIDA), and the roughly one- to two-year period spanned by Instituto's Carreras En Salud's series of bridge programs. Based on these considerations and an interest in coordination with the HPOG Impact Study, ACF decided that the first follow-up survey would start at 15 months (with the goal of obtaining interviews as soon as possible after this milestone).

The second follow-up survey will capture more substantial progress towards major postsecondary credentials such as an associate's degree and transfer to a four-year institution, as well as more definitive evidence on career track employment and earnings. The 60-month follow-up will provide a fuller portrait of developing career pathways (allowing us to document longer-term education and employment trajectories), as well as impacts on more general domains of adult and child well-being.

4.2.2 Content

Each survey wave will focus on domains in the theory of change for career pathways most likely to be influenced at different follow-up intervals. The first follow-up survey concentrates on education and training experiences, service receipt, and intermediate outcomes, with some coverage of employment, income, and debt. Emphasis in the 36- and 60-month surveys will shift to educational attainment, details of employment experiences (e.g., job histories), income and assets, and adult and (particularly at 60 months) child well-being. There will be some coverage of intermediate outcomes, but only minimal detail on service receipt.

As noted above, the PACE and HPOG teams are coordinating the timing of the first follow-up survey (starting call attempts 15 months after random assignment and allowing at least 6 months of effort to maximize response rates). In addition, the PACE and HPOG surveys also share the core module related to education and training service receipt (although, due to a need to keep the HPOG survey within 36 minutes, it contains a subset of the PACE questions).

For several outcomes, there will be some degree of overlap between measures obtained in follow-up surveys and administrative data. As discussed in subsequent sections, in all cases, the overlap will be relatively small, and data obtained via survey and administrative records will differ in important respects. Where there is overlap, the team will compare estimates to understand better the sensitivity of impact estimates to measurement strategies.

4.2.3 Administration

The team will use a mixed-mode approach, with an initial telephone interview stage followed by in-person interviews for those not reached by telephone. The team will design the survey field procedures to generate an 80-percent response rate in each program.

Ensuring a high response rate will require use of contact information collected at baseline, ongoing tracking using a variety of resources, advance communications, and compensation for time. In addition, the survey team will track response rates separately for treatment and control groups

Professional telephone interviewers from Abt SRBI (Abt's survey affiliate) will administer the first stage of interviews by telephone using Computer Assisted Telephone Interviewing (CATI) software. The software maximizes data quality by enforcing question skip logic and by checking data items as they are entered to make sure they are in appropriate ranges and are consistent with previous responses.

In the second stage, the team will attempt to interview first-stage nonrespondents using field staff. Field locators will find sample members through efforts such as talking to neighbors, relatives, postal employees, and others. When a locator finds a sample member, he or she will administer the survey using Computer Assisted Personal Interviewing (CAPI) software.

4.3 Data from Administrative Systems

This section describes plans to collect data from three types of administrative systems: information PACE programs gather to track participation and outcomes, databases colleges maintain to record transcript and other school-related outcomes, and databases containing wage records employers report to state Unemployment Insurance offices. For each of these systems, this section describes the kinds of measures available, primary uses of such measures in PACE analyses, strengths and weaknesses of these measures, and likely data collection approaches.

4.3.1 PACE Program Data

Each PACE program collects and maintains records on participants' characteristics, service receipt and education and training outcomes. Such data generally will be available only for treatment group members, as these are the individuals with whom programs mainly interact and monitor after random assignment.⁹² However, some program providers, particularly I-BEST colleges, Des Moines Area Community College, Madison Area Technical College and Pima Community College, also may provide other (standard) services to control group members. The team will collect control group data to the extent possible from these programs. For some programs, particularly those funded by HPOG, there will be minimal control group records because the HPOG management information system, the Performance

⁹² All programs collect data from basic academic skills assessments administered to some or all participants in both groups at intake. Thereafter, PACE community college programs may collect some data on control group members attending the wider institution but not the PACE program.

Reporting System (PRS), collects information only for the treatment group.⁹³ Exhibit 4.3 depicts data collected from programs.

Exhibit 4.3: Program Records Data

	Des Moines Area Community College	I-BEST Community Colleges	Instituto del Progreso Latino	Madison Area Technical College	Pima Community College	San Diego Workforce Partnership	Workforce Development Council of Seattle-King County	VIDA	Year Up
Educational Background									
High School									
Date awarded high school diploma	✓	✓	✓	✓	✓	✓	✓	✓	✓
Date earned GED	✓	✓	✓	✓	✓	✓	✓	✓	✓
Previous College Enrollment									
Semester/ term enrolled	✓	✓	✓	✓	✓				✓
Name of course	✓	✓	✓	✓	✓				
Credit or non-credit course	✓	✓	✓	✓	✓				
Major or program of study	✓	✓	✓	✓	✓				
Credits earned	✓	✓	✓	✓	✓				
Course grade	✓	✓	✓	✓	✓				
Credential or degree awarded	✓	✓	✓	✓	✓	✓	✓	✓	✓
Core Basic and Occupational Skills Curriculum									
Term Information									
Major or program of study	✓	✓	✓	✓	✓	✓	✓		✓
Credits attempted	✓	✓	✓	✓	✓				
Credits completed	✓	✓	✓	✓	✓				
GPA	✓	✓	✓	✓	✓				
Course Information									
Course term	✓	✓	✓	✓	✓				✓
Name of course	✓	✓	✓	✓	✓	✓	✓		✓

⁹³ See below for further discussion of the PRS.

	Des Moines Area Community College	I-BEST Community Colleges	Instituto del Progreso Latino	Madison Area Technical College	Pima Community College	San Diego Workforce Partnership	Workforce Development Council of Seattle-King County	VIDA	Year Up
Credits earned	✓	✓	✓	✓	✓				
Grade	✓	✓	✓	✓	✓	✓	✓		✓
Credential or degree awarded									
Credential or degree awarded	✓	✓	✓	✓	✓	✓	✓		
Type of credential	✓	✓	✓	✓	✓	✓	✓		
Date or term credential or degree awarded	✓	✓	✓	✓	✓	✓	✓		
Supportive Services									
Type of service receipt	✓		✓	✓			✓	✓	
Date service received	✓		✓	✓			✓	✓	
Amount of assistance received (if applicable)	✓	✓	✓	✓			✓	✓	✓

The main uses of program data in the PACE evaluation will be in implementation study analyses focused on patterns of service receipt and program completion. As described in Chapter 2, an important subset of such analyses will document rates at which participants complete and progress through successive career pathways steps. Career pathways flow analyses require program records on entries and exits from training and employment (potentially supplemented by data from other sources). Finally, as described in Chapter 3, program data on service utilization also are needed to estimate unit costs for the cost-benefit study.

The specific participant characteristics, services, and outcomes recorded by each PACE program reflect differences in program designs as well as in monitoring processes and procedures. For example, most programs provide some form of counseling or guidance, but the frequency, nature, and aspects of interactions with participants that programs deem important to record will vary. Programs variously offer services such as special tutoring, financial assistance, and in-kind assistance with transportation and child care. Programs also vary in what is recorded and how. Some services, such as counseling, are recorded in some sites (e.g., VIDA) but not in most others. For services involving monetary transfers (e.g., payments for tuition or other direct costs of education), there is generally an electronic record of the transaction. However, for services available on a walk-in (e.g., tutoring) or drop-off (e.g., child care) basis, there may only be a sign-in sheet recording who used the service.

Data Use Agreements (DUAs) with each program⁹⁴ document what is measured, how information is obtained, and how data are organized and stored. To ensure an accurate understanding of the program records, PACE staff continue to have ongoing discussions with program staff on data and data sources, including collecting and analyzing test extracts from each key data source.

As to storage, there is considerable variation in the degree to which pertinent data are maintained in centralized locations, as well as how information is recorded. For example, in community college-based sites, there is generally one system for tracking services specific to program participants and multiple databases covering transcripts and academic and support services available to the general student population. Programs often keep some records in paper files and others in spreadsheets maintained by individual counselors and other staff. For the three PACE programs funded by HPOG, the PRS, established for monitoring outcomes in the HPOG evaluation, will be a key data source. The PRS records participant intake, enrollment, services received, training and education, and employment and related outcomes. The PRS records some baseline data on participants on first enrollment into the program, and staff enter information on subsequent services and outcomes at regular intervals thereafter.

4.3.2 Data from College Record Systems

All PACE programs aim to increase college attainment, mostly through for-credit certificate and degree programs and, in some cases, non-credit programs offered at community colleges or proprietary schools. Data from college records offer one potentially attractive way to measure these and related outcomes (e.g., enrollment and credits earned) for estimating impacts on college attainment. As discussed in the last section and Chapter 2, college records on services and outcomes also may provide useful data on treatment group members' program experiences for the implementation study. This section outlines the team's approach to collecting college records and explains how they will complement data on training from PACE follow-up surveys.

Basics of College Records Data

College records systems typically track enrollment, courses taken, credits and grades earned, and certificates and degrees awarded. Colleges also keep records on hours and outcomes for non-credit courses, financial aid eligibility and receipt, and use of student services. These various types of records may be in databases maintained by different college departments—the degree of system integration and provisions for central linkage and extract creation vary by institution. In addition to individual colleges, many state agencies created student record databases containing information from multiple colleges in the state. As explained below, the coverage, content, and access to these data vary by state. A third important source of college records data is the National Student Clearinghouse (NSC). The NSC database contains a very limited number of basic data items—registration status at specific institutions and major degrees received—but has the substantial strength of covering nearly all students at the vast majority of public and private colleges nationwide.

⁹⁴ VIDA and Year Up are providing data under the auspices of its subcontract with Abt and elected to not sign a DUA.

Relative to other data sources (e.g., surveys), college records data potentially offer a higher degree of coverage and can provide measures for a larger percentage of the research sample. They are also not subject to follow-up failure. Since college record systems generally maintain historical records on each student's progress, they may help to augment PACE baseline characteristics data, as well as measure outcomes prospectively. Thus, where available, records data can provide a way to measure detailed outcomes not measured in the survey.

Sources of College Records Data

There are two potential sources of college records data: the NSC and state or local agencies that gather and centralize student records from institutions of higher education within the state or specified geographic area.

The primary source of college records will be the NSC which is available for all programs. Access to NSC data is open to the public for a modest fee, and PACE has obtained an agreement from NSC to use the data. The NSC's coverage is its substantial strength, whereas the minimal information it contains on college experiences is its main limitation. These data cover all states and the vast majority of college enrollments at public (93 percent) and private nonprofit (87 percent) institutions. Coverage of enrollments at private for-profit institutions is substantially lower (53 percent), suggesting a need for careful assessment and potential adjustments for undercoverage if feasible.

For each student in its system, the NSC records periods of enrollment, full/part-time status, schools attended, and degrees granted. Because college enrollment and degrees are important PACE outcomes, the NSC will be a useful source of data for impact analyses. Known gaps in coverage of completions again suggest a need for careful evaluation and adjustment.⁹⁵

For most sites, PACE will not use data from individual colleges, because the programs are not sufficiently isolated to be sure that differences in coverage across the treatment and control groups are minimal. However, there are several programs in which other options may be available. For example, the team obtained permission to access California's system of college records (for San Diego Workforce Partnership and Year Up Bay Area) and is pursuing access to the four colleges in the lower Rio Grande Valley to which and control groups have access (for VIDA).

Summary

The plan outlined here provides minimal college records data for all sites (from the NSC) and varying additional data for a subset of sites. Although it would be ideal to have consistent and detailed information for all programs, the broader plan to conduct separate analyses for each program does not require analysis of exactly the same outcomes.

⁹⁵ NSC has engaged a growing number of schools in its "Degree Verify" program which covers nearly all major degree completions in participating schools. Rates for these schools may provide the basis for adjusting completion rates estimated in non-Degree Verify schools.

4.3.3 Wage Records from the Unemployment Insurance (UI) Reporting System

Wage records employers report quarterly to state UI agencies (accessed through the NDNH) will be the main source of data for estimating impacts on employment and earnings in PACE. The chief virtues of these data, compared with survey-reported employment and earnings, are that they are not subject to nonresponse. They are collected in a fairly uniform way across states and on a continuous basis over time. More than 90 percent of all workers work for employers who are subject to state UI taxes and by law must report earnings in support of tax collection. The main disadvantage of UI data is that information is limited to quarterly earnings (as opposed to hourly). Thus, although the team can measure whether sample members had any employment (non-zero earnings) and their average total earnings in a calendar quarter, UI data will not indicate much more than that. For detail on hours, wages, benefits, and occupation, the project will rely on the follow-up surveys. Because such detail is essential in documenting occupational trajectories in career pathways, the team plans to supplement UI earnings data with information on hours, wages, and occupation from follow-up surveys.

Collecting UI Data via the National Directory of New Hires (NDNH)

PACE will collect UI data from the NDNH maintained by ACF's Office of Child Support Enforcement (OCSE). The NDNH is a national database covering all 50 states and the District of Columbia. In addition to affording much easier access to state data, its national coverage ensures that information for sample members moving out of state will be captured in the database, whereas in previous evaluations that relied on UI data from particular states that information was unavailable. Finally, the NDNH also includes all federal civilian and military employment.

Under the agreement with OCSE the team transmits periodic files including SSNs and other characteristics needed for impact analysis (the "pass-through file") and the agency will match SSNs to its files. OCSE will return the pass-through file attaching wage records for the pertinent follow-up period. To protect the confidentiality of wage data, OCSE will strip the returned files of SSNs but leave identifiers for program, treatment-control status, and other characteristics needed for analysis, as well as a pseudo-identifier to permit matching with later waves of data.

Timing of Matches with NDNH

Since OCSE purges UI records in its main database two years after each calendar quarter of record, the team must transmit match information for PACE sample members immediately following random assignment in order to maximize data on earnings histories prior to random assignment. Such pre-program information provides important covariates for use in regression-adjusting impact estimates, comparing impacts for sample members with varying work histories, and generally describing earnings trajectories over time. Subsequently, OCSE will match the (increasing) PACE sample to its files on a quarterly basis, adding records for each new quarter to records for each sample member. Although the duration of the MOU is for three years, it is expected that a new agreement will be entered into at its expiration and that ultimately ten years of post-random assignment earnings records will be available for analysis.

4.4 Other Possible Data Sources

The core data sources described thus far will provide measures for many basic characteristics and outcomes needed for the PACE implementation and impact studies. In commissioning PACE, ACF charged the evaluation team with identifying additional measurement approaches that would add value to the study and help to chart innovative directions for measurement for use in PACE and future evaluations. This section describes the potential child outcome substudy.

PACE programs might affect children through their parents' educational attainment, if such higher attainment results in better jobs and increased income. It is plausible that PACE participants will place higher value on education for themselves and their children as a result of their program experiences, will attain more education-related self-efficacy, and will use their higher incomes to prioritize educational resources to support their children. In a home environment that is more educationally supportive, children may place higher value on education and consequently experience more academic success. Alternatively, it is feasible that parents who initially devote more time to education and subsequently obtain better jobs as a result of participation in PACE programs may be at home less often, resulting in less supervision for their adolescent children. It may also cause a shifting of responsibilities for younger siblings to adolescent children in the household and ultimately may result in negative and risk-taking behaviors in the older children.

Although these are plausible theoretical hypotheses, and the experimental research on the effects of various state welfare reforms on children provide empirical support for some of them, to be adequately powered tests, they depend on identifying sufficiently large subsamples of treatment and control group members' children in various age groups. Because the PACE samples were not designed for this purpose, many sample members do not have minor children, and for those that do, typically they are spread across all age groups under 18.⁹⁶ As a result, analyses of child outcomes are likely to have very low power. Thus, in addition to measurement of child outcomes at the second follow-up, the team plans to include questions about parenting practices which apply to larger groupings of children than would support direct measurement of children.⁹⁷

⁹⁶ The one exception is Year Up which, because only 18-24 year olds are eligible for it, may have a concentration of younger children.

⁹⁷ The welfare reform studies referenced above suggest there is strong heterogeneity of effects across different age groupings of children.

References

- Abt Associates et al. (2006). *Effects of housing vouchers on welfare families*. Washington, D.C.: HUD. <http://www.huduser.org/portal/publications/commdevl/hsgvouchers.html>.
- Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in Medicine*, 27, 1282-1304.
- Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Science*, 58, 626-639.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444-455.
- Bell, S. H., and Bradley, M C. (2013). Calculating long-run impacts of social programs with staggered implementation. *Empirical Economics*, 44(1), 163-187.
- Bloom, H. (2006). The core analytics of randomized experiments for social research. MDRC Working Papers on Research Methodology. New York, NY: MDRC. Available at - <http://www.mdrc.org/publications/437/full.pdf>
- Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8, 225-246.
- Bloom, H., Hill, C., and Riccio, J. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*, 22(4), 551-575.
- Bloom, H., Michalopoulos, C., Hill, C., and Lei, Y. (2002). Can nonexperimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work program? New York: Manpower Demonstration Research Corporation.
- Bullock, J. C. and Ha, S. E. (2011). Mediation analysis is harder than it looks. In J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 508-521). Cambridge: Cambridge University Press.
- Carnevale, A., Smith, N., and Strohl, J. (2010). Help wanted: Projections of jobs and education requirements through 2018. Washington: Georgetown University Center on Education and the Workforce.
- Clar, M., Rothstein, J. and Schanzenbach, D. W. (2009). Selection bias in college admissions test scores. *Economics of Education Review*, 28(3), 295-307.

- Cook, T. D., Shadish, W. R., and Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4), 724-750.
- Council of Economic Advisers. (2009). Preparing the workers of today for the jobs of tomorrow. Washington, DC: Council of Economic Advisers.
- Creswell, J.W., and Plano Clark, V.L. (2007). Designing and conducting mixed methods research. Thousand Oaks, CA: Sage Publications, Inc.
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19 (2), 294-301.
- Dehejia, R. H., and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 84(1), 151-161.
- Dubow, E.F., Boxer, P. and Huesmann, L.R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill Palmer Quarterly*, 55(3), 224- 249.
- Edgington, E. S., and Onghena, P. (2007). Randomization tests (4th ed.). London: Chapman and Hall.
- Fan, J., and Judkins, D. (2006). Robust covariate control in cluster-randomized trials. Proceedings of the Section on Survey Research Methods of the American Statistical Association, 2988-2994.
- Fan, J., and Judkins, D. (2010). Robust covariate control in cluster-randomized trials with MPLUS and WinBUGS. Proceedings of the Joint Statistical Meetings [CD-ROM], pp. 951-962. Alexandria, VA: American Statistical Association.
- Fein, D. J. (2012). Career pathways as a framework for program design and evaluation: A working paper from the Innovative Strategies for Increasing Self-Sufficiency 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.
- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., and Wallace, F. (2005.) Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida.
- Flick, U. (2006). An introduction to qualitative research, 4th ed.. Thousand Oaks, CA: Sage Publications, Inc.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21-29.
- Frey, W. D., Drake, R. E., Bond, G. R., Miller, A. L., Goldman, H. H., Salkever, D. S., and Holsenbeck, S. (2011). Mental health treatment study: *Final report*. Rockville, MD: Westat.

- Glazerman, S., Levy, D. M., and Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science*, 589(1), 63-93.
- Granger, R. C. (2011). *The big why: A learning agenda for the scale-up movement*. Pathways. Winter.
- Greenberg, D. H., Michalopoulos, C., and Robins, P.K. (2001). *A meta-analysis of government sponsored training programs*. Baltimore: University of Maryland Baltimore County.
- Greenberg, D., Cebulla, A., and Bouchet, S. (2005). *Report on a meta-analysis of welfare-to-work programs*. Baltimore: University of Maryland Baltimore County.
- Greenberg, D., Meyer, R. H., and Wiseman, M. (1994). Multisite employment and training program evaluations: A tale of three studies. *Industrial and Labor Relations Review*, 47(4), 679-691.
- Greenberg, D., Meyer, R. H., Michalopoulos, C., and Wiseman, M. (2001). Explaining variation in the effects of welfare-to-work programs. Institute for Research on Poverty Discussion Paper No. 1225-01, University of Wisconsin-Madison.
- Halle, T., Kurtz-Costes, B., and Mahoney, J. (1997). Family influences on school achievement in low-income, African American children. *Journal of Educational Psychology*, 89, 527-537.
- Harris, E. (2010). Six steps to successfully scale impact in the nonprofit sector. *The Evaluation Exchange*, 15(1).
- Heckman, J. J. (2008). *Schools, skills, and synapses*. NBER Working Paper 14064. Cambridge, MA: National Bureau of Economic Research.
- Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261-294.
- Hendra, R., Dillman, K., Hamilton, G., Lundquist, E., Martinson, K., and Wavelet, M., with Hill, A., and Williams, S. (2010). *How effective are different approaches aiming to increase employment retention and advancement? Final impacts for twelve models*. New York: MDRC.
- Hill, N., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E. et al. (2004). Parent academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development*, 75, 1491-1509.
- Holland, P. W. (1986). Statistics and causal inferences. *Journal of the American Statistical Association*, 81, 945-960.
- Imai, K., Keele, L., and Tingley, D. (2010). A general approach to casual mediation analysis. *Psychological Methods*, 15, 309-334.
- Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 51-71.

- Innovative Strategies for Increasing Self-Sufficiency Project. (2009). Stakeholder views from early outreach. Washington D.C.: Abt Associates Inc. and the Department of Health and Human Services. Available at http://www.acf.hhs.gov/programs/opre/welfare_employ/isis/reports/isis_stakeholder_summary/isis_stakeholder_summary.pdf.
- Jaeger, D. A., and Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in returns to education. *The Review of Economics and Statistics*, 78(4), 733-740.
- Jenkins, D. (2006). Career pathways: Aligning public resources to support individual and regional economic advancement in the knowledge economy. New York, NY: Workforce Strategy Center.
- Jo, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychological Methods*, 13, 314-336.
- Judkins, D., Krenzke, T., Piesse, A., Fan, Z., and Haung, W.C. (2007). Preservation of skip patterns and covariance structure through semi-parametric whole-questionnaire imputation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 3211-3218.
- Judkins, D., St. Pierre, R., Guttman, B., Goodson, B., von Glatz, A., Hamilton, J., Webber, A., Troppe, P., and Rimdzius, T. (2008). A study of classroom literacy interventions and outcomes in even start (NCEE 2008-4028). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Judkins, D.R., and Porter, K.E. (2013). The remarkable robustness of ordinary least squares in randomized clinical trials. (Draft manuscript.)
- Koch, G. G., Amara, I. A., Davis, G. W., and Gillings, D. B. (1982). A review of some statistical methods for covariance analysis of categorical data, *Biometrics*, 38, 563-595.
- Koch, G. G., Tangen, C. M., Jung, J. W., and Amara, I. A. (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.
- Krenzke, T. and Judkins, D. (2008). Filling in Blanks: Some guesses are better than others—Illustrating the impact of covariate selection when imputing complex survey items. *Chance*, 21(3), 7-13.
- LaVange, L. M., Durham, T. A., and Koch, G. G. (2005). Randomization-based nonparametric methods for the analysis of multi-centre trials. *Statistical Methods in Medical Research*, 14, 281-301.
- Lesaffre, E., and Senn, S. (2003). A note on non-parametric ANCOVA for covariate adjustment in randomized clinical trials. *Statistics in Medicine*, 22, 3583-3596.
- Lin, W. (2013), Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7(1), 295-318.

- Magnuson, K. (2007). Maternal education and children's academic achievement during middle childhood. *Developmental Psychology*, 43 (6), 1497-1512.
- Magnuson, K., Sexton, H. R., Davis-Kean, P.E., and Huston, A.C. (2009). Increases in maternal education and young children's language skills. *Merrill-Palmer Quarterly*, 55 (3), 319-350.
- Manly, B. F. J. (2007). *Randomization, bootstrap, and Monte Carlo methods in biology* (3rd ed.). London: Chapman and Hall.
- Marshall, C., and Rossman, G.B. (2006). *Designing qualitative research*. Thousand Oaks, CA: Sage Publications, Inc.
- McGroder, S., and Gardiner, K. (2006). *Creating and Using the Logic Model for Performance Management*. U.S. Department of Health and Human Services Office of Child Support Enforcement. <http://www.acf.hhs.gov/programs/css/resource/creating-and-using-the-logic-model-for-performance-management>
- McLendon, L., Jones, D., and Rosin, M. (2011). *The return of investment (ROI) from adult education and training: Measuring the economic impact of a better educated and trained U.S. workforce*. New York: McGraw-Hill Research Foundation. <http://www.mcgraw-hillresearchfoundation.org/wp-content/uploads/the-return-on-investment-from-adult-education-and-training.pdf>
- National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st Century*. Washington DC: The National Academy Press.
- Nisar, H., Klerman, J. and Juras, R. (2013). *Designing training evaluations: New estimates of design parameters and their implications for design (working paper)*. Cambridge, MA: Abt Associates Inc.
- Orr, L. L. (1999). *Social experiments: Evaluating social programs with experimental methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Orr, L. L., Bloom, H. S., Bell, S. H., Doolittle, F., and Lin, W. (1996). *Does training for the disadvantaged work? Evidence from the national JTPA study*. Washington, DC: The Urban Institute Press.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. New York: Cambridge University Press.
- Pearl, J. (2001). Direct and indirect effects. In J. Breese and D. Koller (Eds.), *Uncertainty in artificial intelligence, proceedings of the seventeenth conference* (pp. 411-420). San Francisco: Morgan Kaufmann.
- Pearl, J. (2012). *The causal mediation formula—A guide to the assessment of pathways and mechanisms*. *Prevention Science*. (Prepublication. Doi: 10.1007/s11121-011-0270-1)
- Pearl, J. (2012a). *The causal mediation formula—A guide to the assessment of pathways and mechanisms*. *Prevention Science*, 13(4), 426-436.

- Pearl, J. (2012b). Interpretable conditions for identifying direct and indirect effects. Technical report R-389 of the Computer Science Department of the University of California, Los Angeles, October 2012, 3rd revision.
- Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation*, 24(2), 157-187.
- Preacher, K. J. and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891.
- Program for the International Assessment of Adult Competencies. (2013). PIAAC: A new strategy to assess adult competencies and their social and economic impact in the United States and internationally. May 2013. <http://www.caalusa.org/PIAACFactSheets5-13.pdf>
- Puma, M. J., Olsen, R. B., Bell, S. H., and Price, C. (2009). What to do when data are missing in group randomized controlled trials. Washington, DC: Institute of Education Sciences
- Quint, J., D.C. Byndloss, H. Collado, A. Gardenhire, A. Magazinnik, G. Orr, and Welbeck, R. (2011). Scaling up is hard to do. New York: MDRC.
- Roder, A., and Elliott, M. (2011). A promising start: Year Up's initial impacts on low-income young adults' careers. New York: Economic Mobility Corporation.
- Rosenthal, R. (1984). Parametric measures of effect size. In H. Cooper and L.V. Hedges (EDs.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Rosenzweig, M. R., and Wolpin, K. I. (1994). Are there increasing returns to intergenerational production of human capital? *Journal of Human Resources*, 29, 670-693.
- Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473-489.
- Scheirer, M. A. and Dearing, J. W. (2011). An agenda for research on the sustainability of public health programs. *American Journal of Public Health*, 101(11).
- Schochet, P. Z. (2008). Technical methods report: Guidelines for multiple testing in impact evaluations. (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Smith, J. A., and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125, 305-353.
- Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics*, 33(2), 230-251.

- Stevenson, D. L., and Baker, D. P. (1987). The family-school relation and the child's school performance. *Child Development*, 58(5), 1348-1357.
- Strawn, J. (2010). *Shifting gears: State innovation to advance workers and the economy in the Midwest*. Chicago: The Joyce Foundation.
- Sullivan, T. A., Mackie, C., Massy, W. F., and Sinha, E. (Eds.). (2012). Improving measurement of productivity in higher education, panel on measuring higher education productivity: Conceptual framework and data needs. Committee on National Statistics, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education, National Research Council.
- ter Braak, C. J. F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In K. J. Jockel (Ed.), *Bootstrapping and related techniques* (pp. 79-86). Berlin: Springer-Verlag, Berlin.
- Tsiatis, A. A., Davidian, M., Zhang, M., and Lu, X. (2008). Covariate adjustment for two-sample comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27, 4658-4677.
- U.S. Department of Education, Office of Vocational and Adult Education. (2011). *Promoting college and career readiness: Bridge programs for low-skill adults*. Washington, DC.
- Wachen, J., David, J., Belfield, C., and Van Noy, M. (2012). *Contextualized college transition strategies for adult basic skills students: Learning from Washington state's I-BEST program model*. New York: Community College Research Center.
- Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for P-value adjustment*. New York: John Wiley and Sons.
- Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple comparisons and multiple tests using SAS*, 2nd ed. Cary, NC: SAS Institute.
- Wilde, E. T., and Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management*, 26(3), 455-477.
- Zeidenberg, M., Cho, S. W., and Jenkins, D. (2010). *Washington state's basic education and skills training program (I-BEST): New evidence of effectiveness (CCRC Working Paper #20)*. New York: Community College Research Center.

Appendix A: Analysis of Mediation in PACE using Acyclic Graphs

Many of the mediators and outcomes will be binary instead of continuous variables with roughly normal distributions. This, of course, requires additional complexities and refinements to the analysis plans. The variances of the residual errors will no longer be constant since the variance of a binary variable depends on its mean. The econometric framework used in equation systems (3.4) through (3.7) in Chapter 3 to estimate the direct and indirect effects no longer works to even define what is meant by the terms “direct effect” and “indirect effect via” For systems of variables that are not normally distributed and do not have linear relationships to each other, “directed acyclic graphs”⁹⁸ such as the one shown below can help the team think about definitions of direct and indirect effects and how to estimate them.

Imagine that a treatment works by initiating changes down several separate pathways from randomization to primary outcome and that it were somehow possible to artificially block all but one of the pathways while allowing changes to roll down the single open pathway in exactly the same way that they would in the absence of blocking all the other pathways. Each pathway represents a mediator. In this hypothetical world, changes in the unblocked mediator would still ripple through the system, changing successor variables and ultimately the primary outcome. Any change caused to the primary outcome in this hypothetical world would be defined as the indirect effect of treatment via the (unblocked) mediator. One can do this for each mediator in the system. Finally, any residual changes in the primary outcome that are not mediated by other variables represent the direct effect of treatment.

To illustrate this more concretely, consider the career pathways theory of change in Exhibit 1.2. This figure shows six classes of intermediate outcomes. Imagine that it was possible to block the effects of being admitted into a program on five of these six: foundational academic skills, occupational skills, career orientation and knowledge, resource constraints, other personal and family challenges, and any unnamed intermediate outcomes. This would leave only changes to psychosocial factors unblocked. If there were still effects of treatment (in this hypothetical world) on primary outcomes such as performance and persistence in training, performance and advancement in jobs, income and assets, and child and adult well-being, then these effects would be indirect effects of the local PACE program via psychosocial factors.

These intuitive concepts can be written out more formally in a potential outcomes framework.⁹⁹ Following the notation in both Imai, Tingley and Yamamoto (2010) and Imai, Keele, and Tingley (2010),

⁹⁸ A directed acyclic graph is one with directed paths between nodes such that none of the paths have arrows at both ends and it is impossible following the directed pathways to return to any node once having left it. For an introduction to these graphs, see Pearl (2000).

⁹⁹ The concept of potential outcomes was developed by Jerzy Neyman in the early 1920s and is central to Rubin’s Causal Model. See Holland (1986).

let $M_i(1)$ be the value of the mediator if the student is assigned to treatment and $M_i(0)$ be the value of the same mediator if the student is assigned to control. Obviously, only one of these will be observed for each student; the other one is counterfactual. Prior to random assignment, they are both potential outcomes of treatment. Further, let:¹⁰⁰

- $Y_i(1, M(1))$ be the value of the primary outcome if the student is assigned to treatment and there is no blocking of any pathways;
- $Y_i(1, M(0))$ be the value of the primary outcome if the student is assigned to treatment but natural treatment-induced change to the mediator is blocked (while other natural treatment-induced changes are allowed to occur);
- $Y_i(0, M(1))$ be the value of the primary outcome if the student is assigned to treatment but all treatment-induced changes other than those that occur through the mediator are blocked; and
- $Y_i(0, M(0))$ be the value of the primary outcome if the student is assigned to control.

Continuing, the total effect of the intervention is then formally defined as

$$\bar{\tau} = \frac{1}{n} \sum_i [Y_i(1, M(1)) - Y_i(0, M(0))] \tag{A.1}$$

the direct effect of treatment is

$$\bar{\zeta}(0) = \frac{1}{n} \sum_i [Y_i(1, M(0)) - Y_i(0, M(0))] \tag{A.2}$$

and the indirect effect of treatment mediated by M is

$$\bar{\delta}(1) = \frac{1}{n} \sum_i [Y_i(1, M(1)) - Y_i(1, M(0))] \tag{A.3}$$

¹⁰⁰ The symbols are the same and have the same meaning as in the named papers by Imai and his coauthors, but the interpretation is given in terms of Pearl's language about path blocking.

In the language of directed acyclic graphs (blocked and unblocked pathways), the direct effect of treatment, $\bar{\zeta}(0)$, is the difference between the population average outcome under treatment if change to the mediator is blocked and the population average outcome under control conditions, while the indirect effect of treatment mediated by M , $\bar{\delta}(1)$, is the difference between the population average outcome under treatment with the population average outcome under treatment if change to the mediator is blocked. With these definitions, it is easy to see that $\bar{\tau} = \bar{\zeta}(0) + \bar{\delta}(1)$, so that the sum of direct and indirect effects is equal to the total effect.¹⁰¹

Estimating direct and indirect effects in this context (where linear models are inadequate to represent the relationships among the variables) requires some additional complexity over the methods discussed in Chapter 3. Instead of forming products of selected coefficients from a system of linear regressions, counterfactual simulations are drawn from a system of more general models and then appropriately combined. This process is illustrated with the same example as from the Chapter 3 where there is one binary baseline covariate (X), two program inputs (D_1 and D_2), two intermediate outcomes (W_1 and W_2), and one primary outcome (Y).

The figure below shows a directed acyclic graph of the system. The solid lines reflect causal relationships between conditions where the originating node can be manipulated (either directly or indirectly) and the destination node is a malleable outcome (intermediate or primary). The dashed lines represent fixed relationships between baseline covariates and outcomes. Both sets of relationships can be estimated, but the focus here is on the solid lines. As mentioned earlier, the effort to measure these is sometimes referred to figuratively as “getting inside the black box.” In this case, the black box consists of the program components, intermediate outcomes, the relationships between them, and their relationships to the primary outcome.

¹⁰¹ Pearl prefers to work with $\bar{\delta}(0) = \frac{1}{n} \sum_i [Y_i(0, M(1)) - Y_i(0, M(0))]$ and refers to it as the “natural indirect effect,” while Imai and his coauthors prefer to work with $\bar{\delta} = (\bar{\delta}(0) + \bar{\delta}(1)) / 2$ and to call it the “average causal mediation effect.” If there is no interaction between treatment and the mediator, then $\bar{\delta}(0) = \bar{\delta}(1)$ and there are no differences between A.3 and these other two definitions. If, however, there is interaction—that is, the effect of the mediator on the primary outcome is different under treatment conditions than under control conditions—then the three definitions are different. The team prefers the stated approach because it forces the direct and indirect effects to sum to the total effect (unlike with Pearl’s definition), and because we are not really interested in the effect of mediators under control conditions, so that there is no benefit of averaging $\bar{\delta}(0)$ in with $\bar{\delta}(1)$ as Imai and his coauthors do.

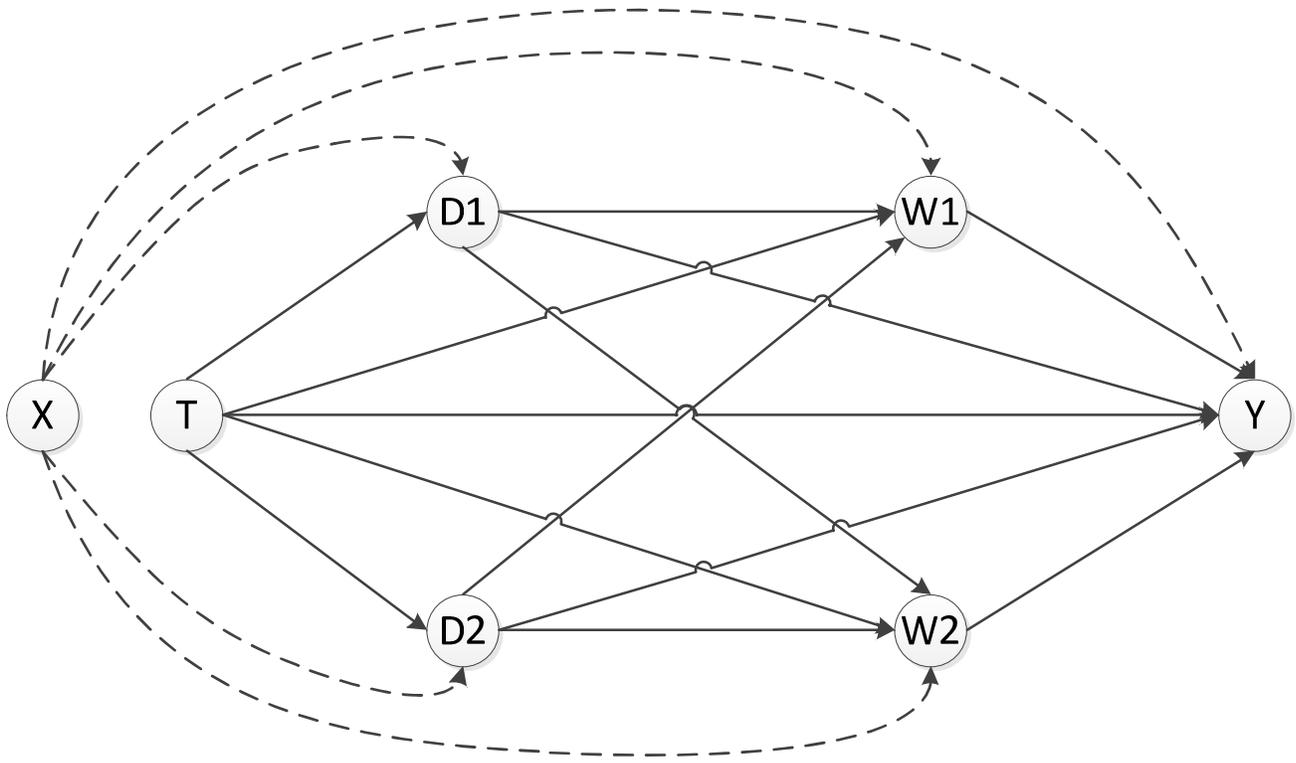


Figure: Illustrative directed acyclic graph with baseline covariates on the far left, treatment on the left, two program inputs slightly to the right, two intermediate outcomes further to the right, and the primary outcome on the far right

The estimation process starts with fitting appropriate models for each mediator and outcome:

$$\begin{aligned}
 D_{1i} &= \Phi_{D1}(X_i, T_i) \\
 D_{2i} &= \Phi_{D2}(X_i, T_i) \\
 W_{1i} &= \Phi_{W1}(X_i, T_i, D_{1i}, D_{2i}) \\
 W_{2i} &= \Phi_{W2}(X_i, T_i, D_{1i}, D_{2i}) \\
 Y_i &= \Phi_Y(X_i, T_i, D_{1i}, D_{2i}, W_{1i}, W_{2i})
 \end{aligned}
 \tag{A.4}$$

These models can be nonlinear and include interaction terms among their inputs. Once these models have been fit, they can be used to simulate values for the five variables for any possible values of their inputs without regard to what they were actually assigned or observed to experience. This is how the team will estimate counterfactual conditions where some paths are blocked. For example, if the team generates a stream of Monte Carlo values for D_1 using $\hat{\Phi}_{D1}$ with $T=0$, the team obtains a set of values for this program input that are consistent with what would have occurred under control conditions; i.e., blocking the path $T \rightarrow D_1$. These can then be fed into $\hat{\Phi}_{W1}$ and $\hat{\Phi}_{W2}$ to get simulated values of W_1

and W_2 , respectively, under the hypothetical condition that treatment had no effect on the first program input. These, in turn, could be fed into $\hat{\Phi}_Y$ to get simulated values of Y under the same conditions. If the team also blocked $T \rightarrow Y$ in conjunction with the simulated values of W_1 and W_2 based on blocked $T \rightarrow D_1$, then it will obtain simulated values of Y that reveal the indirect effect of treatment via D_2 when compared with a stream of simulated values of Y based on blocking all paths out from T .

With a total of 13 solid paths in Figure A.1 that can be either blocked or not blocked, it is possible to generate $2^{13} = 8,192$ distinct streams of simulated values for Y corresponding to different patterns of joint and partial mediation by the program inputs and intermediate values. Based on the program's theory of change, the team will only explore a few of these. If there is moderated mediation, the team can estimate this by simply tabulating the various simulated values of Y on subgroups defined by baseline covariates. This allows the team to directly address the important question of "what works for whom?"

As in the simpler approaches described in Chapter 3, strong assumptions are still required although they are weaker than those required for the simpler approaches. The assumptions basically require that each solid path in the causal graph be unconfounded. This means that other than those joint causes identified in the graph, there can be no joint causes of any pair of variables that are linked by a direct solid path. As in the simpler approaches, this assumption would be violated for example if there were any unmeasured baseline covariates that were joint causes of Y and any of the mediators. As is discussed in Chapter 3 with regard to linear models, this assumption would also be violated if there were any post-randomization events that both triggered any of the program inputs and had distinct effects on Y .