

Design and Analysis Plan for the Impact Study of Work Success

OPRE 2023-212

November 2023



Design and Analysis Plan for the Impact Study of Work Success

OPRE Report 2023-212

November 2023

April Yanyuan Wu, Owen Schochet, Kristen Joyce, and Nicardo McInnis

Submitted to:

Project Officers: Hilary Bruck, Marie Lawrence, Sarita Barton, and Megan Reid
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP233201500035I/HHSP23337035T

Submitted by:

Project Director: Annalisa Mastri
Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Reference Number: 50718

This report is in the public domain. Permission to reproduce is not necessary.

Suggested citation:

Wu, April Yanyuan, Owen Schochet, Kristen Joyce, and Nicardo McInnis (2023). *Design and analysis plan for the impact study of Work Success*, OPRE Report 2023-212, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer:

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

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

Connect with OPRE



Contents

| | |
|---|------------|
| Overview | v |
| Executive Summary | vii |
| 1. Introduction | 1 |
| 2. Background | 2 |
| 3. Work Success | 3 |
| A. Program eligibility | 3 |
| B. Coaching and other services..... | 3 |
| C. Theory of change..... | 4 |
| 4. Research Design | 6 |
| A. Research questions | 6 |
| B. Study enrollment | 6 |
| C. Random assignment..... | 6 |
| D. Data collection and uses..... | 7 |
| 5. Analysis Plan..... | 9 |
| A. Analysis priority and outcome selection | 9 |
| B. Categorizing analysis priority | 9 |
| C. Outcomes in the confirmatory analysis..... | 10 |
| D. Outcomes in the secondary analysis | 12 |
| E. Outcomes in the exploratory analysis..... | 14 |
| F. Main approach to estimating impacts | 16 |
| G. Approach to secondary analysis..... | 19 |
| H. Exploratory analysis..... | 21 |
| I. Interpreting impact estimates | 23 |
| References..... | 24 |
| Appendix A. Psychometric Analysis of Outcomes in the Self-Regulation and Goal-Setting Skills Domain..... | 27 |

Tables

| | | |
|-------|--|----|
| ES.1. | Classifying confirmatory, secondary, and exploratory analysis | ix |
| 1. | Classifying confirmatory, secondary, and exploratory analysis | 10 |
| 2. | Confirmatory outcomes..... | 11 |
| 3. | Secondary outcomes..... | 13 |
| 4. | Exploratory outcomes..... | 14 |
| 5. | Availability of baseline versions of confirmatory outcomes..... | 17 |
| 6. | Subgroups to include in exploratory analysis | 21 |
| A.1. | Goal-related items in the first follow-up survey..... | 28 |
| A.2. | Items capturing self-regulation skills in the employment context in the first follow-up survey | 28 |
| A.3. | Criteria used for assessing reliability and validity..... | 29 |
| A.4. | Reliability of measures of self-regulation skills..... | 30 |
| A.5. | Model fit statistics of measures of self-regulation skills..... | 30 |
| A.6. | Correlations between measures of self-regulation skills | 31 |

Figure

| | | |
|----|---|---|
| 1. | Program logic model for Work Success..... | 5 |
|----|---|---|

Overview

Many adults with low incomes face complex challenges that make it difficult for them to secure and keep a job or advance in their careers. Coaching by trained staff as an approach to assist adults become economically secure has garnered increasing interest among policymakers and employment program operators. Work Success is a novel program for adults with low incomes that combines intensive group and individual coaching with workshops on topics such as resume writing, job search strategies, and interviewing skills.

The Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families at the U.S. Department of Health and Human Services is sponsoring an evaluation of Work Success as part of the Next Generation of Enhanced Employment Strategies (NextGen) Project (see [project website](#)). This evaluation, which Mathematica is conducting, complements the evaluations of other coaching programs conducted through the OPRE-sponsored Evaluation of Employment Coaching for Temporary Assistance for Needy Families (TANF) and Related Populations (see [project website](#)). The evaluation of Work Success will explore how the program was implemented and its impacts on the self-regulation skills, employment, earnings, self-sufficiency, and other well-being measures of study participants.

A. Primary research questions

The overall purpose of the study is to explore whether Work Success improves the outcomes of people with low incomes. More specifically, the research questions address the following areas:

- Does Work Success affect participants' outcomes related to goal pursuit and other skills associated with labor market success?
- Does Work Success affect participants' employment and economic security outcomes?
- How do the impacts of Work Success change over time?
- Is Work Success more effective for some groups of participants than others?

B. Purpose

This report describes the design of the impact evaluation of Work Success. It discusses employment coaching and why a program such as Work Success that provides employment coaching is expected to affect participants' employment and economic security outcomes. It also provides details on the study's random assignment processes and data sources. The report concludes with a description of the analytic approach to estimating intervention impacts.

C. Key takeaways and highlights

The research design will provide rigorous estimates of the effectiveness of Work Success, a coaching program for adults with low incomes.

D. Methods

The study uses an experimental research design to assess the effectiveness of Work Success in improving employment-related outcomes, economic security, self-regulation, and other measures of well-being. Eligible adults who consented to participate in the study were randomly assigned either to a program group, which was offered access to Work Success, or to a control group, which was not offered access to Work Success but was offered a list of services available from the employment center and in the community. Both groups could access any other services available in the center or community. The effectiveness of Work Success will be assessed based on differences in outcomes between members of the program and control groups.

The study's confirmatory outcomes—the key test of Work Success' effectiveness—include goal-setting and attainment skills (a primary measure of self-regulation skills); earnings; and economic hardship, such as needing to go without food to pay bills. The outcomes will be measured at two points in time: 6 months after study enrollment and 24 months after study enrollment. The data to measure these outcomes were collected from (1) two follow-up surveys of study participants, (2) administrative employment and unemployment insurance (UI) records from the National Directory of New Hires (NDNH), and (3) administrative records on TANF and Supplemental Nutrition Assistance Program (SNAP) benefit receipt.

Executive Summary

Poverty and other chronic stressors can hinder the development and use of skills that are critical in finding, keeping, and advancing in a job. As an approach to assist adults in becoming economically independent, coaching has garnered increasing interest among policymakers and employment program operators. Employment coaching involves trained staff working with participants to set individualized goals—directly or indirectly related to employment—and develop individualized action steps for meeting those goals. Coaches take a collaborative approach by working in partnership with participants, as opposed to telling the participants what goals they should pursue or what action steps to take in pursuing them. This nondirective approach to goal setting can help people use and strengthen the skills they need to stay organized, finish tasks, and control emotions, which we refer to as self-regulation skills. By helping participants practice these skills, coaching could improve employment outcomes and hence economic security. Despite the growing interest in employment coaching programs for adults with low incomes, there is limited rigorous evidence of their effectiveness.

Work Success is a novel employment coaching program designed to help adults facing complex employment challenges use and strengthen the skills needed to find and keep a job. The program combines intensive group and individual coaching with workshops on topics such as resume writing, job search strategies, and interviewing skills.

To learn more about the potential of Work Success to improve outcomes of people with low incomes, the Administration for Children and Families (ACF) funded an experimental study of the Work Success coaching program, conducted as part of the Next Generation of Enhanced Employment Strategies (NextGen) Project. The evaluation of Work Success will explore how the program was implemented and its impacts on self-regulation skills, employment, earnings, self-sufficiency, and other well-being measures of study participants. This report describes the design and analysis plan for the impact study of Work Success.

The impact study seeks to address four main research questions:

- Does Work Success affect participants' outcomes related to goal pursuit and other skills associated with labor market success?
- Does Work Success affect participants' employment and economic security outcomes?
- How do the impacts of Work Success change over time?
- Is Work Success more effective for some groups of participants than for others?

A. Work Success program

Work Success has been administered by Utah's Department of Workforce Services since 2011. The program is offered throughout Utah through its employment centers (American Job Centers) that have one or two coaches per center. When it was first implemented in 2011, Work Success was offered to Temporary Assistance for Needy Families (TANF) recipients. In 2013, it broadened its eligibility to include, in addition to TANF recipients, other job seekers who use the employment centers.

Work Success aims to provide intensive coaching and other assistance for a few weeks to help participants obtain a job. Coaching occurs both in groups and one-on-one. In the short term, Work Success aims to reduce participants' employment challenges, improve their job search and life skills, and improve their self-regulation skills. In the long term, the program aims to increase participants' employment, job quality, job satisfaction, self-sufficiency, and well-being and reduce their need for TANF and other public assistance. To this end, the program provides comprehensive services including coaching sessions, workshops on job-related topics and self-regulation skills, and support services. Coaches work with participants to identify, develop action steps, and make progress on both long-term goals (such as those related to careers) and short-term goals that serve as stepping stones toward achieving their long-term goals. This process helps participants practice self-regulation skills. Coaches may also work with participants in developing strategies to address weaknesses in self-regulation skills that impede the participants' ability to make progress toward their goals. Participants also receive peer-to-peer support in a group setting.

B. Research design

We use an experimental research design to assess the effectiveness of the Work Success employment coaching program. Between April 2019 and February 2020, we randomly assigned, with equal probability, adults who were eligible for Work Success and who consented to participate in the study either to the program group, which was offered Work Success, or to a control group that did not have access to Work Success but was offered a list of services available from the employment center in the community. Both groups could access any other services available in the center or community.

This Work Success impact study will use data from four main sources: (1) a baseline survey administered just before study enrollment that collected data on characteristics of the study participants and information needed to locate them for follow-up surveys; (2) two follow-up surveys conducted at 6 and 24 months after study enrollment that collected data on study participants' outcomes; (3) the National Directory of New Hires (NDNH), a database maintained by ACF's Office of Child Support Services that provides data on earnings reported by Unemployment Insurance (UI) agencies as well as data on new hires and receipt of UI benefits; and (4) program administrative data on receipt of TANF and Supplemental Nutrition Assistance Program.

C. Analysis plan

We will examine the impact of Work Success on a broad set of outcomes. The risk of finding a statistically significant result by chance, rather than one representing a true effect of the program, increases with the number of outcomes tested (Schochet 2009). To minimize concerns about multiple comparisons, we will categorize analysis as confirmatory, secondary, or exploratory and set rules for reporting the impacts (Table ES.1). The confirmatory analysis will be used as the core test of program effectiveness. Secondary analysis includes impacts in domains that are less central to program goals and that are estimated using the core multivariate regression model. It also includes a Bayesian analysis approach applied to confirmatory outcomes. All other analysis will be exploratory.

Table ES.1. Classifying confirmatory, secondary, and exploratory analysis

| Type of analysis | Analysis and interpretation: Multivariate weighted least-squares regression model; statistical significance of estimates based on <i>p</i> -values | Analysis and interpretation: Bayesian approach to estimate the probability that a program had an effect of more than a specific amount on study outcomes | Analysis and interpretation: Supplement main impact estimations, such as through subgroup analysis and robustness checks | Presentation of findings |
|------------------------------|--|--|--|--|
| Confirmatory analysis | Confirmatory outcomes | n.a. | n.a. | Overview, executive summary, and main report |
| Secondary analysis | Secondary outcomes | Confirmatory outcomes | n.a. | Main report |
| Exploratory analysis | Exploratory outcomes | n.a. | Confirmatory outcomes | Appendices |

n.a. = not applicable.

Confirmatory outcomes. Confirmatory outcomes are the main outcomes that the program is expected to change. The main test of the program's effectiveness is based on whether the program had a favorable impact on the confirmatory outcomes. The confirmatory outcomes for the Work Success impact study include measures in three domains:

- 1. Self-regulation and goal-related skills.** Setting goals and working to attain them require self-regulation skills, which are the centerpiece of employment coaching. We use an eight-item scale on goal-setting and attainment skills designed to measure a person's ability to set and work toward attaining employment goals as a confirmatory outcome.
- 2. Labor market outcomes.** We use earnings as the confirmatory measure of labor market success because they encompass three ways that employment coaching could influence labor market success: obtaining a job, working more regularly or more hours, or earning higher wages. We measure earnings using both the responses to the follow-up survey and NDNH administrative records. Earnings reported on the survey cover all jobs the study participant may have had, but they may be subject to error if study participants remember jobs incorrectly. NDNH records are not subject to this error but exclude jobs not reported to UI agencies—such as self-employment or gig work—that are becoming more common.
- 3. Economic well-being.** Work Success intends to improve economic well-being. This may be accomplished through improved labor market outcomes, access to other material supports (such as assistance programs), or better financial management. We use a six-item economic hardship scale to assess the extent to which scarce economic resources affected key aspects of material well-being, such as food, housing, and medical care.

Other outcomes. We will also examine the impact of Work Success on other outcomes not deemed as confirmatory. These will include impacts on outcomes in domains likely to be of interest to readers but less central to program goals (secondary outcomes) and thus not part of the central test of program effectiveness. These may also include a broader set of outcomes (exploratory outcomes) that will not be considered key indicators of program effectiveness but could be used to broaden our understanding of overall program effects.

Estimating and interpreting impacts. We will assess the effectiveness of Work Success based on differences in average outcomes between members of the program and control groups. With random assignment, the program and control group members will have similar characteristics and experiences, on average, before participating in the program, so any differences in observed outcomes can be attributed to Work Success. We will estimate the impact on each outcome using a multivariate statistical model to control for baseline characteristics and improve the precision of the impact estimates.

For our main impact estimates, we will deem impact estimates to be statistically significant if the associated p -value of the estimate falls below 5 percent based on a two-tailed hypothesis test. We will also note if the associated p -value falls between 5 percent and 10 percent, classifying these impacts as statistically significant at the 10 percent level.

For impacts on confirmatory outcomes, we will complement this analysis of statistical significance with a Bayesian analysis, which will estimate the probability that the program's impact is greater than a specified amount. Bayesian analysis interprets programs' impacts on outcomes by accounting for the prior evidence on the effectiveness of similar programs. It also guards against the possible misunderstanding that a lack of statistical significance means a low probability of a program having an effect.

In addition to estimating impacts on exploratory outcomes, the exploratory analysis will also include (1) estimating impacts by subgroup, (2) conducting robustness checks, (3) conducting mediation analysis, (4) estimating impacts on those who participated in Work Success as opposed to all those who were assigned to the Work Success group, and (5) estimating the effect of the COVID-19 pandemic on the estimated impacts of Work Success.

Guided by this design and analysis plan, a report on the program's impacts at 6- and 24-months after study enrollment, anticipated in 2024, will address how Work Success affects participants' short- and longer-term outcomes, and whether the impacts change over time.

1. Introduction

Many adults with low incomes face complex challenges that make it difficult for them to secure and keep a job or advance in their careers. As an approach to assist adults in becoming economically secure, employment coaching by trained staff has garnered increasing interest among policymakers and employment program operators. Work Success is a novel program for adults with low incomes that combines intensive group and individual coaching with workshops on topics such as resume writing, job search strategies, and interviewing skills.

The Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families at the U.S. Department of Health and Human Services is sponsoring an evaluation of Work Success as part of the Next Generation of Enhanced Employment Strategies (NextGen) Project (see [project website](#)). This evaluation, which Mathematica is conducting, complements the evaluations of other coaching programs conducted through the OPRE-sponsored Evaluation of Employment Coaching for Temporary Assistance for Needy Families (TANF) and Related Populations (see [project website](#)). The evaluation of Work Success will explore how the program was implemented and its impacts on the self-regulation skills, employment, earnings, self-sufficiency, and other well-being measures of study participants.

This report describes the design and analysis plan for the impact study of Work Success. We begin the report by describing why employment coaching is of interest. We then describe the Work Success program; present the evaluation's research design; and, lastly, present a detailed plan for analyzing the data collected as part of the impact study.

2. Background

Poverty and other chronic stressors can hinder the development and use of the self-regulation skills that are critical in finding and maintaining employment (Mullainathan and Shafir 2013; Cavadel et al. 2017). Sometimes referred to as soft skills, executive functioning skills, or executive skills, self-regulation skills are the skills needed to stay organized, finish tasks, and control emotions (Nyhus and Pons 2005; Hogan and Holland 2003; Störmer and Fahr 2013; Caliendo et al. 2015). Examples of self-regulation skills relevant to employment include, among others, the persistence and self-efficacy needed to keep at a task despite setbacks, the time management skills that make it possible to consistently show up to work on time, and the emotional understanding and regulation to deal productively with co-workers or supervisors. Research has revealed that setting goals and developing action steps to meet them can cultivate self-regulation skills (Locke and Latham 1990; Zimmerman et al. 1992).

Coaching—in which trained staff members work with program participants to set individualized goals and then provide support and feedback as participants work toward their goals—has been shown to be an effective method for changing behavior and improving self-regulation skills (Jones et al. 2015; Fletcher and Mullen 2012). Studies of coaching have focused mostly on professional and educational settings such as in financial management (Collins and Murrell 2010; Theodos et al. 2015), higher education (Bettinger and Baker 2011), and health (Pirbaglou et al. 2018). However, there is little evidence on whether employment coaching is effective (Martinson et al. 2020). The evaluation of Work Success will help build the evidence about the effectiveness of employment coaching.

Employment coaching involves trained staff working with participants to set individualized goals—directly or indirectly related to employment—and develop individualized action steps for meeting the goals. The coach is not directive—that is, the coach does not specify goals for participants, develop plans to achieve those goals, or tell them what to do next. Rather, coaches guide participants in a collaborative process in which the participants determine their goals and develop plans to achieve them. The coach helps participants learn the skills to set goals on their own and work toward meeting those goals, reinforces participants' motivation to meet their goals, and holds participants accountable by regularly discussing with the participants their progress toward reaching goals. In this way, coaches help participants practice the self-regulation skills needed to find, keep, and advance in a job and prepare to use them after leaving the program.

3. Work Success

Work Success has been administered since 2011 by Utah's Department of Workforce Services (DWS)—an agency that oversees TANF, SNAP, programs funded by the Workforce Innovation and Opportunity Act (WIOA), and other workforce programs. The program is offered throughout Utah through its employment centers (American Job Centers), with one or two coaches per center. Each coach works with about 10 participants at a time. The program aims to provide intensive coaching and other assistance for a few weeks to help participants obtain a job. When it was first implemented in 2011, Work Success was offered to TANF recipients. In 2013, it broadened its eligibility to include other job seekers who use the employment centers in addition to TANF recipients.

Work Success is funded by multiple programs, with funding approximately proportional to the number of each program's participants who participate in Work Success. Wagner-Peyser Act Employment Services, which provides employment services to any job seeker, provides the majority of the funding. TANF provides the second-largest amount of funding. WIOA and SNAP Employment and Training also provide funding.

A. Program eligibility

Eligible participants for the program must be age 18 or older, be seeking employment, and commit to participating in the program each day. They must also be “work-ready” by having appropriate clothing for an interview in their profession, such as business casual for an office job or scrubs for a medical position, and having sufficient child care and transportation to enable them to work. Participants may face other challenges to economic security (such as mental health issues or justice system involvement) and still be eligible for the program. In addition, eligible participants were excluded from the study (but were permitted to participate in the program) if they were an active military service member, a veteran, or a spouse of a veteran; had previously participated in Work Success; or had a spouse or household member already participating in Work Success.

B. Coaching and other services

Coaching occurs both in groups and one-on-one. Each participant is assigned a coach. Four days a week, participants meet together as a group with their coach at 8:00 a.m. for a 30- to 60-minute group activity called Morning Motivation. Participants set their own short-term goals for what they will achieve that day using goal-setting worksheets. On the worksheets, participants write down their top three actionable and individualized goals for that day. Typically, these goals are small steps in their job search, such as contacting three people on LinkedIn or attending a specific job fair. The group meets again at 3:30 p.m. for 30 minutes to review participants' activities and progress on their goals, and to celebrate each other's successes.

The coaches generally meet with participants one-on-one twice a week: once for 20 to 30 minutes and once for a shorter check-in. Coaches talk to participants about realistic time frames for completing longer-term goals, emphasizing that the skills participants learn in Work Success can apply to their future goals.

All participants work toward (1) developing a resume and gathering general information needed for job applicants in a “master application,” such as contact information for references and previous employers’ addresses; (2) practicing interviewing skills; (3) completing a mock interview; (4) learning job search strategies such as networking and using websites; (5) learning budgeting skills; and (6) compiling a career portfolio—a binder that includes a targeted resume, the master application, and a letter of recommendation.

Work Success participants are required to participate in 32 hours per week of activities at the employment center and 8 hours per week of individual job search activities that can take place at another location. These activities could include attending meetings with their coach (individually or in groups), attending workshops or presentations at the employment center, working on their resume or job application materials, attending job fairs, or applying for jobs. Work Success also allows some participants to participate part-time in the program for 20 hours per week if they are unable to attend 40 hours per week due to having other commitments during the week or because they are seeking only part-time work.

The employment centers offer a variety of workshops. Available workshops, sometimes facilitated by coaches, include topics such as resume writing, job search strategies, interviewing skills, communication skills, networking strategies, budgeting and credit, LinkedIn, parenting, and relationships. The workshops are offered in person or online, and most are available to all employment center customers. In-person workshops are generally offered two or three times each month, while online workshops are available to be viewed at any time. In addition, a licensed clinical therapist delivers a weekly presentation at the employment center to Work Success participants on stress-reduction techniques and work-life balance. The therapist is available for crisis counseling and referrals to other mental health supports.

Participants can continue to attend Work Success until they find a job, which typically takes two to four weeks. Participants receive a certificate of completion if they complete the program, along with a professional portfolio to hold career-related documents.

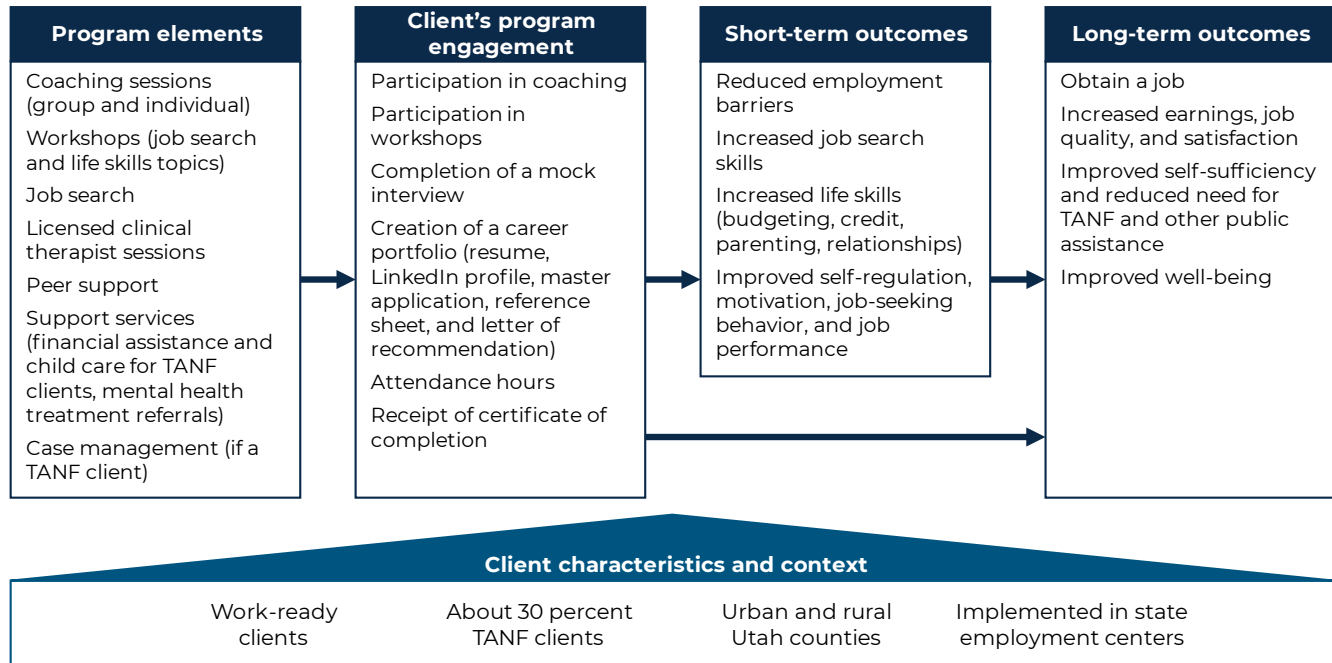
If participants are also receiving TANF benefits, they continue to work with their TANF employment counselor (Utah’s version of a TANF case manager). The employment counselor assists participants with their needs for support services, while the Work Success coach helps manage their employment-related services and work participation requirements. The hours participants spend in Work Success count toward meeting their TANF work participation requirements.

C. Theory of change

Work Success aims to reduce participants’ employment challenges, improve their job search and life skills, and improve their self-regulation skills in the short term. In the long term, the program aims to increase participants’ employment, job quality, job satisfaction, self-sufficiency, and well-being and to reduce their need for TANF and other public assistance (Figure 1).

Coaches work with participants to set individual goals and develop action steps each day. In the group setting, participants also receive peer-to-peer support, are motivated by each other’s successes, and are often informally coached by other participants.

Figure 1. Program logic model for Work Success



TANF = Temporary Assistance for Needy Families.

4. Research Design

To examine the impacts of Work Success, we used an experimental research design. Eligible participants who enrolled in the study were assigned randomly to one of two groups: (1) a program group that was offered participation in Work Success (which we refer to as the Work Success group) or (2) a control group that was not offered Work Success but was offered a list of services available from the employment center and in the community. Both groups could access any other services available in the center or community.

A. Research questions

The overall purpose of the study is to explore whether Work Success improves the outcomes of people with low incomes. More specifically, the research questions address the following areas:

- Does Work Success affect participants' outcomes related to goal pursuit and other skills associated with labor market success?
- Does Work Success affect participants' employment and economic security outcomes?
- How do the impacts of Work Success change over time?
- Is Work Success more effective for some groups of participants than others?

B. Study enrollment

The sample for the impact analysis was recruited over an 11-month period from April 2019 through March 2020 in 11 centers. Recruitment strategies included one-on-one and group presentations to community partners and to potential participants at DWS as well as marketing materials such as flyers, posters, brochures, videos, and a Work Success web page.

The evaluation aimed to enroll 1,000 participants during a 12-month enrollment period. The final study analysis sample was 752 study participants—smaller than planned due to challenges associated with the onset of the COVID-19 pandemic. Because of the pandemic, DWS canceled all services effective March 18, 2020. We therefore excluded 29 study participants who were enrolled in March 2020 from the analysis, because program services were canceled before these participants could have received two weeks of services. Some other study participants who were enrolled before March 2020 also may not have had time to finish Work Success.

C. Random assignment

When DWS intake staff identified an interested person who met the Work Success eligibility requirements, they conducted an intake appointment at a DWS office to enroll the person in the study. Program staff used the Random Assignment, Participant Tracking Enrollment, and Reporting (RAPTER™) system, a modular, web-based management information system, to document informed consent, collect baseline data on the participant's characteristics, and randomly assign consenting participants to either the Work Success group or the control group. After completing this process, the staff member informed the participant of the results of random assignment and discussed the implications.

Participants had equal probabilities of being randomly assigned to either the Work Success group or the control group. Of the 752 study sample members, 377 participants (50.1 percent) were assigned to the Work Success group and 375 participants (49.9 percent) were assigned to the control group.

D. Data collection and uses

The evaluation collected data from the following sources to support the impact study: (1) a baseline survey; (2) two follow-up surveys of study participants; (3) the National Directory of New Hires (NDNH); and (4) TANF and SNAP administrative records. This section discusses each source and how these data will be used.

1. Baseline survey

The evaluation used a 20-minute survey to collect baseline data about study participants, including both Work Success and control group members. Data collected included demographic and socioeconomic characteristics, employment status and history, self-regulation and goal-setting skills, measures capturing the propensity to receive program services, and information to locate study participants for follow-up surveys. These data will be used to (1) describe the characteristics of study participants and check that random assignment created research groups with similar characteristics, (2) provide control variables for regression models that will increase the statistical precision of impact estimates, (3) construct weights to adjust for survey nonresponse, (4) locate study participants for follow-up surveys, and (5) support exploratory analyses of program impacts within baseline subgroups and the mediating factors driving intervention impacts.

2. Follow-up surveys

The evaluation collected data from two follow-up surveys of study participants. The first follow-up survey was administered approximately 6 months after random assignment, and the second follow-up survey was administered approximately 24 months after random assignment. Data collected as part of the first and second follow-up surveys were similar. Both surveys collected information on participants' outcomes related to employment, self-sufficiency, self-regulation, and receipt of services. Data from the two follow-up surveys will be used to assess the effectiveness of Work Success, examine differences in the services received by the Work Success and control groups, and examine the mechanisms through which the program operates.

3. National Directory of New Hires

NDNH data will be used to provide information on participants' employment, earnings, and receipt of UI benefits from a year before study enrollment through the follow-up period. NDNH is a database of wage and employment information maintained by the Administration for Children and Families' Office of Child Support Services (OCSS). These data include earnings from all jobs reported to the UI agency. They exclude jobs not reported to the UI agency, such as jobs by independent contractors (such as on-call employees or temp agency employees) and other gig economy jobs (such as ride and food app delivery workers), which is a growing employment sector that is concentrated in the bottom half of the income distribution (Lim et al. 2019). The NDNH data also exclude other self-employed workers; federal government employees; members of the military; and jobs that are not reported to any government agency (such as babysitting, day laborer, or odd jobs). They also exclude earnings that employers should report but do not to avoid UI taxes, which is frequently the case for low-paying jobs (Abraham et al. 2013; Blakemore et al. 1996).

We collected NDNH data that covered the year before random assignment and two years after random assignment for all study participants. OCSS identified the NDNH records for Work Success program participants using names and Social Security numbers.

4. TANF and SNAP administrative data

TANF and SNAP administrative records from Utah's DWS will provide information on receipt of program benefits received one year before study enrollment and two years after study enrollment.

5. Analysis Plan

This section describes how we will use the data to estimate the impacts of Work Success.

A. Analysis priority and outcome selection

Work Success might affect a broad range of outcomes related to self-regulation, employment, economic stability, and other domains. However, the risk of finding a result that is statistically significant by chance, rather than one representing a true effect of the program, increases with the number of outcomes tested (Schochet 2009). Therefore, we must balance the need to examine the range of outcomes Work Success aims to affect with the need to minimize multiple comparison concerns.

We will follow two approaches to minimize the risk of focusing on findings that are statistically significant by chance. First, we will restrict the number of outcomes used for determining program effectiveness. The confirmatory tests of program effectiveness will be identified through categorizing analysis, as described below, and will not be adjusted for multiple comparisons. Before conducting the analysis, we will establish a hierarchy of reporting that places findings into one of three categories: (1) confirmatory analysis findings that will primarily feature in summaries and the main report, (2) secondary analysis findings that will be presented in the main report and also included in summary sections of the report if they contribute to understanding the confirmatory findings, and (3) exploratory analysis findings that will be reported in an appendix but will only be included in the main report or summaries if they add to our understanding of impacts on the confirmatory outcomes. Second, we will conduct multiple comparison adjustments as robustness checks. We plan to use the results that are unadjusted for multiple comparisons for the main analysis because the statistical adjustments reduce statistical power, or the likelihood of identifying a true effect (Schochet 2009). However, we will assess whether the results of the confirmatory analysis are robust to multiple comparison adjustments. We will qualify findings that are not robust to these adjustments as having weaker evidence of effectiveness.

B. Categorizing analysis priority

We plan to use a reporting hierarchy in which analysis is specified in advance as confirmatory, secondary, or exploratory (Table 1). The confirmatory analysis will be used as the core test of program effectiveness. As such, it will be presented prominently in the main report and in summary and bottom-line descriptions of program effectiveness. Secondary analysis includes impacts in domains that are less central to program goals and that are estimated using the core multivariate regression model. It also includes a Bayesian analysis approach applied to confirmatory outcomes. Secondary analysis findings will be presented in the main report and in summary sections of the report if they contribute to understanding the confirmatory findings. All other analysis will be exploratory. Because exploratory findings are not part of the main test of program effectiveness, there is flexibility in what outcomes are included in the exploratory analysis and in how they are specified. In addition, we will include selected exploratory findings in the main report and summary sections of the report if they help us interpret the confirmatory findings. We will present all other exploratory findings in an appendix.

Table 1. Classifying confirmatory, secondary, and exploratory analysis

| Type of analysis | Analysis and interpretation: Multivariate weighted least-squares regression model; statistical significance of estimates based on p-values | Analysis and interpretation: Bayesian approach to estimate the probability that a program had an effect of more than a specific amount on study outcomes | Analysis and interpretation: Supplement main impact estimations, such as through subgroup analysis and robustness checks | Presentation of findings |
|-----------------------|---|--|--|--|
| Confirmatory analysis | Confirmatory outcomes | n.a. | n.a. | Overview, executive summary, and main report |
| Secondary analysis | Secondary outcomes | Confirmatory outcomes | n.a. | Main report |
| Exploratory analysis | Exploratory outcomes | n.a. | Confirmatory outcomes | Appendices |

n.a. = not applicable.

C. Outcomes in the confirmatory analysis

The confirmatory analysis will focus on a small set of outcomes across domains that are central to Work Success’ goals and are feasible to assess, given the study’s sample size and the length of its follow-up period (Table 2). In selecting outcomes for the confirmatory analysis, we considered Work Success’ logic model and the outcomes most central to the program’s goals. The outcome domains for the confirmatory analysis include (1) self-regulation and goal-related skills, (2) labor market outcomes, and (3) economic well-being. We considered including receipt of public assistance as a confirmatory outcome, but decided against doing so because only about one-fifth of study enrollees received TANF at the time of study enrollment. As previously described, the evaluation collects survey and administrative data for study participants at two follow-up points. Similar confirmatory outcomes will be used for the analysis of the first and second follow-ups, although, as described in Table 2, the time frame for their measurement differs.

Table 2. Confirmatory outcomes

| Outcome and data source | Measure | Justification for selection and other comments |
|---|---|--|
| Self-regulation and goal-related skills | | |
| Goal-setting and attainment skills <ul style="list-style-type: none"> First and second follow-up survey data | Eight-item scale ($\alpha = 0.85$) indicating the average level of agreement—from “strongly disagree” (= 0) to “strongly agree” (= 3)—that a respondent reports on items about goal-related skills: <ul style="list-style-type: none"> I know I need to get a job or a better job and really think I should work on finding one I set employment goals based on what is important to me or my family I set long-term employment goals that I hope to achieve (such as finding a job, finding a better job, getting promoted, or enrolling in further education) I set specific short-term goals that will allow me to achieve my long-term employment goals Based on everything I know about myself, I believe I can achieve my employment goals When I set employment goals, I think about barriers that might get in my way and make specific plans for overcoming those barriers Even when I face challenges, I continue to pursue my employment goals I keep track of my overall progress toward my long-term employment goals and adjust my plans if needed These are study-developed questions. This measure is available on both the first and second follow-up surveys. | Work Success intends to improve goal-setting and attainment skills as participants receive program services and for these improvements to persist over time. We therefore will include this measure in the confirmatory analysis for both the first and second follow-up periods. The psychometric properties of this measure indicate that it is appropriate for the evaluation populations, as described in the appendix. <p>Other aspects of self-regulation skills measured in the survey, such as task management, might be improved through setting goals. One or more such skills might be a focus for some Work Success participants, as needed. However, the program does not set out to improve any specific skill. Thus, we propose examining other self-regulation skills as part of the exploratory analysis rather than the main test of program effectiveness in this domain.</p> |
| Labor market outcomes | | |
| Earnings <ul style="list-style-type: none"> First and second follow-up survey data | Average monthly earnings during the follow-up period <p>For the first follow-up period, we will define the reference period for the measure as the first 6 months after study enrollment.</p> <p>For the second follow-up period, we will define the reference period as Months 7 through 24 after study enrollment.</p> | Work Success aims to improve labor market outcomes and would expect these improvements to emerge by the time of the first follow-up and persist over time. For this reason, we will include both first and second follow-up measures in the confirmatory analysis. <p>We selected earnings for the confirmatory analysis of labor market outcomes because it encompasses a wide range of ways in which the intervention could affect the labor market success of participants. These include increasing their likelihood of working at all; working more regularly (more weeks, months, or quarters); working more hours when they do work (full-time instead of part-time, for example); and earning higher wages when they do work. All these effects would show up as an increase in total earnings.</p> |

| Outcome and data source | Measure | Justification for selection and other comments |
|--|--|--|
| <p>Earnings</p> <ul style="list-style-type: none"> First and second follow-up survey data; administrative data (NDNH) | <p>Average monthly earnings during the follow-up period</p> <p>For the first follow-up period, we will define the reference period for the measure as the first two quarters after study enrollment.</p> <p>For the second follow-up period, we will define the reference period as Quarters 3 through 8 after study enrollment.</p> | <p>We recommend examining earnings using both administrative records data (NDNH data) and survey data because these data sources have both strengths and limitations. Unlike survey data, NDNH data have no recall error, and they are available for a longer reference period than the follow-up survey. In contrast, the main advantage of the survey data is that they will pick up additional kinds of employment not covered by the administrative data. Survey data cover self-employment (such as ride-app driving and gig economy employment) and under-the-table or informal employment, which can be common among low-wage workers and often not covered by unemployment insurance benefits and thus not included in NDNH data.</p> <p>By examining earnings from both data sources, we reduce the risk of missing the effect that the program may have had on earnings in either formal or informal jobs. The proposed approach—focusing on earnings and using administrative and survey data—has been widely used in prior studies of the impacts of employment and training programs on low-income individuals. Earlier studies that have used this approach include PACT, Parents' Fair Share, Job Corps, WIA Gold Standard Evaluation, and CSPED (see Barnow and Greenberg [2015] for a comprehensive review of many of these earlier studies).</p> |
| Economic well-being | | |
| <p>Economic hardship</p> <ul style="list-style-type: none"> First and second follow-up survey data | <p>A count ranging from 0 to 6 of the number of the following coping strategies used to stretch budgets:</p> <ul style="list-style-type: none"> Cut the size of meals or skipped meals because couldn't afford enough food Moved in with other people because of financial problems Asked to borrow money from friends or family Went without a phone because could not afford to pay the bill Sold belongings or took out a payday loan Went without medical care because of cost | <p>Work Success intends to improve participants' material well-being. This economic hardship scale reflects the extent to which scarce economic resources affected key aspects of material well-being, such as food, housing, and medical care.</p> |

Note: The first follow-up period is the first 6 months after study enrollment. The second follow-up period is 24 months after study enrollment.

CSPED = National Child Support Noncustodial Parent Employment Demonstration; NDNH = National Directory of New Hires; PACT = Parents and Children Together; WIA=Workforce Investment Act.

D. Outcomes in the secondary analysis

In the secondary analysis, we will examine impacts on outcomes in domains likely to be of interest to readers but that were less central to program goals and thus not part of the central test of program effectiveness. This category does not include domains in the confirmatory analysis. For the secondary analysis, we propose outcomes in four domains (Table 3).

Table 3. Secondary outcomes

| Outcome and data source | Measure | Justification for selection and other comments |
|--|---|--|
| Service receipt | | |
| Ever received one-on-one job assistance (percentage) | This is a binary variable that equals 1 if the respondent reported receiving one-on-one job assistance. This outcome is available at both first and second follow-up. | Work Success' impacts on service receipt could lead to impacts on other outcomes. |
| Number of times received one-on-one job assistance | This measures the number of times the respondent received one-on-one job assistance. This outcome is available at both first and second follow-up. | Work Success' impacts on service receipt could lead to impacts on other outcomes. |
| Hard skill acquisition | | |
| Completion of an education program <ul style="list-style-type: none"> First and second follow-up survey data | This is a binary variable that equals 1 if the respondent reported completing an education program and 0 otherwise. This outcome is available at both first and second follow-up. | Work Success' impacts on education could lead to impacts on earnings and other outcomes. |
| Completion of a training program <ul style="list-style-type: none"> First and second follow-up survey data | This is a binary variable that equals 1 if the respondent reported completing a training program and 0 otherwise. This outcome is available at both first and second follow-up. | Work Success' impacts on training could lead to impacts on earnings and other outcomes. |
| Job quality | | |
| Employment in jobs offering fringe benefits <ul style="list-style-type: none"> First and second follow-up survey data | Whether the respondent was employed in a job offering fringe benefits at the end of the follow-up. The reference period for these outcomes will align with the earnings measures in Table 2. | Understanding impacts on job quality is key to understanding impacts on earnings. Moreover, improved job quality could improve other aspects of economic stability and well-being. Receipt of fringe benefits is an important measure of the quality of a job. |
| Employment challenges | | |
| Specific challenges that impeded employment <ul style="list-style-type: none"> First and second follow-up survey data | Seven separate outcomes corresponding to a particular challenge: <ul style="list-style-type: none"> The first is whether (yes or no) the survey respondent reported having a current driver's license at the time of the follow-up survey. The other six outcomes are a binary variable indicating whether the respondent reported that having the challenge made it very hard or extremely hard to find and keep a good job at the time of the follow-up survey: <ul style="list-style-type: none"> Child care Transportation Right clothes or tools Right skills or education Criminal record Limiting health reason | Work Success' services and potential impacts on self-regulation skills could influence participants' ability to address employment challenges. Such effects could be important for understanding impacts on earnings and receipt of public assistance. |
| Housing stability | | |
| Unstable housing <ul style="list-style-type: none"> First and second follow-up survey data | A binary variable indicating whether the respondent was unhoused, living in a shelter, or living rent-free at the time of the follow-up surveys. | Work Success' services and potential impacts on labor market outcomes could influence participants' housing stability. |

Note: The first follow-up period is the first 6 months after study enrollment. The second follow-up period is 24 months after study enrollment.

CSPED = National Child Support Noncustodial Parent Employment Demonstration; NDNH = National Directory of New Hires; PACT = Parents and Children Together.

E. Outcomes in the exploratory analysis

In the exploratory analysis, we will examine a broader set of outcomes, which will not be considered key indicators of program effectiveness but could be used to broaden our understanding of overall program effects. The main function of this analysis is to further investigate the confirmatory findings and identify how they emerged.

We provide a list of outcomes to be included in the exploratory analysis below (Table 4). These include supplementary outcomes in the domains with confirmatory outcomes (self-regulation and goal-related skills, labor market outcomes, and economic well-being) as well as outcomes from additional domains such as education and training, job search, receipt of public assistance or UI benefits, criminal activity, and marital status. In addition to secondary outcomes in the service receipt domain, the exploratory analysis will include impact estimates on other service receipt outcomes drawn from survey reports collected from both Work Success and control group members. These findings will inform interpretation of other findings because impacts on these outcomes would supply information about the contrast between the services received by the Work Success and control groups. The list of exploratory outcomes is likely to expand as we further investigate patterns in the confirmatory analysis.

Table 4. Exploratory outcomes

| Outcome | Data source |
|---|--|
| Self-regulation and goal-related skills | |
| Task monitoring, planning, and initiation at the time of the first and second follow-up surveys | First and second follow-up surveys |
| Emotional control and self-monitoring at the time of the first and second follow-up surveys | First and second follow-up surveys |
| Employment self-regulation at the time of the first and second follow-up surveys | First and second follow-up surveys |
| Self-esteem at the time of the first and second follow-up surveys | First and second follow-up surveys |
| Set an employment goal at the time of the first and second follow-up surveys | First and second follow-up surveys |
| Labor market outcomes | |
| Labor market participation | |
| Whether in labor market at time of the follow-up surveys (employed or looking for a job) | First and second follow-up surveys |
| Whether actively engaged at the time of the follow-up surveys (employed, looking for a job, in school or training, or caring for a family member) | First and second follow-up surveys |
| Employment outcomes | |
| Whether employed by month after study enrollment | First and second follow-up surveys |
| Number of months employed | First and second follow-up surveys |
| Whether employed by quarter after study enrollment | NDNH data for the first and second follow-up periods |
| Number of quarters employed | NDNH data for the first and second follow-up periods |
| Hours worked per week during the first and second follow-up periods | First and second follow-up surveys |
| Whether a new hire during the first and second follow-up periods | NDNH data for the first and second follow-up periods |
| Number of months employed in a full-time job during the first and second follow-up periods | First and second follow-up surveys |
| Number of months self-employed during the first and second follow-up periods | First and second follow-up surveys |

Chapter 5. Analysis Plan

| Outcome | Data source |
|---|--|
| Job quality | |
| Number of months employed in job with wage rate higher than the U.S. 25th percentile (about \$14) during the first and second follow-up periods | First and second follow-up surveys |
| Number of jobs with each of the following benefits: health insurance, paid leave, retirement benefits during the first and second follow-up periods | First and second follow-up surveys |
| Number of months employed in a (full- or part-time) wage and salary job (excluding seasonal, contract, on-call, and odd jobs) during the first and second follow-up periods | First and second follow-up surveys |
| Whether employed in a job with high perceived likelihood of promotion in next 12 months at the time of the follow-up surveys | First and second follow-up surveys |
| Whether satisfied with job held at the time of the follow-up surveys | First and second follow-up surveys |
| Education and training | |
| Participation in an education program during the first and second follow-up periods | First and second follow-up surveys |
| Participation in a training program during the first and second follow-up periods | First and second follow-up surveys |
| Receipt of a diploma or degree from an education program during the first and second follow-up periods | First and second follow-up surveys |
| Receipt of a certificate, license, or diploma from a training program during the first and second follow-up periods | First and second follow-up surveys |
| Highest level of education at time of the follow-up surveys | First and second follow-up surveys |
| Job search | |
| Number of job search activities conducted since random assignment (updated resume, explored requirements for a job, found child care, looked into training, looked into transportation) during the first and second follow-up periods | First and second follow-up surveys |
| Number of job offers received when working and number of job offers received when not working during the first and second follow-up periods | First and second follow-up surveys |
| Intensity of job search (frequency of activities) when working and intensity of job offers received when not working during the first and second follow-up periods | First and second follow-up surveys |
| Receipt of public assistance or social insurance benefits | |
| Average monthly TANF benefit amount during the first and second follow-up period | Administrative records data for the first and second follow-up periods |
| Receipt of any income from public assistance or social insurance programs (TANF, SNAP, UI, SSI, SSDI, WIC, or housing assistance) during the first and second follow-up periods | First and second follow-up surveys |
| Amount of TANF benefits received by month after study enrollment | Administrative records data for the first and second follow-up periods |
| Amount of SNAP benefits received by month after study enrollment | Administrative records data for the first and second follow-up periods |
| Total amount of UI benefits received during the first and second follow-up periods | NDNH data for the first and second follow-up periods |
| Criminal activity | |
| Whether convicted of a crime since study enrollment during the first and second follow-up periods | First and second follow-up surveys |
| Whether convicted of a felony since study enrollment during the first and second follow-up periods | First and second follow-up surveys |
| Marital status | |
| Whether married at time of the follow-up surveys | First and second follow-up surveys |

| Outcome | Data source |
|---|------------------------------------|
| Service receipt | |
| Total time spent in one-on-one employment services (average and frequency) during the first and second follow-up periods | First and second follow-up surveys |
| Number of times received group employment services (average and frequency) during the first and second follow-up periods | First and second follow-up surveys |
| Received job assistance focused on setting long-term goals during the first and second follow-up periods | First and second follow-up surveys |
| Received job assistance focused on setting short-term goals during the first and second follow-up periods | First and second follow-up surveys |
| Received job assistance focused on planning to achieve your goal during the first and second follow-up periods | First and second follow-up surveys |
| Received a career assessment during the first and second follow-up periods | First and second follow-up surveys |
| Received job leads from a program during the first and second follow-up periods | First and second follow-up surveys |
| Received child care services during the first and second follow-up periods | First and second follow-up surveys |
| Received transportation assistance during the first and second follow-up periods | First and second follow-up surveys |
| Received clothes, uniforms, tools, or other supplies and equipment during the first and second follow-up periods | First and second follow-up surveys |
| Received tuition assistance during the first and second follow-up periods | First and second follow-up surveys |
| Received assistance finding stable housing during the first and second follow-up periods | First and second follow-up surveys |
| Received assistance with budgeting, credit, banking, or other financial matters during the first and second follow-up periods | First and second follow-up surveys |
| Received assistance expunging a criminal record or other legal assistance during the first and second follow-up periods | First and second follow-up surveys |
| Received help related to domestic violence during the first and second follow-up periods | First and second follow-up surveys |
| Received help with marital and other family relationships during the first and second follow-up periods | First and second follow-up surveys |
| Received help with child behavioral issues during the first and second follow-up periods | First and second follow-up surveys |
| Received cash or a gift card during the first and second follow-up periods | First and second follow-up surveys |
| Received substance use counseling during the first and second follow-up periods | First and second follow-up surveys |
| Received mental health treatment during the first and second follow-up periods | First and second follow-up surveys |

NDNH = National Directory of New Hires; SNAP = Supplemental Nutrition Assistance Program; SSDI = Social Security Disability Insurance; SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families; UI = Unemployment Insurance; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

F. Main approach to estimating impacts

The main impact estimates for all outcomes will be based on the study's experimental design. As previously described, participants eligible for the coaching services were randomly assigned to one of two groups: (1) a Work Success group offered coaching services or (2) a control group not offered coaching services. With this design, the research groups should be very similar in terms of their characteristics before being randomly assigned. Our basic analytic approach is to compare the outcomes of members of the Work Success and control groups. Because of random assignment, differences in observed outcomes between the groups that are large enough to be unlikely due to chance can be attributed to the Work Success employment coaching.

1. Multivariate estimation and covariates

We will estimate the impact of Work Success on each outcome by using a multivariate weighted least-squares regression model. There are two reasons for estimating a regression model rather than using the difference in the average value of an outcome between the Work Success and control groups. First, the regression model will enable us to adjust for any differences in baseline characteristics between the Work Success and control groups that emerge by chance, despite the random assignment design. Second, including covariates in the model that are correlated with the outcome measure will improve the statistical precision of the impact estimates (Orr 1999).

We will select covariates based on likely correlation with program outcomes, because highly correlated covariates will increase the precision of the impact estimates. We will include baseline versions of all confirmatory outcomes that are available because baseline and follow-up versions of the same measure are likely to be highly correlated (Table 5). This means that for the analysis of earnings outcomes based on follow-up surveys, we will include earnings reported on the baseline survey as a covariate; for analysis of earnings outcomes based on NDNH data, we will include pre-enrollment earnings from the NDNH data as covariates.

In addition, we will include a set of covariates to control for other baseline characteristics with statistically significant differences between the Work Success and control groups. We have tested for average differences between Work Success and control group members in a wide range of baseline characteristics, including age, sex, race and ethnicity, marital status, number of adults with whom the respondent lives, number of children the respondent lives with, whether the study participant has a high school diploma or General Educational Development (GED) credential, and challenges the study participant faces in seeking or maintaining employment. As expected, given the random assignment research design, there were no significant differences in baseline characteristics between the program and control groups.

Table 5. Availability of baseline versions of confirmatory outcomes

| Outcome | Availability |
|--|--|
| Self-regulation and goal-related skills | |
| Goal-setting and attainment skills, first and second follow-up surveys | Three of the eight items ^a |
| Labor market outcomes | |
| Earnings, survey data | Earnings in 30 days before enrollment |
| Earnings, administrative data | Available only for analysis of outcomes based on NDNH data |
| Economic well-being | |
| Economic hardship | No |

Note: The first follow-up period is the first 6 months after study enrollment. The second follow-up period is 24 months after study enrollment.

^a See Appendix Table 1 for specific items.

NDNH = National Directory of New Hires.

The covariates will also include a set of indicator variables related to the timing of study enrollment. These covariates will control for common timing effects experienced by enrollment cohorts, such as those related to labor market conditions, the COVID-19 pandemic, or other factors.

2. Statistical significance

For all outcomes (confirmatory, secondary, and exploratory), we plan to report statistical significance of impact estimates based on p -values. This approach will allow us to determine whether Work Success has an effect on an outcome and to quantify the magnitude of any effects.

For our main impact estimates, we will deem impact estimates to be statistically significant if the associated p -value of the estimate falls below 5 percent based on a two-tailed hypothesis test. We will also note if the associated p -value falls between 5 percent and 10 percent, classifying these impacts as statistically significant at the 0.10 level.

To help interpret the magnitude of the impact estimates, we will calculate an effect size for each outcome. We will report effect sizes in the main report for outcomes measured as scales because the magnitude of impacts on these outcomes is not easily interpretable. We will report effect sizes for all outcomes in the appendices. For continuous outcomes, we will calculate the effect size as Hedges' g , which equals the impact estimate from the regression model divided by the unadjusted pooled standard deviation of the outcome for respondents across both program and control groups (Hedges 1981). For binary outcomes, we will calculate the effect size as the Cox index, which equals the log odds ratio divided by the constant 1.65 (Cox 1970).

3. Treatment of missing baseline data

When one or more covariates have missing data, we will use dummy variable adjustment, which involves setting any missing baseline values to a single constant value and including flag variables for missing values as additional covariates in the regression model. This approach is appropriate when the covariates are not correlated with research groups, as is the case in evaluations with a random assignment design (Deke and Puma 2013; Puma et al. 2009).

4. Treatment of missing outcome data

We will estimate all regressions using weights to account for sample members who did not complete the follow-up survey or who could not be matched to the administrative data because of missing or inaccurate Social Security numbers. The nonresponse weights will adjust the data to be representative of all sample members, not just those who completed the survey or could be matched to an administrative record. Using regression analysis, we will calculate the weights by estimating the probability of nonresponse for study participants as a function of their baseline characteristics. We will adjust the standard errors of the impact estimates to account for the variability associated with these weights.

We will use imputation to address item nonresponse that affects a subset of items used to create survey outcomes. The following are examples:

- If a sample member responded to at least two-thirds of the items on a scale, we will use the average scale score for that person based on the available items.
- We will impute minor missing items from the job grid with midpoint values when constructing earnings measures.
- For more substantive missing items, such as length of time in the job, we will use hot-decking or other appropriate procedures to impute based on job characteristics and other relevant follow-up and baseline data.

G. Approach to secondary analysis

In addition to estimating impacts on secondary outcomes, we will conduct Bayesian analysis on confirmatory outcomes as a secondary analysis.

1. Bayesian analysis

To help readers interpret the findings, we will complement our reporting of statistical significance on earnings outcomes with a Bayesian analysis. The Bayesian analysis will provide a probability that the true effect of Work Success is positive or greater than a specified amount—this is nuanced information that is helpful to practitioners and policymakers rather than just a conclusion that the program is probably effective or not. The Bayesian analysis also guards against the frequent misunderstanding about the meaning of statistical significance, which can lead to serious misinterpretation of study findings. Many people misinterpret statistical significance (p -value < 0.05) to mean that there is at most a 5 percent chance that the program had no effect rather than the correct conclusion that when the true effect is zero, there is a 5 percent chance that the impact estimate is statistically significant. A statistically significant impact does not necessarily imply a high probability that the program had an effect. Similarly, a lack of statistical significance does not necessarily mean that there is a low probability that the program had an effect.

Overview of the BASIE approach. We plan to present findings from a Bayesian approach known as BASIE (BAyEgian Interpretation of Estimates) (Deke and Finucane 2019).¹ We will apply this approach to estimate the probability that coaching had an effect of more than a specific amount on key study outcomes, rather than an indication of whether coaching had an effect at all. This approach applies Bayesian methods, drawing on both the effect directly estimated from the study's data and prior evidence about how common it is for programs to have effects.

The BASIE approach directly estimates the probability that the true effect of a program is a certain size. For example, we could draw conclusions about the likelihood that the impact is positive, such as, “There is a 75 percent chance that Work Success had a positive effect on average monthly earnings.” In addition, we could draw conclusions about the probability that the program had a large effect that readers are likely to regard as meaningful, such as, “There is a 50 percent chance that Work Success boosted average monthly earnings by \$250 or more.”

How the BASIE approach compares to other Bayesian methods. The BASIE approach differs from how researchers often apply Bayesian methods in two key ways. First, a common concern with Bayesian methods is that they can be subjective. Instead of drawing on prior evidence, they sometimes rely on prior beliefs about the effects of a program (Cooper et al. 2009). The BASIE framework avoids this concern by drawing only on prior evidence from similar evaluations, rather than on a researcher's beliefs about a program's effects. Second, under the standard Bayesian approach, researchers often report only the Bayesian shrunken estimate (which is a weighted average of the traditional effect estimate and prior evidence). If the Bayesian statistical model is correct, the shrunken estimate is less susceptible to statistical noise. In contrast, the BASIE approach encourages researchers to report both the main impact estimate (based only on study data) and the Bayesian shrunken estimate (Cooper et al. 2009; Gelman et al. 2012).

¹ The components of BASIE draw on guidance from many sources (Gigerenzer and Hoffrage 1995; Gelman and Weakliem 2009; Gelman 2001, 2012, 2015a, 2015b, 2016; Gelman and Shalizi 2013).

Information required to implement BASIE. The BASIE approach requires information that will come from our main analysis and additional information from other sources. In particular, the approach requires (1) the effect estimate and standard error and (2) prior information on how common it is for generally similar programs to have effects. The additional prior information will allow us to quantify how common it is to achieve effects of different sizes, such as how common it is to achieve positive effects or effects greater than a particular size.

Guidelines to selecting prior information. BASIE applies five guidelines for selecting and analyzing prior information: (1) use evidence from past evaluations, as opposed to beliefs about the effectiveness of programs that are not based in evidence; (2) select prior evidence that meets systematic standards for quality, such as studies reviewed by evidence clearinghouses; (3) statistically adjust evidence for variation in precision and possible biases that arise from how effects are reported; (4) select evidence that is similar to the programs and populations in the evaluation; and (5) examine and report sensitivity of findings to the selection of prior evidence.

There is not a general guideline for the number of past evaluations to assess. We will aim for 30 studies for this analysis. It is not necessary for the evidence we draw on for this work to come from evaluations of coaching programs, so long as we are able to clearly articulate what the evidence represents. For example, the evidence could represent programs intended to help people with low incomes improve their employment outcomes.

Recommended source of prior information. We will base the priors on OPRE's Pathways to Work Evidence Clearinghouse (Pathways). Pathways aligns closely with the Work Success evaluation because it focuses on studies of employment and training interventions for populations with low incomes. As described in more detail below, Pathways has enough studies to enable us to form prior information.

Selecting priors from Pathways. In advance of conducting the analyses, we plan to identify (1) a set of prior information that we will use for the Bayesian analysis to be presented in the body of the report and (2) other sets of prior information that we will use for sensitivity analyses to be presented in the appendix. We will identify the prior information using parameters in the Pathways database, including the quality of the study, the populations served by the intervention, and the intervention services. We recommend basing the priors for the main analysis on studies that are rated as high quality. However, we do not recommend limiting them further (such as by whether the intervention serves particular populations or provides particular services); because all studies in Pathways focus on employment and training programs for low-income populations, we view them as relevant prior evidence for this study. This choice also enables us to use a larger set of evidence as the basis for the priors and will be simple to describe.

Selecting outcome domains and timing of outcome measurement. Pathways categorizes outcomes into four possible categories: (1) benefit receipt, (2) education and training, (3) earnings, and (4) employment. We suggest focusing on earnings because these are confirmatory outcomes in this evaluation. Pathways does not include information on self-regulation skills or economic hardship, the other confirmatory outcomes in this study. Thus, we plan to exclude self-regulation skills and economic hardship from the Bayesian analyses.

Pathways also categorizes the timing of measurement of outcomes as either short-term (18 months or less after participants are first offered services), long-term (between 18 months and 5 years after participants are first offered services), or very long-term (more than 5 years after participants are first offered services). We suggest focusing on categories that align with the timing of our data

collection. For analysis of the first follow-up survey, we will focus on short-term outcomes. For analysis of the second follow-up survey, we will focus on long-term outcomes.

H. Exploratory analysis

In addition to estimating impacts on exploratory outcomes, the exploratory analysis will also include (1) estimating impacts by subgroup, (2) conducting robustness checks, (3) conducting mediation analysis, (4) estimating impacts on those who participated in Work Success as opposed to all those who were assigned to the Work Success group, and (5) estimating the effect of the COVID-19 pandemic on the estimated impacts of Work Success. Additional exploratory analysis may also be conducted as findings emerge.

1. Subgroup analysis

As part of the exploratory analysis, we will check the consistency of impacts on confirmatory outcomes across subgroups. Examining differences in impacts by groups of participants, as identified by their characteristics at study enrollment, can help our understanding of the magnitude of impacts as well as help programs think through whether there is a need to change the services they offer.

Although arguments can be made for why impacts might differ between many groups of participants owing to their characteristics at baseline, there is no hypothesis or expectation that the impacts would differ significantly by subgroup. Thus, all the subgroup analysis is part of the exploratory analysis. We will present findings from the subgroup analysis in the technical appendix, referencing them in the text of the main report as appropriate.

For adequate statistical power, we will not estimate separate impacts for subgroups with fewer than 300 study participants. A sample of 300 study participants would provide a minimum detectable effect size of 0.27 for outcomes based on administrative records and 0.32 for those based on survey reports. Table 6 presents some subgroups we may examine and the data source we would use to identify the subgroups. We may expand the set of subgroups included in the exploratory analysis as needed to investigate patterns in the confirmatory analysis.

2. Robustness checks

We will conduct analytic robustness checks to verify that the findings from our confirmatory analysis are not overly sensitive to specific analytic decisions that we made. We will perform robustness checks by re-running the confirmatory analysis using different specifications. For example, we will compare the findings of our confirmatory analysis from when we apply weights and when we do not, and we will compare the findings of our confirmatory analysis from when we use regression models with and without covariates. We will present findings from these sensitivity analyses in the technical appendix and reference them in the text of the main report as appropriate.

Table 6. Subgroups to include in exploratory analysis

| Subgroup | Data source |
|--|-----------------|
| Demographic and socioeconomic characteristics | |
| Age | Baseline survey |
| Race and ethnicity | Baseline survey |
| Whether has children | Baseline survey |
| Received income from any public assistance or unemployment insurance program | Baseline survey |
| Education, employment, and self-regulation and goal-related skills | |
| Challenges to employment at baseline | Baseline survey |
| Education level at baseline | Baseline survey |
| Goal-setting skills at baseline | Baseline survey |

3. Mediation analysis

As part of the exploratory analysis, we will consider conducting a mediation analysis, which will shed light on the mechanisms through which impacts emerge. We will use a two-step procedure to estimate this decomposition (Heckman and Pinto 2015; Kautz and Zanoni 2022). This analysis will focus on outcomes with statistically significant impacts in the confirmatory analysis. For example, if there were statistically significant impacts on average monthly earnings, we could decompose the overall program impact on earnings into (1) a component attributable to impacts on potentially relevant outcomes such as self-regulation skills, goal-setting skills, and hard skills such as education and training and (2) a component attributable to changes in other, unmeasured variables. In another example, if there were a statistically significant impact on goal-setting and attainment skills, we could decompose the overall program impact into (1) a component attributable to impacts on service-receipt outcomes, such as receipt of one-on-one job assistance related to goal-setting and attainment skills, and (2) a component attributable to changes in other, unmeasured variables.

We will use findings from the confirmatory analysis to determine the specific outcomes to include in the mediation analysis. We will not conduct this analysis if there are no statistically significant impacts in the confirmatory analysis. If conducted, this analysis will include intermediate outcomes for which there are statistically significant impacts.

4. Estimating impacts on participants

Our main impact analysis will compare outcomes for all those assigned to the program group to those assigned to the control group and provide estimates of the intent to treat (ITT) impact. However, policymakers and program administrators are also interested in estimates of the impact of the intervention on those who actually participated in the intervention—the treatment on the treated (TOT) impact. We will consider estimating impacts for those who received program services or certain amounts or types of services if doing so would be useful for interpreting findings from the confirmatory analysis.

5. Analysis of the impacts of COVID-19

It is possible that the COVID-19 pandemic influenced program impacts of Work Success. The direction of this influence is unknown. If the skills taught by Work Success helped program group members weather the changes caused by the pandemic while control group members struggled, the impacts could have been larger than under more typical economic conditions. Alternatively, because of business and school closures and workers being concerned about the health risks of some jobs, the pandemic could have eliminated work opportunities for both program and control group members, leading to smaller program impacts than would be found under more typical conditions.

We will conduct exploratory analysis to assess how program impacts might have been influenced by the pandemic. We will consider an analysis of earnings based on calendar date to assess how the pandemic might have affected impacts over time. The differences between program and control group members in average monthly and/or quarterly earnings by calendar dates over time provide suggestive evidence of how the pandemic affected the impact of the Work Success program on participants' earnings and whether the pattern changed during the pandemic. When interpreting pandemic-related differences in impacts on earnings, we caution that we cannot attribute those differences to causal effects of the pandemic because over time there are differences in the composition of the sample of study participants, which might have contributed to the changes.

I. Interpreting impact estimates

The estimates of the impact of Work Success will be interpreted based on how Work Success was implemented and the context in which it was implemented. For example, the interpretation of the estimates will vary by whether Work Success was implemented with fidelity to its design and the public health issues and economic situation faced by study participants after participating in the program. Both the implementation of Work Success and the context in which it was implemented will be described in a forthcoming report.

References

- Abraham, K.G., J.C. Haltiwanger, K. Sandusky, and J.R. Spletzer. "Exploring Differences in Employment Between Household and Establishment Data." *Journal of Labor Economics*, vol. 31, no. S1, 2013, pp. 129–172.
- Barnow, B.S., and D. Greenberg. "Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments." *Evaluation Review*, vol. 39, no. 2, 2015, pp. 179–228.
- Bettinger, E., and R. Baker. "The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring." NBER Working Paper #16881. National Bureau of Economic Research, 2011. <http://www.nber.org/papers/w16881>.
- Blakemore, A.E., P.L. Burgess, S.A. Low, and R.D. St. Louis. "Employer Tax Evasion in the Unemployment Insurance Program." *Journal of Labor Economics*, vol. 14, no. 2, 1996, pp. 210–230.
- Caliendo, M., D.A. Cobb-Clark, and A. Uhlendorff. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics*, vol. 97, no. 1, 2015, pp. 88–103.
- Cavadel, E.W., J.F. Kauff, M.A. Anderson, S. McConnell, and M. Derr. "Self-Regulation and Goal Attainment: A New Perspective for Employment Programs." OPRE Report #2017-12. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2017.
- Collins, J., and K. Murrell. "Using a Financial Coaching Approach to Help Low-Income Families Achieve Economic Success: Challenges and Opportunities for the Field." Center for Financial Security, 2010. <https://fyi.extension.wisc.edu/financialcoaching/files/2010/07/Using-a-Financial-Coaching-Approach.pdf>.
- Cooper, H.M., L.V. Hedges, and J. Valentine (eds.). *The Handbook of Research Synthesis and Meta-Analysis*, 2nd ed. Russell Sage Foundation, 2009.
- Cox, D.R. *Analysis of Binary Data*. Chapman and Hall/CRC, 1970.
- Deke, J., and M. Finucane. "Moving Beyond Statistical Significance: The BASIE (BAYesian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations." OPRE Report #2019-35. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2019.
- Deke, J., and M. Puma. "Coping with Missing Data in Randomized Controlled Trials." Evaluation Technical Assistance Brief #3. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Adolescent Health, 2013.
- Fletcher, S.J., and C.A. Mullen (eds.). *The SAGE Handbook of Mentoring and Coaching in Education*. Sage Publications, 2012.
- Gelman, A. "What Is the 'True Prior Distribution'? A Hard-Nosed Answer." *Statistical Modeling, Causal Inference, and Social Science* blog, 2016. <http://andrewgelman.com/2016/04/23/what-is-the-true-prior-distribution-a-hard-nosed-answer/>. Accessed June 7, 2018.
- Gelman, A. "Prior Information, Not Prior Belief." *Statistical Modeling, Causal Inference, and Social Science* blog, 2015a. <http://andrewgelman.com/2015/07/15/prior-information-not-prior-belief/>. Accessed June 7, 2018.
- Gelman, A. "The General Problem I Have with Noninformatively-Derived Bayesian Probabilities Is That They Tend to Be Too Strong." *Statistical Modeling, Causal Inference, and Social Science* blog, 2015b. <http://andrewgelman.com/2015/05/01/general-problem-noninformatively-derived-bayesian-probabilities-tend-strong/>. Accessed June 7, 2018.

References

- Gelman, A. "Ethics and Statistics: Ethics and the Statistical Use of Prior Information." *CHANCE*, vol. 25, no. 4, 2012, pp. 52–54.
- Gelman, A. "Induction and Deduction in Bayesian Data Analysis." *Rationality, Markets and Morals*, vol. 2, 2001, pp. 67–78.
- Gelman, A., and C.R. Shalizi. "Philosophy and the Practice of Bayesian Statistics." *British Journal of Mathematical and Statistical Psychology*, vol. 66, no. 1, 2013, pp. 8–38. <https://doi.org/10.1111/j.2044-8317.2011.02037.x>.
- Gelman, A., and D. Weakliem. "Of Beauty, Sex and Power." *American Scientist*, vol. 97, no. 4, 2009, pp. 310–316.
- Gigerenzer, G., and U. Hoffrage. "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats." *Psychological Review*, vol. 102, no. 4, 1995, pp. 684–704.
- Heckman, J., and R. Pinto. "Causal Analysis After Haavelmo." *Econometric Theory*, vol. 31, no. 1, February 2015, pp. 115–151.
- Hedges, L.V. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics*, vol. 6, no. 2, 1981, pp. 107–128.
- Hogan, J., and B. Holland. "Using Theory to Evaluate Personality and Job-Performance Relations: A Socioanalytic Perspective." *Journal of Applied Psychology*, vol. 88, no. 1, 2003, pp. 100–112.
- Jones, R.J., S.A. Woods, and Y.R.F. Guillaume. "The Effectiveness of Workplace Coaching: A Meta-Analysis of Learning and Performance Outcomes from Coaching." *Journal of Occupational and Organizational Psychology*, vol. 89, no. 2, 2015, pp. 249–277.
- Kautz, T., and W. Zanoni. "Measuring and Fostering Non-Cognitive Skills in Adolescence: Evidence from Chicago Public Schools and the OneGoal Program." Working paper, 2022.
- Lim, K., A. Miller, M. Risch, and E. Wilking. "Independent Contractors in the U.S.: New Trends from 15 Years of Administrative Tax Data." U.S. Department of the Treasury, Internal Revenue Service, 2019.
- Locke, E.A., and G.P. Latham. *A Theory of Goal Setting and Task Performance*. Prentice Hall, 1990.
- Martinson, K., K. Gardiner, E. Harvill, and R. Cook. "Implementation and Impact of a Goal-Oriented Coaching Program for Cash Assistance Recipients in Michigan." OPRE Report #2020-73. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2020. <https://www.acf.hhs.gov/opre/report/implementation-and-impact-goal-oriented-coaching-program-cash-assistance-recipients>.
- Mullainathan, S., and E. Shafir. *Scarcity: Why Having Too Little Means So Much*. Henry Holt, 2013.
- Nyhus, E.K., and E. Pons. "The Effects of Personality on Earnings." *Journal of Economic Psychology*, vol. 26, no. 3, 2005, pp. 363–384.
- Orr, L.L. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Sage, 1999.
- Pirbaglou, M., J. Katz, M. Motamed, S. Pludwinski, K. Walker, and P. Ritvo. "Personal Health Coaching as a Type 2 Diabetes Mellitus Self-Management Strategy: A Systemic Review and Meta-Analysis of Randomized Controlled Trials." *American Journal of Health Promotion*, vol. 32, no. 7, 2018, pp. 1613–1626.
- Puma, M.J., R.B. Olsen, S.H. Bell, and C. Price. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE #2009-0049. U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, 2009.

References

- Schochet, P.Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*, vol. 33, no. 6, 2009, pp. 539–567.
- Störmer, S., and R. Fahr. "Individual Determinants of Work Attendance: Evidence on the Role of Personality." *Applied Economics*, vol. 45, no. 19, 2013, pp. 2863–2875.
- Theodos, B., M. Simms, M. Treskon, C. Stacy, R. Brash, D. Emam, R. Daniels, and J. Collazos. "An Evaluation of the Impacts and Implementation Approaches of Financial Coaching Programs." Urban Institute, 2015.
- Zimmerman, B.J., A. Bandura, and M. Martinez-Pons. "Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting." *American Educational Research Journal*, vol. 29, no. 3, 1992, pp. 663–667.

Appendix A. Psychometric Analysis of Outcomes in the Self-Regulation and Goal-Setting Skills Domain

We conducted an analysis to determine how to define measures of self-regulation and goal-setting skills using the responses from questions on the surveys and to assess whether these measures were reliable and valid. We previously identified four measures through a psychometric analysis of the data collected on the baseline survey: (1) goal-setting skills; (2) self-esteem (Rosenberg 1965); (3) emotional control and self-monitoring (Roth et al. 2005); and (4) task monitoring, planning, and initiation (Roth et al. 2005).

This appendix describes a similar analysis we conducted using responses to items in the first follow-up survey. Our analyses proceeded in three steps. First, we posited how to group individual survey items into measures of specific self-regulation skills. In this step, we developed new groupings of items on goals and self-regulation skills that were included in the follow-up surveys but not on the baseline survey. For measures that also appeared on the baseline survey, we used the groupings suggested by our analysis of the baseline survey. Second, we examined the reliability and validity of the measures using data from the first follow-up survey, including those that we previously had examined using data from the baseline survey. Third, informed by the results, we revised the measures slightly and confirmed the reliability and validity of the final, proposed version of the measures.

Step 1: Grouping items into measures

In this step, we focused on grouping items that were on the first follow-up survey but not on the baseline survey. Three measures of self-regulation skills were developed from items common across the baseline and follow-up surveys. Because these three measures were based on existing instruments and performed well using data from the baseline survey, we did not revisit their definitions but did confirm their reliability and validity using data from the follow-up survey. The follow-up survey also included two sets of study-developed items that did not appear on the baseline survey. The new items were designed to capture aspects of goal-related skills and self-regulation skills as demonstrated in the context of employment.

Goal-related skills. In addition to the three goal-related items on the baseline survey, the follow-up survey included five other items related to goals (Table A.1). We considered two ways to group the items into scales: (1) an overall measure of goal-setting and attainment skills and (2) separate measures of goal-setting and goal-attainment skills (see the last column of Table A.1). Because we did not have prior evidence on these items, we tested both options to see which grouping fit the data better. As described in the next section, our analyses supported a single measure of goal-setting and attainment skills.

Self-regulation skills in the employment context. In addition to the one item in the baseline survey designed to capture self-regulation skills (as demonstrated in the context of employment), the follow-up survey included five other items designed to capture those skills (Table A.2). We posited that these six items together would capture a single skill: employment-related self-regulation.

Table A.1. Goal-related items in the first follow-up survey

| # | Item | In baseline survey? | Goal-setting or goal-attainment skills? |
|---|---|---------------------|---|
| 1 | I know I need to get a job or a better job and really think I should work on finding one. | Yes | Goal-setting skills |
| 2 | I set employment goals based on what is important to me or my family. | No | Goal-setting skills |
| 3 | I set long-term employment goals that I hope to achieve (such as finding a job, finding a better job, getting promoted, or enrolling in further education). | Yes | Goal-setting skills |
| 4 | I set specific short-term goals that will allow me to achieve my long-term employment goals. | Yes | Goal-setting skills |
| 5 | Based on everything I know about myself, I believe I can achieve my employment goals. | No | Goal-attainment skills |
| 6 | When I set employment goals, I think about barriers that might get in my way and make specific plans for overcoming those barriers. | No | Goal-attainment skills |
| 7 | Even when I face challenges, I continue to pursue my employment goals. | No | Goal-attainment skills |
| 8 | I keep track of my overall progress toward my long-term employment goals and adjust my plans if needed. | No | Goal-attainment skills |

Table A.2. Items capturing self-regulation skills in the employment context in the first follow-up survey

| # | Item | In baseline survey? |
|---|---|---------------------|
| 1 | Lost your temper with someone other than friends or family | No |
| 2 | Said something that you later regretted to someone other than friends or family | No |
| 3 | Decided not to apply for a job because you didn't think you would get an interview | No |
| 4 | Overcame a barrier that could have prevented you from finding or keeping a job | No |
| 5 | Been late for a job, interview, program meeting, class, or training session | Yes |
| 6 | Missed an appointment related to work, looking for a job, a program, school, or training for a reason other than you were sick or ill | No |

Step 2: Examining the reliability and validity of the measures

To examine whether the candidate measures performed well, we conducted two analyses. First, we estimated Cronbach's alpha for each of the five measures. This is a statistic that provides evidence on internal consistency reliability—that is, the degree to which different items for a given measure produce similar results. Second, we conducted a confirmatory factor analysis that sheds light on aspects of validity—that is, whether the measures capture what they were designed to measure. The confirmatory factor analysis examined (1) factor loadings that capture the extent to which each item relates to the underlying self-regulation skill (related to internal consistency reliability), (2) model fit statistics that summarize whether the groupings of items into measures fit the data well overall (exhibit model validity), and (3) correlations between the pairs of skills that suggest whether separate measures capture different constructs (exhibit discriminant validity). We conducted the analysis twice, once with a single goal-setting and attainment measure, and once with separate measures of goal-setting skills and goal-attainment skills. In both cases, the models also included the four other measures. We restricted the analysis sample to the control group to limit the possible perception that we selected definitions of outcome measures based on the results of the impact analysis. We assessed whether the hypothesized measures met standard criteria for reliability and validity by examining how well corresponding statistics compared to target values (Table A.3). As discussed in Kautz and Moore (2020), we viewed these criteria as guidelines, not strict rules.

Table A.3. Criteria used for assessing reliability and validity

| Type of reliability or validity | Statistic | Target value |
|------------------------------------|---|---|
| Internal consistency reliability | Cronbach's alpha | At least 0.65 (DeVellis 2017) |
| Internal consistency reliability | Factor loading | 0.40 or above (Stevens 2012), particularly in cases when Cronbach's alpha is low and when the sign matches the theoretical relationship between the item and factor |
| Model validity (overall model fit) | Root mean square error of approximation (Steiger and Lind 1980) | 0.05 or below for a close fit and 0.08 or below for a reasonable fit, as Browne and Cudeck (1992) suggested on the basis of practical experience |
| Model validity (overall model fit) | Comparative Fit Index (Bentler 1990a) | 0.90 or above as suggested by Brown (2015) based on analysis by Bentler (1990b) |
| Model validity (overall model fit) | Tucker–Lewis Index (Tucker and Lewis 1973) | 0.90 or above, as suggested by Brown (2015) on the basis of analysis by Bentler (1990b) |
| Discriminant validity | Correlation between factors | Less than 0.80 and they are theoretically distinct (Brown 2015) |

Overall, the initial grouping of items into measures performed well. Across all measures and specifications, Cronbach's alpha was close to or above our minimum target of 0.65, ranging from 0.64 to 0.92. Similarly, all of the model fit statistics met our criteria in both specifications. The analyses had two implications for the final measures:

- 1. Supported a single goal-setting and attainment measure.** We estimated that the correlation between the two separate measures of goal-setting skills and goal-attainment skills was 0.91. This estimate exceeded our cutoff for discriminant validity, suggesting that the two measures captured the same underlying skill. We also conducted a separate analysis based on an exploratory approach that suggested that the goal-setting and attainment items captured a single skill.² Based on this evidence, we suggest using a single goal-setting and attainment measure.
- 2. Suggested removing one item from the employment self-regulation scale.** Our confirmatory factor analysis produced another statistic—called the factor loading—that captured the extent to which each item related to the corresponding self-regulation skill. A factor loading ranges from -1 to 1, and a positive (negative) factor loading indicates that higher values on the item are positively (negatively) associated with the overall measure. The sign of the factor loading (negative or positive) should match the expected relationship with the skill. However, the factor loading had an unexpected sign for one item in the employment self-regulation measure: “Overcome a barrier that could have prevented you from finding or keeping a job.” The factor loading suggested that participants who overcame barriers more frequently had lower self-regulation skills. One possible explanation for this is that people who have higher levels of self-regulation skills experience fewer barriers to begin with, so they also report overcoming fewer barriers. Based on this evidence, we suggest removing the item from the scale.

² In particular, we estimated the number of skills captured by the group of items, using the Kaiser criterion (Kaiser 1960). We found evidence that supported a single skill.

Step 3: Confirming the reliability and validity of the revised measures

After adjusting the measures in the two ways described above, we re-estimated the Cronbach's alpha and the confirmatory factor model. For all scales, Cronbach's alpha met our criteria for internal consistency reliability (Table A.4). Similarly, the overall fit statistics and correlations between skills met our criteria for model validity and discriminant validity (Tables A.5 and A.6). Based on these analyses, we do not suggest additional revisions to the measures.

Table A.4. Reliability of measures of self-regulation skills

| All | Cronbach's alpha | Meets criterion |
|--|------------------|-----------------|
| Goal-setting and attainment skills ^a | 0.85 | Yes |
| Self-esteem ^b | 0.68 | Yes |
| Emotional control and self-monitoring ^c | 0.86 | Yes |
| Task monitoring, planning, and initiation ^c | 0.91 | Yes |
| Employment self-regulation ^d | 0.62 | Yes |

Source: Evaluation of Employment Coaching first follow-up survey.

^a A 0- to 3-point scale based on the extent to which respondents agree with statements that reflect a high level of goal-setting and attainment skills. The scale indicates whether they (0) strongly disagree, (1) disagree, (2) agree, or (3) strongly agree.

^b A 0- to 3-point scale based on the extent to which respondents agree with statements that reflect a high level of self-esteem. The scale indicates whether they (0) strongly disagree, (1) disagree, (2) agree, or (3) strongly agree.

^c A 0- to 2-point scale that indicates whether respondents have problems related to the skill (0) often, (1) sometimes, or (2) never.

^d A 0- to 3-point scale based on the frequency with which respondents exhibit behaviors that reflect a lack of employment self-regulation skills. The scale indicates whether they exhibit specific behaviors (0) a few times a week, (1) a few times a month, (2) about once a month, or (3) hardly ever or never.

Table A.5. Model fit statistics of measures of self-regulation skills

| All | Value | Meets criterion |
|---|----------------|-----------------|
| Root mean square error of approximation | | |
| Estimate | 0.045 | Yes |
| 95 percent confidence interval (lower and upper bounds) | 0.038 to 0.051 | Yes |
| Comparative Fit Index (CFI) | 0.963 | Yes |
| Tucker-Lewis Index (TLI) | 0.960 | Yes |

Source: Evaluation of Employment Coaching first follow-up survey.

Note: The estimates for each sample come from a single confirmatory factor model that assumes five factors that correspond to the five self-regulation skills. The items corresponding to each skill are constrained to relate only to that skill. The factors are not constrained to be independent.

Table A.6. Correlations between measures of self-regulation skills

| Skill 1 | Skill 2 | Correlation | Meets criterion |
|--|--|-------------|-----------------|
| Goal-setting and attainment skills ^a | Self-esteem ^b | 0.52 | Yes |
| Goal-setting and attainment skills ^a | Emotional control and self-monitoring ^c | 0.27 | Yes |
| Goal-setting and attainment skills ^a | Task monitoring, planning, and initiation ^c | 0.42 | Yes |
| Goal-setting and attainment skills ^a | Employment self-regulation ^d | 0.15 | Yes |
| Self-esteem ^b | Emotional control and self-monitoring ^c | 0.59 | Yes |
| Self-esteem ^b | Task monitoring, planning, and initiation ^c | 0.64 | Yes |
| Self-esteem ^b | Employment self-regulation ^d | 0.49 | Yes |
| Emotional control and self-monitoring ^c | Task monitoring, planning, and initiation ^c | 0.74 | Yes |
| Emotional control and self-monitoring ^c | Employment self-regulation ^d | 0.67 | Yes |
| Task monitoring, planning, and initiation ^c | Employment self-regulation ^d | 0.55 | Yes |

Source: Evaluation of Employment Coaching first follow-up survey.

Note: The estimates come from a single confirmatory factor model that assumes five factors that correspond to the five self-regulation skills. The items corresponding to each skill are constrained to load only on that skill. The factors are not constrained to be independent.

^a A 0- to 3-point scale based on the extent to which respondents agree with statements that reflect a high level of goal-setting and attainment skills. The scale indicates whether they (0) strongly disagree, (1) disagree, (2) agree, or (3) strongly agree.

^b A 0- to 3-point scale based on the extent to which respondents agree with statements that reflect a high level of self-esteem. The scale indicates whether they (0) strongly disagree, (1) disagree, (2) agree, or (3) strongly agree.

^c A 0- to 2-point scale that indicates whether respondents have problems related to the skill (0) often, (1) sometimes, or (2) never.

^d A 0- to 3-point scale based on the frequency with which respondents exhibit behaviors that reflect a lack of employment self-regulation skills. The scale indicates whether they exhibit specific behaviors (0) a few times a week, (1) a few times a month, (2) about once a month, or (3) hardly ever or never.

References for Appendix A

- Bentler, P.M. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin*, vol. 107, no. 2, 1990a, pp. 238–246.
- Bentler, P.M. "Fit Indexes, Lagrange Multipliers, Constraint Changes and Incomplete Data in Structural Models." *Multivariate Behavioral Research*, vol. 25, no. 2, 1990b, pp. 163–172.
- Brown, T.A. *Confirmatory Factor Analysis for Applied Research*. Guilford Publications, 2015.
- Browne, M.W., and R. Cudeck. "Alternative Ways of Assessing Model Fit." *Sociological Methods Research*, vol. 21, no. 2, 1992, pp. 230–258.
- DeVellis, R.F. *Scale Development: Theory and Applications*, 4th ed. Sage Publications, 2017.
- Kaiser, H.F. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement*, vol. 20, no. 1, 1960, pp. 141–151.
- Kautz, T., and Q. Moore. "Selecting and Testing Measures of Self-Regulation Skills Among Low-Income Populations." OPRE Report #2020-138. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2020.
- Rosenberg, M. *Society and the Adolescent Self-Image*. Princeton University Press, 1965.
- Roth, R.M., P.K. Isquith, and G.A. Gioia. *Behavior Rating Inventory of Executive Function—Adult Version: Professional Manual*. Psychological Assessment Resources, 2005.
- Steiger, J.H., and J.C. Lind. "Statistically Based Tests for the Number of Common Factors." Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Stevens, J.P. *Applied Multivariate Statistics for the Social Sciences*. Routledge, 2012.
- Tucker, L.R., and C. Lewis. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika*, vol. 38, no. 1, 1973, pp. 1–10.



To learn more, please visit

<https://www.acf.hhs.gov/opre/project/next-generation-enhanced-employment-strategies-project-2018-2023>

Questions? NextGenProject@mathematica-mpr.com