



REPORT

The External Review of Job Corps: Directions for Future Research

March 2018

Joanne Lee
Peter Z. Schochet
Jillian Berk

Submitted to:

U.S. Department of Labor
Chief Evaluation Office
200 Constitution Ave., NW
Washington, DC 20210

Project Officer: Jessica Lohmann
Contract Number: DOLQ129633249/DOL-OPS-16-U-00122

Submitted by:

Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Jillian Berk
Reference Number: 50290.01.440.478.001

This page has been left blank for double-sided copying.

DISCLAIMER

This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office, by Mathematica Policy Research, under contract number DOLQ129633249/DOL-OPS-16-U-00122. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

This page has been left blank for double-sided copying.

CONTENTS

I	INTRODUCTION.....	1
II	OVERVIEW OF EVALUATION PROCESS FOR YOUTH SERVICES INNOVATIONS.....	3
III	EVALUATION DESIGN CONSIDERATIONS FOR STRATEGIES TO ENHANCE THE EXPERIENCES OF YOUTH PARTICIPANTS.....	11
IV	EVALUATION DESIGN CONSIDERATIONS FOR DISCIPLINARY APPROACHES TO ENHANCE THE CLIMATE AND SAFETY OF CENTERS.....	17
V	DESIGN OPTIONS FOR WAYS TO ENHANCE THE EMPLOYMENT SKILLS OF JOB CORPS GRADUATES.....	21
VI	DESIGN OPTIONS FOR CHANGES TO GROUP DYNAMICS IN JOB CORPS CENTERS.....	27
VII	CONCLUSIONS.....	31
	REFERENCES.....	33

This page has been left blank for double-sided copying.

EXHIBITS

II.1	Overview of the evaluation cycle	4
II.2	Overview of intervention selection	5
II.3	Overview of considerations for evaluation design	6
III.1	What strategies can Job Corps pursue to enhance the experiences and growth of participants?	11
III.2	Measuring success for interventions to enhance the experiences and growth of Job Corps participants	16
IV.1	What strategies can Job Corps adopt to enhance the climate and safety of centers?.....	18
IV.2	Measuring success for interventions to enhance the climate and safety of Job Corps centers	20
V.1	How can Job Corps enhance the employment skills and job readiness of graduates?	22
V.2	Measuring success for interventions to enhance the employment skills and job readiness of Job Corps graduates.....	25
VI.1	How can Job Corps change group dynamics in centers to enhance program outcomes for youth?.....	28
VI.2	Measuring success for interventions to change group dynamics in Job Corps centers.....	30

TABLES

II.1	Considerations for evaluation design.....	6
II.2	Example calculations for MDIs for an RCT focusing on education, employment, and training outcomes.....	9
III.1	Key design considerations for example interventions to enhance the experiences and growth of Job Corps participants	13
III.2	Potential data items for matching for a comparison group design.....	15
IV.1	Key design considerations for example interventions to enhance the climate and safety of Job Corps centers	19
V.1	Key design considerations for example interventions to enhance the employment skills and job readiness of Job Corps graduates	24
VI.1	Key design considerations for example interventions to change group dynamics in Job Corps centers.....	30

This page has been left blank for double-sided copying.

I. INTRODUCTION

Job Corps, a program administered by the U.S. Department of Labor’s (DOL) Employment and Training Administration (ETA), is the nation’s largest and most comprehensive residential education and job training program for at-risk youth. Originally established by the Economic Opportunity Act of 1964, the program currently operates under the provisions of the Workforce Innovation and Opportunity Act (WIOA), which Congress enacted in 2014. A rigorous impact evaluation found promising impacts on educational attainment, arrests, and short-term earnings for the full sample of youth Job Corps served, and positive long-term earnings impacts for older youth (Schochet et al. 2008).

DOL has worked for the past several years to make improvements to the Job Corps program, with the goal of better serving youth. Although progress has been made, important issues and questions remain about how best to structure and deliver services in the program. DOL recognizes the need to assess current best practices for serving youth as it considers options for enhancing the Job Corps program. To fulfill this need, DOL’s Chief Evaluation Office (CEO) contracted with Mathematica Policy Research and its subcontractor, Decision Information Resources, Inc., to conduct an external review of the Job Corps program. The goals are to (1) document what is known about Job Corps and other similar programs, (2) identify promising evidence-based practices that Job Corps might consider for the future, and (3) present options for future research and evaluation. The external review covers a broad range of topics that are relevant to the Job Corps program, including program operations and services. It does not include an implementation or impact analysis of the Job Corps program.

Mathematica addresses the first and second goals in an accompanying report, “The External Review of Job Corps: An Evidence Scan Report” (Berk et al. 2018), which we hereafter refer to as the “Evidence scan.” The Evidence scan summarizes the findings from previous research and identifies promising evidence-based practices relevant to Job Corps. The research questions that guide the Evidence scan were informed by discussions with CEO and national Job Corps staff.

Research questions for program improvement

- ❖ What strategies can Job Corps pursue to enhance the experiences and growth of participants?
- ❖ What strategies can Job Corps adopt to enhance the climate and safety of centers?
- ❖ How can Job Corps enhance the employment skills and job readiness of graduates?
- ❖ How can Job Corps change group dynamics in centers to enhance program outcomes for youth?
- ❖ How could changes to program organization and management practices enhance Job Corps?

An expert working group provided information on current practices that other organizations are implementing that the Job Corps program could consider, and the Evidence scan assesses the evidence base for relevant practices. The covered topics are organized into two key domains: (1) provision of youth services and (2) program organizational structures and practices.

This report addresses the third goal of the external review by providing a high-level summary of evaluation design options for the first four research questions covered by the

Evidence scan, which relate to the provision of youth services (see box). The intended audience for this report is people who are knowledgeable about Job Corps, have some familiarity with evaluation methods, and have interest in evaluating whether potential innovations to Job Corps

improve participants' outcomes. To keep the presentation manageable, our focus is on quantitative impact evaluation design considerations, not implementation analysis.

We tailor the discussion of design options to each research question by using examples of interventions that apply the promising practices described in the Evidence scan. We do not present design details for specific evaluations; we provide an overview of key design considerations that could be refined in a technical design plan if DOL selects specific interventions for rigorous evaluation.

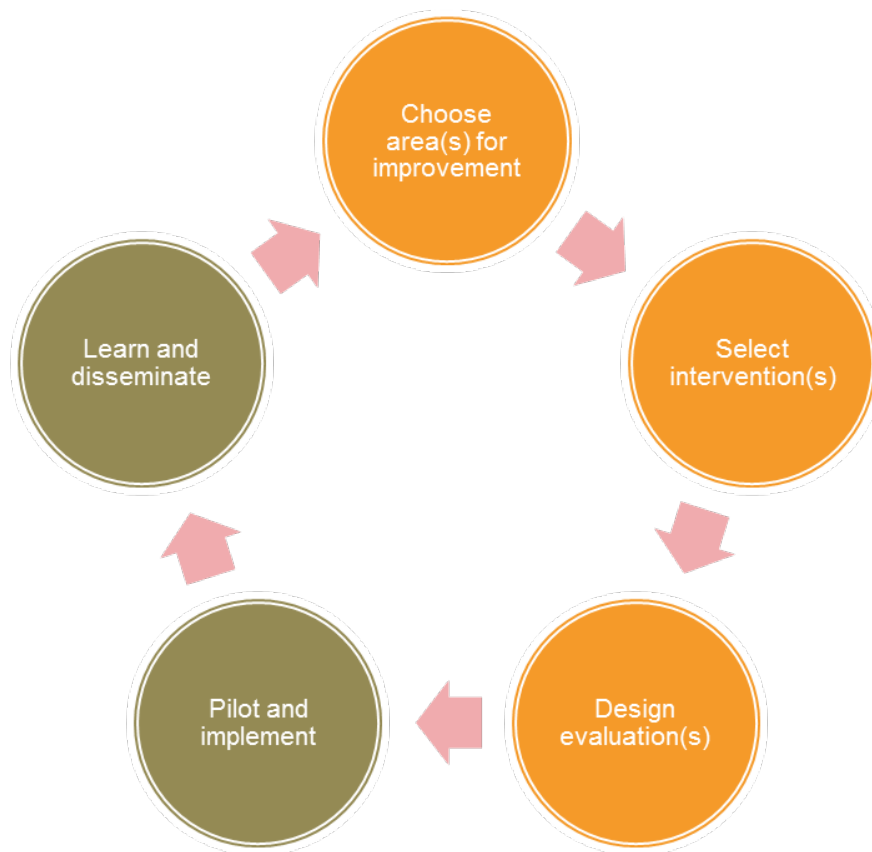
The remainder of the report is in six sections. In the first section, we provide an overview of the evaluation process, including considerations for selecting the research questions and interventions to evaluate. In the subsequent four sections, we present examples of interventions and design options for each research question identified in the Evidence scan. In each section, we start by showing examples of promising interventions and practices. We then describe the considerations for evaluation designs before presenting the potential design option(s) most relevant for these practices. We end each section with a discussion of outcomes and data sources that could be used to identify successful interventions. The final section summarizes design options for future research on Job Corps.

II. OVERVIEW OF EVALUATION PROCESS FOR YOUTH SERVICES INNOVATIONS

Evaluation is a tool that programs such as Job Corps can use to help guide their decisions about the effectiveness of the services that are tested. The basic goal of an impact evaluation is to learn whether an intervention produced outcomes that are better than some benchmark. The intervention can be an individual service or practice or a bundle of them, and common benchmarks are what would have happened in the absence of the intervention (sometimes referred to as "business as usual") or under a different service or practice. For example, the national evaluation of Job Corps assessed whether participating in the Job Corps program (a bundle of services) yielded better outcomes than no participation (Schochet et al. 2006, 2008). The Evidence scan identifies promising youth services innovations, some of which researchers have evaluated rigorously. However, none have been tested in Job Corps' unique context. Through evaluation, DOL can measure whether these innovations improve the Job Corps experiences and outcomes of youth relative to the benchmark condition.

Generally, the evaluation process can be thought of as a cycle of research questions and answers about potential improvements (Exhibit II.1). DOL, implementers, and evaluators can collaborate and make substantial contributions to every step. Starting at the top of Exhibit II.1, the first step of this cycle involves identifying an area for improvement, then, moving clockwise, selecting an intervention within that area for improvement, and designing an evaluation to test it.

Once an intervention or bundle of interventions has been selected, it is valuable to pilot and refine the intervention(s) before full-scale implementation and evaluation. A pilot phase can help identify ways to tailor the intervention to the Job Corps context and improve the chances that the intervention will be implemented with fidelity during the evaluation. This will increase the chances that the evaluation can find positive impacts from full-scale implementation and that the impact findings can be replicated more broadly. Ideally, evaluations should be conducted after there is confidence that the intervention can be implemented well in the Job Corps context so that the outcomes the evaluation measures are a fair test of intervention effects.

Exhibit II.1. Overview of the evaluation cycle

The final step is learning from the findings and disseminating the information. This is vital for continuous program improvement, for assessing which interventions should be implemented and for whom, and for generating future research agendas.

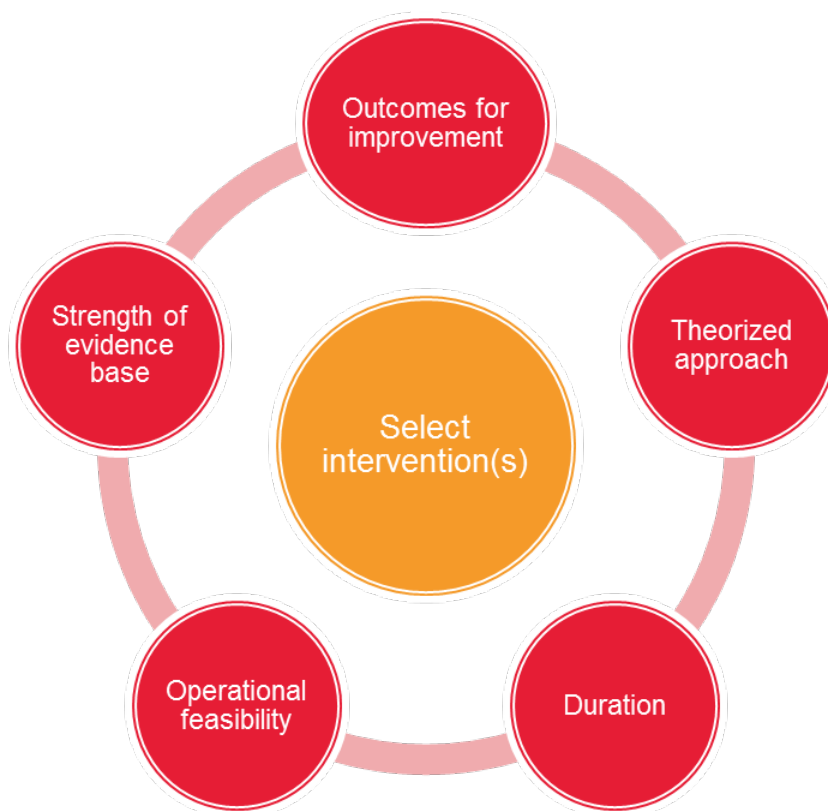
Each cycle adds to an evidence base that can help identify or refine which improvements to pursue next. Although this report focuses on impact evaluations, a formal implementation evaluation is an important complementary tool for interpreting the effects of the intervention. An implementation evaluation can use systematic methods to reveal how the intervention was implemented, the variation in implementation, the challenges experienced, and the lessons learned by centers, staff, and youth. DOL can use both impact and implementation evaluations to build evidence that refines the Job Corps model to best achieve its mission of helping participants become responsible, employable, and productive citizens.

In the rest of this section, we focus on considerations for the first three steps of the cycle (which we highlighted in orange). We describe the second and third steps in greater detail in Exhibits II.2 and II.3. These figures are helpful resources for weighing or prioritizing the different research questions, interventions, and evaluation design options we describe in the rest of this report.

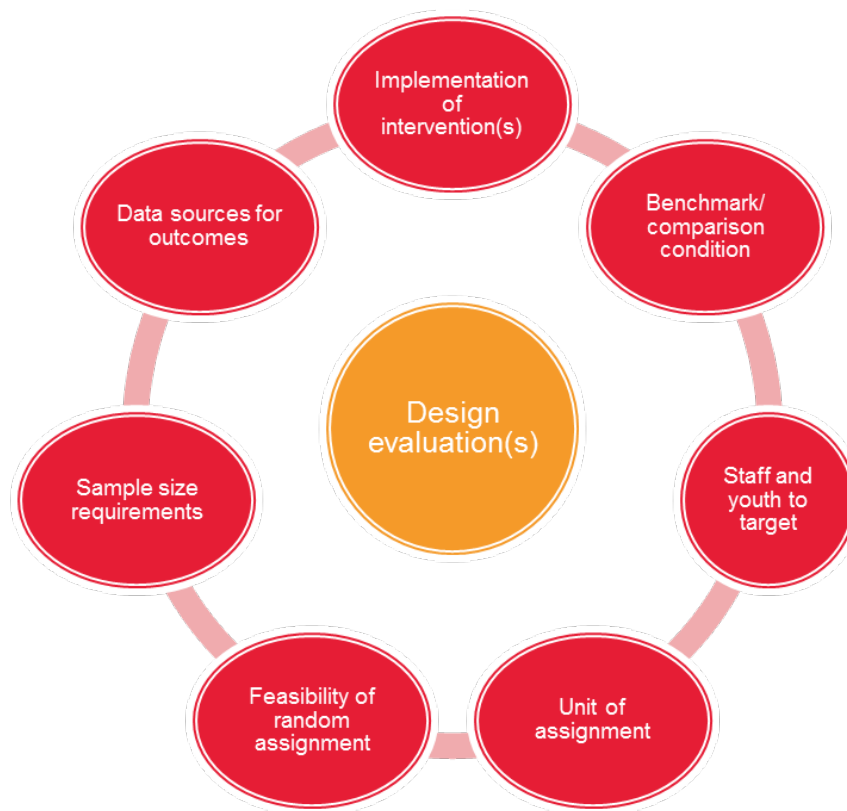
DOL can use the Evidence scan and this report to begin the first step of the evaluation cycle: selecting areas of improvements for the youth services innovations. The areas of improvement

could be further narrowed or prioritized based on feedback from DOL, Job Corps staff, and other stakeholders. After selecting an area for improvement, the next step is to select an intervention within the domain of interest (Exhibit II.2). We have identified the following key, interrelated factors that DOL could consider when selecting an intervention or set of interventions for evaluation. The factors are: (1) the outcome domain or specific outcomes DOL is targeting for improvement, (2) the theory underlying the intervention, (3) the duration and intensity of the intervention, (4) the operational feasibility of the intervention, including the availability of resources necessary to implement the intervention with fidelity, and (5) the strength of the evidence base.

Exhibit II.2. Overview of intervention selection



The selection of the intervention relates to how the intervention will be implemented, and that, in turn, influences the evaluation design options that are possible. DOL could choose to focus on designs with a single intervention or pursue multiple research questions simultaneously, weighing the potential advantage of timing against costs. DOL could also examine the overall effect of a package of services or the effect of a narrower component. Exhibit II.3 shows key intervention implementation and other design considerations affected by these factors. As a reference, we briefly expand on these considerations in Table II.1. Key considerations include the services that the comparison group will receive, the staff and youth to target for the study, the unit of intervention assignment, the feasibility of random assignment, the evaluation's sample size requirements, and the data sources for outcomes. We return to these design considerations in more detail in the next sections, when we discuss design options for the program innovations identified in the Evidence scan.

Exhibit II.3. Overview of considerations for evaluation design**Table II.1. Considerations for evaluation design**

Design consideration	Example questions for reflection
Implementation of intervention(s)	What problem or issue will this intervention solve? How will it improve program operations or youths' experiences and outcomes? How prevalent is this intervention in Job Corps centers? How will the intervention fit into centers' existing service delivery processes? Who delivers the intervention, and what supports are needed? How long will the intervention be delivered, and in what size group? Will it be delivered in person? Is it necessary to tailor the intervention to the staff and youth in Job Corps?
Benchmark/comparison condition	What will be the experiences of staff and youth who do not receive the intervention? Will the contrast involve the status quo for Job Corps, no participation in Job Corps, or another intervention? Is the expected contrast stark enough to generate impacts the evaluation can detect? Which centers, staff, and youth should expect a similar impact with the intervention?
Staff and youth to target	Which staff and youth will the intervention target? How will individuals be identified for the intervention, and what will be the intensity of their involvement? Do they have prior experiences with practices similar to the intervention, prior to their involvement in Job Corps?
Unit of assignment	Should the unit of assignment be the Job Corps center or more disaggregated level, such as the student, staff, class, or entering group? If one unit is assigned to the intervention group and another is assigned to the comparison group, is it likely that they will influence each other's behavior ("spillovers")? Is there a more aggregated unit of assignment (such as the entire center) where spillovers are unlikely?

Design consideration	Example questions for reflection
Feasibility of random assignment	Could the intervention be randomly assigned in the Job Corps context? Will it be necessary to adjust the contracts with center operators? What will be the timing of consent and random assignment to condition? Who will conduct random assignment? How will the fidelity of assignment and compliance to the randomized condition be observed? If random assignment is not possible, is it feasible to construct a comparison group of similar youth or centers not receiving the intervention?
Sample size requirements	For key outcomes, what are reasonable impacts to expect from the intervention(s)? Does relevant data exist to use for statistical power calculations? What are the sample sizes necessary to detect statistically significant impacts of that magnitude?
Data sources for outcomes	Is it possible to measure outcomes with the same level of fidelity for both the intervention and comparison groups? To what extent will data collection procedures vary for the intervention and comparison groups? What are the costs of obtaining the data, and for what time frame? What is the expected level of attrition in the intervention and comparison groups? Is there a plan for analyzing outcomes available in multiple data sources?

In our discussion of evaluation designs, we focus on randomized controlled trials (RCTs) and briefly summarize alternative designs if an RCT is not feasible or to supplement the RCT. Researchers regard RCTs, also referred to as experimental designs, to be the gold standard for impact evaluation. When implemented carefully, random assignment creates balanced groups in terms of their characteristics before the intervention, differing only in their intervention service offerings. Thus, any resulting differences in average outcomes between the research groups are causally attributable to the interventions, and we can be confident that the RCT provides rigorous impact estimates. The details of implementation for the RCT also determine what we can learn from the RCT results. Four crucial aspects of implementation that we touch on here are whether there is excess demand for intervention services (“oversubscription”), site selection, when random assignment will occur, and required sample sizes.

1. *Oversubscription.* A key feature regarding the feasibility and ethics of an RCT is the extent to which there are more individuals who can benefit from the tested interventions than can be served. This requirement is necessary to support the formation of a comparison group. In the Job Corps context, this condition is likely to be met for the evaluations that we are considering because they involve testing specific enhancements in selected settings only, and not nationwide until their efficacy is shown. Furthermore, study resources may only support an evaluation conducted in select centers and/or with a subsample of Job Corps participants and staff.
2. *Site selection.* Findings can be nationally generalizable if sites are chosen randomly and a high percentage of them participate. Conversely, if sites are selected purposively, the impact findings will formally pertain to those sites only, and ancillary study information (for example, results from the implementation study or subgroup analyses) will be necessary to assess potential intervention effects in other centers and settings. An extreme case is if only one site is in the intervention group, which may occur for interventions that involve large structural changes or are very costly. In those cases, the evaluation would not be able to disentangle the effects of the intervention from other factors specific to that site, making it impossible to rigorously generalize the impact estimates beyond the single site.

Once sites have been selected for the evaluation, DOL and the evaluators can work with site staff before finalizing the design to understand sites’ concerns about study participation.

This is crucial to ensure adequate sample sizes, generalizability (if applicable), and site staffs' adherence to evaluation procedures. The evaluation could address some concerns by adding flexibility to the design. For example, different DOL regions could include different proportions of centers to be randomly assigned to the intervention group. The analysis can statistically adjust for differences in the probabilities of random assignment so that they do not weaken the RCT.

3. *Timing of random assignment and selecting student samples.* The ideal timing of random assignment relates to the types of interventions being assessed and the base population targeted for services. In all cases, it is important to conduct random assignment early enough to ensure sufficient time to properly implement the intervention (for example, factoring in time for staff training).

If randomization is at the center level, care must be taken to ensure that the sample of students in the intervention and comparison centers have similar characteristics. For example, if the intervention targets a subset of students, the evaluation team must identify the same subset of students in the comparison group centers. Similarly, care must be taken to ensure that the introduction of the intervention does not change the types of students who enroll in the intervention centers, which could undermine the benefits of the RCT design. Some options for selecting the sample of students are when a certain number of Job Corps participants show up at the center or after youth are assessed using the Test of Adult Basic Education (TABE) or other assessment tools used at both the intervention and comparison centers. For all designs, the evaluation team must establish procedures to collect study consent, contact information for follow-up data collection after Job Corps participation, and any background information that the impact analysis requires that are not available in program data.

4. *Sample size.* Sample size requirements to test an intervention depend on the desired level of precision for impact estimates on key outcomes. Too much or not enough precision will lead to an inefficient use of evaluation resources. The desired level of precision can be based on impact magnitudes that are both realistically attainable and meaningful in the Job Corps context. One advantage of analyzing intermediate outcomes rather than focusing on long-term outcomes only is that we can potentially learn about impacts more quickly and with smaller targeted sample sizes. However, any intervention effects on mediating outcomes need to be interpreted carefully because they may not be as linked with later-term outcomes as the conceptual model might suggest, or based on evidence from related evaluations in other contexts. Some options for how to determine a desired level of precision include using results from previous studies, calculating impacts such that program benefits offset program costs, or identifying a policy-relevant improvement for a center (for instance, the level of improvement a low-performing center must demonstrate to be reclassified as a high-performing center).

A common method for examining required sample sizes for an RCT is to examine “minimum detectable impacts” (MDIs), which are the smallest true impacts that we have a high probability of detecting. The smaller the MDI, the greater the statistical power of the design. Table II.2 provides an example of MDIs calculated for center- or youth-level RCTs examining education, training, and employment outcomes, and compares them to impacts measured from previous studies.

A key observation from this table is that the level of random assignment strongly influences the MDIs: required sample sizes are much larger when assignment is at the center level. This design feature emphasizes the importance of a thorough, flexible plan for site selection and sample enrollment. To illustrate, for a center-level RCT involving 20 centers with 400 youth each (8,000 youth total), we can expect to detect a significant quarterly earnings impact if the true intervention impact were \$930 or more. In contrast, an individual-level RCT with only 2,000 youth split across fewer centers (5 centers with 400 youth each, for example) could detect a significant quarterly earnings impact if the true intervention impact were \$501 or more. Thus, evaluations that randomize students rather than centers will typically require much smaller sample sizes, but this advantage must be balanced against potential “spillover” effects where the intervention influences both intervention and comparison students because of their interactions in the same center.

Table II.2. Example calculations for MDIs for an RCT focusing on education, employment, and training outcomes

Outcome	MDI with individual-level assignment (2,000 youth)	MDI with center-level assignment (20 centers with 400 youth each)	Measured impact from literature
Education and training outcomes			
Received a high school diploma or GED certificate	6.0	11.2	15.1 ^a , 24.1 ^b
Received a vocational, technical, or trade certificate	4.5	8.3	22.3 ^a
Employment and earnings outcomes			
Employed two years after random assignment	6.3	12.0	4.9 ^b
Quarterly earnings two years after random assignment	\$501	\$930	N/A

Note: All measured impacts from the literature were identified as statistically significant in the respective studies. Calculations of MDIs when individuals are randomized are based on an annual intake rate of 2,000 study participants who are evenly assigned to intervention conditions with a sample loss rate of 20 percent. Calculations of MDIs when centers are randomized are based on an annual intake rate of 4,000 study participants across 20 centers that are evenly assigned to intervention conditions, with a youth sample loss rate of 20 percent. Additional parameter assumptions are that baseline characteristics explain 20 percent of the variation in outcomes within and across sites, the intraclass correlation is 0.04, and Type I and Type II errors are set to 0.05 and 0.80, respectively. We assume the rate of receiving a high school diploma or GED for this population is 36 percent; the rate of receiving a vocational, technical, or trade certificate is 15 percent; the rate of being employed is 50 percent; and the standard deviation of quarterly earnings is \$4,000.

^a Impacts are from the National Job Corps Study at 48 months (Schochet et al. 2008) and are based on comparisons between a intervention group who received an offer of admission to Job Corps and a comparison group who did not receive an admission offer.

^b Impacts are from the interim evaluation of the National Guard Youth ChalleNge program at 21 months (Millenky et al. 2010) and are based on comparisons between an intervention group who received an offer of admission to the ChalleNge program and a comparison group who did not receive an offer of admission.

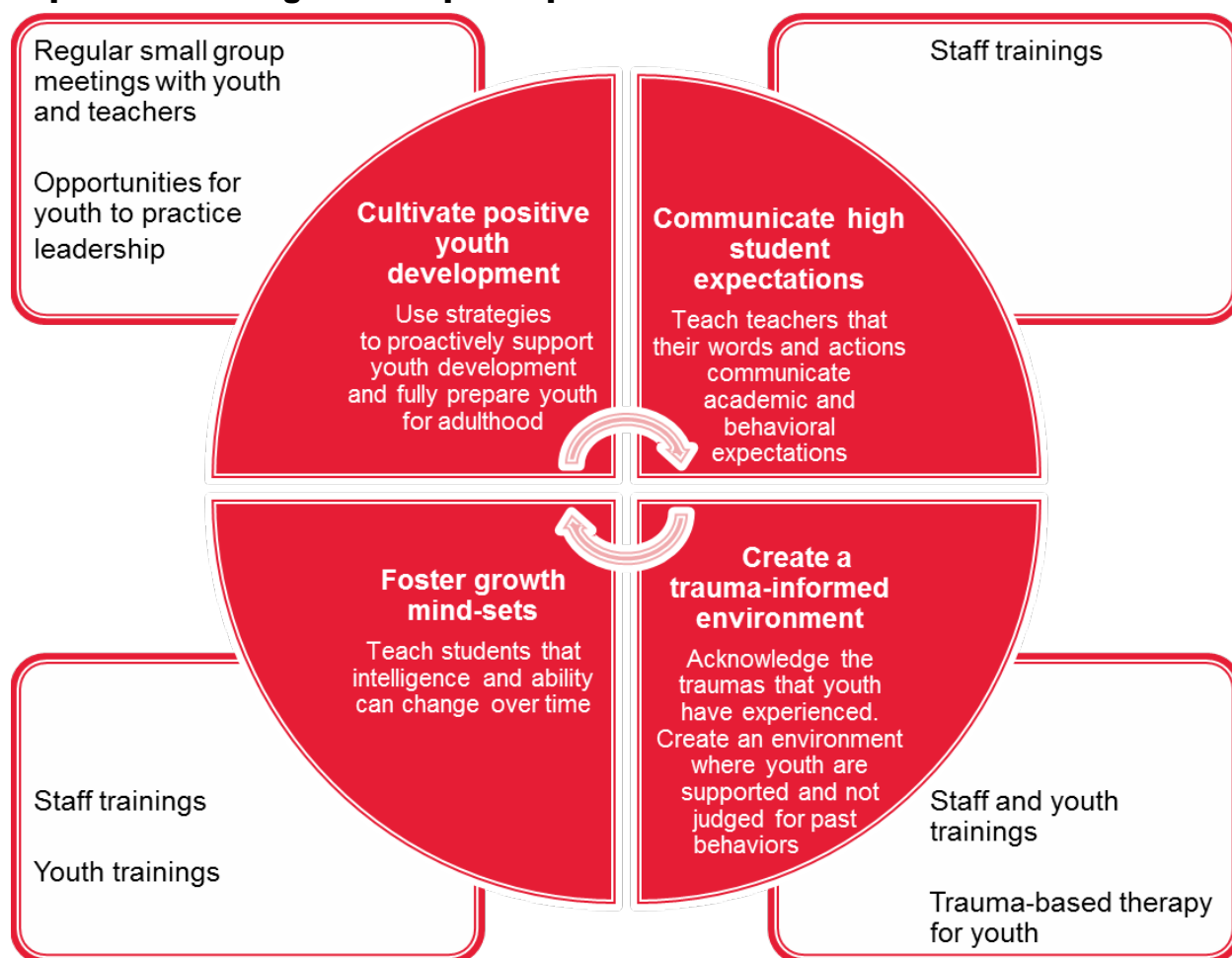
MDI = minimum detectable impact, RCT = randomized controlled trial, N/A = not available.

This page has been left blank for double-sided copying.

III. EVALUATION DESIGN CONSIDERATIONS FOR STRATEGIES TO ENHANCE THE EXPERIENCES OF YOUTH PARTICIPANTS

The Evidence scan identifies four approaches to enhance the experiences and growth of youth participants. These approaches have different underlying theories for how to increase youth engagement, so we present each approach as a quadrant in Exhibit III.1. Examples of interventions that correspond to these approaches are in call-out boxes. In this section, we focus on the design considerations for these examples.

Exhibit III.1. What strategies can Job Corps pursue to enhance the experiences and growth of participants?



Below, we discuss broad design considerations for these interventions using the framework we showed in Exhibit II.3 and Table II.1 above; a summary of the design issues is in Table III.1.

- Implementation of intervention(s).** The interventions vary in terms of which staff and youth would be involved and the intensity of their involvement. In planning implementation, the evaluation team would need to identify how to coordinate any training the staff requires, how and where to deliver services to youth (for example, in a classroom setting, small groups, or individually), when to start the intervention, and what services in existing Job

Corps regimens will be affected. Additional investigation would be required to document which interventions Job Corps centers already have implemented, and how these centers incorporate them into their schedules.

- **Benchmark/comparison condition.** A rigorous evaluation of any of the interventions will require an understanding of the services that the comparison group will receive. This benchmark/comparison condition is crucial for defining the contrast for the evaluation, which could affect the size of the likely impacts and the required sample sizes to obtain precise impacts. The contrast will depend on what business-as-usual services the targeted youth can receive, which might not be under the control of the investigator. For example, to test a new proactive youth development practice, the comparison group might use other proactive youth development practices that Job Corps currently uses, so the evaluation would essentially be testing one intervention relative to another. In other instances, the comparison group might not be using any practices related to the intervention at hand. In multi-armed designs, some contrasts could be controlled by the investigator so that the evaluation would test multiple interventions against each other. In that case, a single study could simultaneously examine various intervention effects. Regardless of the design, we typically expect that having bigger contrasts between the intensity of services the research groups receive will generate bigger impact estimates.
- **Staff and youth to target.** All of the identified interventions to enhance the experiences and growth of youth participants appear appropriate for the full range of Job Corps staff and youth, except for the provision of trauma-based therapy to targeted youth. For these types of interventions, it might be necessary to identify the targeted individuals through a trauma assessment administered to all youth. All individuals in the intervention and comparison groups would have to complete this assessment to identify the samples for the study.
- **Unit of assignment.** If only some individuals (or groups of individuals) within a center were assigned to the intervention group and others to the comparison group, the interactions between the youth could lead to “contamination” of the comparison group and bias the impacts downward. Another consideration is whether assigning individuals to different services within a center could change group dynamics at the center and influence the measured impact estimates in a manner that is inconsistent with the intervention’s underlying theory. This could also lead to biased impact estimates. To avoid bias, the center could be selected as the unit of assignment for nearly all of the example interventions. Individual-level random assignment, however, could be conducted for evaluations that test interventions that directly target individuals, such as youth training sessions or opportunities for youth leadership, where contamination issues might be less severe.
- **Feasibility of random assignment.** All of these interventions could be randomized, but one potential barrier to a random assignment evaluation is if many centers already implement the practices that are to be tested. In such a circumstance, it might not be cost-effective for centers that are already implementing the practices to stop using the practices. In these cases, one approach to potentially increase the number of centers in the study sample would be a design where centers assigned to the comparison group are assigned to temporarily stop offering the services or to offer reduced or other services.
- **Sample size requirements.** Youth outcomes will typically be correlated within centers due to shared center environments and demographic backgrounds. Thus, center-level RCTs

require larger sample sizes than individual-level RCTs to detect a statistically significant effect of the same magnitude on study outcomes (see example calculations in Table II.2). Studies in which random assignment is conducted at the center level but where the analysis directly compares youth outcomes in the intervention and comparison groups will overstate the precision of the impact estimates (Schochet 2008). Statistical methods can be used to adjust the precision of estimates. Statistical power is a bigger concern for analyses of longer-term employment outcomes than intermediate (mediating) outcomes because of the possibility that intervention effects fade over time. Furthermore, for outcomes collected using surveys, response rates could become smaller over time for the longer-term outcomes, further reducing statistical power. Therefore, with center-level random assignment, some evaluations may only have sufficient power to examine impacts on intermediate outcomes if required sample sizes for the longer-term outcomes are prohibitive. Importantly, justification for a design that focuses on intermediate outcomes would require a strong theoretical framework that aligns the mediating outcomes with longer-term outcomes. Additional support for this design strategy could be provided by findings from similar interventions that had positive impacts on both the mediating and longer-term outcomes.

- Data sources for outcomes.** The theory of change for these interventions supports intermediate impacts on youth engagement with staff and the Job Corps program. Administrative data can capture some outcomes related to engagement with Job Corps, such as retention and academic achievement, but programs might not collect data on youth perceptions of staff and the program or staff attitudes and behaviors targeted by interventions. A survey of youth or staff could fill these gaps by providing reliable engagement measures, such as normed scales with established psychometric properties. Findings of beneficial impacts on the mediating outcomes could justify a longer-term follow-up of the sample using administrative earnings records (for example, Unemployment Insurance (UI) records or National Directory of New Hires (NDNH) data) or using surveys (if study resources permit).

Table III.1. Key design considerations for example interventions to enhance the experiences and growth of Job Corps participants

Theorized approach	Example intervention	Targeted staff or youth	Random assignment feasible?	Unit of assignment
Cultivating positive youth development	Regular small group meetings with youth and teachers	All	Y	Center
	Opportunities for youth to practice leadership	All	Y	Center or youth
Communicating high student expectations	Staff trainings	All	Y	Center
Fostering growth mind-sets	Staff trainings	All	Y	Center
	Youth trainings	All	Y	Youth
Creating a trauma-informed environment	Staff and youth trainings	All	Y	Center
	Trauma-based therapy for youth	Trauma-affected youth	Y	Center or youth

Based on these design considerations, it appears that an RCT with center-level randomization will typically be the design of choice. The evaluation team would need to develop more detailed RCT design plans tailored to the specific interventions DOL selects. As we described in Section II, this plan would entail delineating strategies for site selection, the timing and unit of random assignment, determining required sample sizes, monitoring adherence to study procedures by site staff, and data collection. If a survey or other non-administrative data needs to be collected from sites, the design plan should describe how the evaluation team will measure outcomes with the same level of fidelity for both the intervention and comparison groups. The evaluation team could, for instance, plan more intensive data collection assistance for comparison sites or implement a differential incentive structure to help boost response rates in the comparison group.

Alternatives or supplements to an RCT. As evaluators are assessing the feasibility of an RCT and selecting sites, they might find that the timing is right for evaluation but the process of randomization is not feasible. For example, for a study with individual-level randomization, centers may not expect enough youth to enroll during the study intake period to support random assignment. As another example, there may be instances when program staff want full control of who receives the intervention (for instance, for new interventions that target the most needy students or centers). In such cases where a rigorous RCT is not feasible but a comparison group could be formed by other means, evaluators could use a quasi-experimental design (QED) as an alternative to an RCT or as a supplement to an RCT to help improve precision and provide corroborating evidence of intervention effects. The evaluation team would need to adjust their design plans, including a discussion of their preferred QED and its advantages and limitations relative to other ones. Importantly, there may not be time and cost savings from conducting a QED versus an RCT in terms of study implementation, data collection, and the increased complexity of the analytic methods.

One possible QED that the evaluation team could use is a matched comparison group design. This design would compare youth or centers in the intervention group to other, similar youth or centers not using the intervention. Matching centers and/or individuals in the intervention and comparison groups on observable characteristics that correlate with key outcomes will narrow the set of potential explanations for systematic differences in outcomes (although biases could still remain due to unobservable factors).

Matching is likely to be a feasible evaluation option in the Job Corps context because of the rich program data collected by Job Corps in the Student Pay Allotment and Management Information System (SPAMIS) used for program performance and other reasons. These data can be supplemented with publicly available data on the geographic locations of sites (see examples in Table III.2). A detailed evaluation plan for a matched comparison group design would describe the process used to select the matching variables, the statistical methods used for matching (for example, propensity score methods), and the procedures to assess that the matching created balanced groups. In general, matched comparison groups based on administrative data can have larger sample sizes than RCTs, leading to more precise estimates. However, the estimates may be biased because of unobservable factors that differ across the intervention and comparison groups (for example, motivation for signing up for the intervention), and the estimates may also be more susceptible to modeling assumptions used for matching and impact estimation. Consequently, evaluators across a wide range of fields tend to

have less confidence that the estimates from a matched comparison group design are causal, relative to estimates from an RCT.

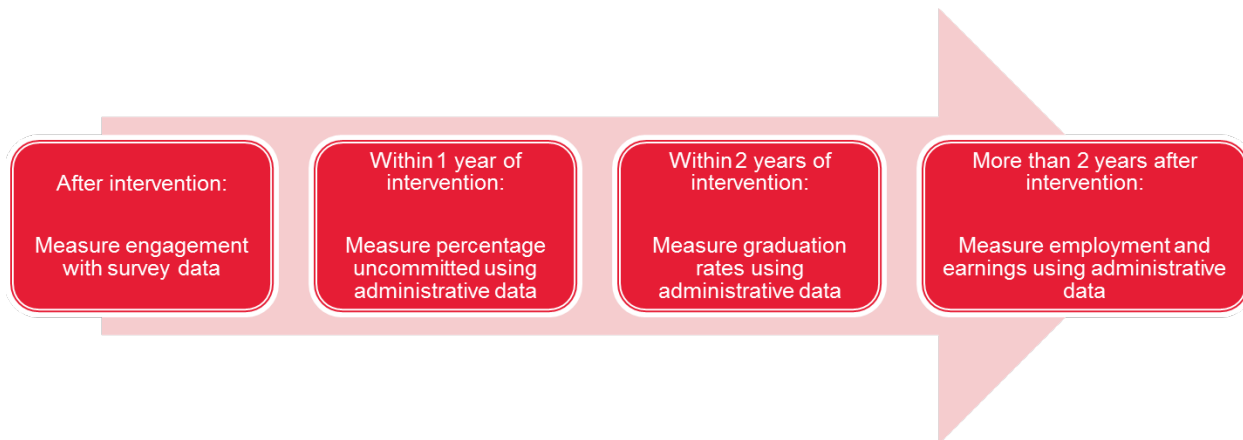
Table III.2. Potential data items for matching for a comparison group design

Characteristics in Job Corps SPAMIS data for Program Year 2016	Local area data from publicly available sources
Percentage female Age distribution Race distribution Percentage ever arrested or charged with a delinquency Percentage of residents in Primary Metropolitan Statistical Area (PMSA or MSA) Distribution of scores on the TABE Vocational completion rates GED attainment rates High school diploma attainment rate Average literacy gain Average numeracy gain Post-program placement rate into a job or the military Distribution of infractions Center size Center contractor/operator size (U.S. Department of Agriculture, large contractor, or small contractor) Number of years the contractor has operated the center	Employment and wages by industry (Census of Employment and Wages; www.bls.gov/cew) Unemployment rates (Local Area Unemployment Statistics; www.bls.gov/lau) Urbanicity measure (Missouri Census Data Center; https://census.missouri.edu/) Poverty rates (Small Area Income and Poverty Estimates; https://www.census.gov/programs-surveys/saipe.html)

SPAMIS = Student Pay Allotment and Management Information System, TABE = Test for Adult Basic Education.

In certain cases, it may also be feasible to consider alternative designs, such as a regression discontinuity (RD) design. RD designs may be appropriate when assignment to the intervention is determined by a cutoff score. For example, if Job Corps were to use a quantitative trauma assessment tool to determine which youth would be good candidates for trauma-based therapy, it might be possible to evaluate the impact of the therapy by comparing the outcomes of youth with assessment scores just above the cutoff to youth with scores just below the cutoff. This design would require Job Corps staff to use the assessment tool, accurately score youth, and make intervention assignments based on the scores. Youth should also not have an incentive or ability to change their scores in order to influence their assignment status. In general, RD designs can generate rigorous estimates for the group of youth with scores near the cutoff. However, this would not provide direct information about the average causal impact for all students, which would be estimated by an RCT. As a result, when testing interventions that are intended to become broadly available to all Job Corps youth, the causal estimates from an RD design may be less policy-relevant than the estimates from an RCT. Further, RD designs require about 3-4 times the sample to achieve the same statistical power as an RCT (Schochet 2009).

Exhibit III.2. Measuring success for interventions to enhance the experiences and growth of Job Corps participants



Measurement of outcomes. For either RCT or quasi-experimental designs, the most comprehensive evaluation design would assess impacts on outcomes using both administrative and survey data. The evaluation team could use administrative data to examine short- and longer-term impacts on measures such as the rate of uncommitted youth, the graduation rate, and employment and earnings. If study resources allowed for conducting a survey, a primary survey-based outcome could be youths' self-reported engagement. Using measures similar to those collected in past evaluations of similar interventions could help provide perspective on the impact findings. If a survey were conducted, it could focus on measures that are not available in administrative data and that are motivated by the theory of change. A sample timeline for measuring outcomes is in Exhibit III.2. DOL could consider making the later administrative data analyses optional if short-term impacts along the causal pathway are not observed.

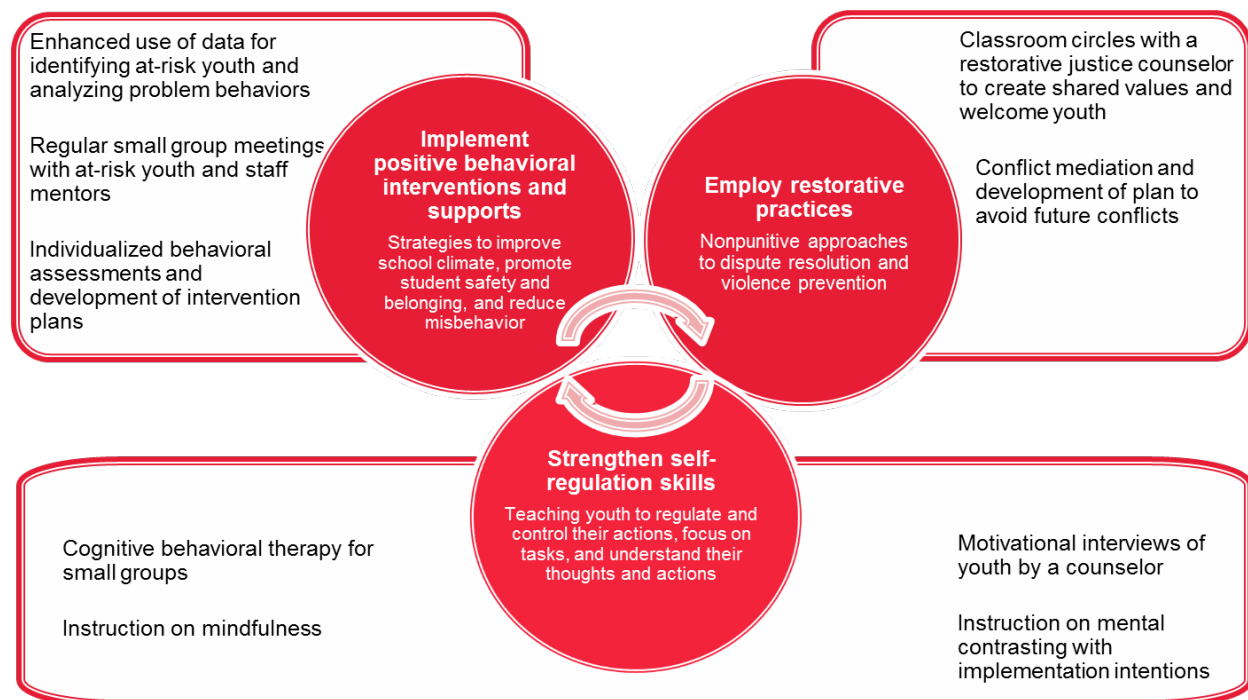
IV. EVALUATION DESIGN CONSIDERATIONS FOR DISCIPLINARY APPROACHES TO ENHANCE THE CLIMATE AND SAFETY OF CENTERS

The Evidence scan identifies three approaches to enhancing the climate and safety of Job Corps centers. These approaches build on theories that adolescents are more influenced by peers and act more impulsively than adults, so they aim to change community climates and/or increase self-control. The approaches and examples of interventions in this group are listed in Exhibit IV.1.

This section focuses on the design considerations for these example interventions using Exhibit II.3 and Table II.1 for structure. We focus on key design considerations with less detail than in Section III to avoid repetition. We summarize key issues in Table IV.1.

- **Implementation of intervention(s).** Nearly all of the interventions in this group focus on delivering services to youth, most often in small groups. The interventions vary in terms of which staff and youth would be involved and the intensity of staff training or cooperation with other experts. The enhanced use of data is an intervention that may require additional technical staff training relative to the other interventions, and could be delivered to youth at an individual level. Additional investigation would need to document which interventions Job Corps centers already have implemented. DOL would also have to ensure that the planned intervention and comparison conditions comply with congressional mandates for existing disciplinary policies.
- **Benchmark/comparison condition.** The benchmark/comparison condition could be business as usual, another intervention, or another disciplinary approach. Because Job Corps has residential living rules and offers conflict resolution training and other disciplinary-related services, the business-as-usual contrast would involve a comparison of the intervention to existing disciplinary approaches (that could differ somewhat by center). Careful documentation of the comparison condition will be important for interpreting the impact estimates.
- **Staff and youth to target.** Five of the nine example interventions apply to all Job Corps staff and youth as potential participants. The remaining four emphasize the development of relationships with staff or counselors and goals for youth at risk of problem behaviors. These targeted interventions vary in terms of which staff members would work with the youth and whether the staff would deliver the intervention to individuals or small groups. However, staff and youth not participating directly in the targeted interventions could still benefit if the interventions, as theorized, improve center-level safety and climate. DOL could target youth for interventions based on SPAMIS data or on a separate assessment tailored to the intervention. If random assignment were at the center level, the evaluation team would need to use the same assessments to target youth in both the intervention and comparison centers.
- **Unit of assignment.** Because the goal of the interventions is to increase group safety, the preferred unit of assignment for these interventions is the center. If just some individuals within a center were assigned to this intervention and reduced their misbehavior, their interactions with other individuals who did not receive it would likely contaminate the comparison group. This would bias the impacts downward.

Exhibit IV.1. What strategies can Job Corps adopt to enhance the climate and safety of centers?



- Feasibility of random assignment.** Random assignment appears to be feasible for these interventions. There is likely to be excess demand by centers for the identified safety-related enhancements, and DOL could select some of these centers for inclusion in the evaluation. One potential barrier is if these practices are already commonly implemented by sites, and it does not make sense to randomly assign sites that already implement the intervention to the comparison group. If this issue is pervasive, a quasi-experimental design might be necessary.
- Sample size requirements.** Because the unit of assignment would likely be at the center level, sample size requirements will be greater than if the unit of assignment were at the youth level (see Table II.2 and the discussion in Section III). Interventions that target subgroups instead of all Job Corps participants would also require larger numbers of centers to participate in the study in order to detect a statistically significant effect of a given size. One way to increase statistical power is to initially focus on intermediate outcomes such as measures of center safety, appropriate behavior, and climate. Improvements in these outcomes could lead to improved in-program experiences of students and staff, and the theories of change underlying these interventions suggest they could correlate with students' post-program employment measures. Statistical power in both center- and individual-level RCTs could also be increased by controlling for the baseline level of the outcome in the analysis if these data are available.
- Data sources for outcomes.** The theory of change for these interventions supports intermediate impacts on safety, appropriate behavior, and center climate. Each type of intervention relates to different aspects of climate. For example, positive behavioral

interventions could lead to a more supportive center climate, and self-regulation interventions could lead to a more mindful climate. Safety and behavior could be measured with program data, but because incidents might be underreported at some centers, an evaluation could have inaccurate findings if it relied solely on program data regarding incidents and infractions (Office of Inspector General 2015). A more comprehensive approach could include a survey of youth and staff to gather detailed information on perceived safety and climate. Longer-term outcomes could be collected to examine the extent to which impacts on the mediating outcomes translated into longer-term earnings gains. Examining impacts on long-term measures would be ideal inasmuch as the Evidence scan finds that few rigorous studies have examined effects of the climate and safety interventions on academic or employment outcomes.

Table IV.1. Key design considerations for example interventions to enhance the climate and safety of Job Corps centers

Theorized approach	Example intervention	Targeted staff or youth	Random assignment feasible?	Unit of assignment
Positive behavioral interventions and supports	Enhanced use of data to identify youth at risk of problem behaviors and analyze patterns within centers	All	Y	Center
	Regular small group meetings with youth at risk of problem behaviors and staff mentors	Youth at risk of problem behaviors	Y	Center
	Develop individualized behavioral assessments and intervention plans	Youth with chronic behavioral issues	Y	Center
Restorative practices	Classroom circles with a restorative justice counselor to create shared values and welcome youth	All	Y	Center
	Conflict mediation and development of plan to avoid future conflicts	All or youth with Level II infractions	Y	Center
Building self-regulation skills	Cognitive behavioral therapy for small groups	All	Y	Center
	Instruction on mindfulness	All	Y	Center
	Motivational interviews of youth by a counselor	All or based on interests and skills	Y	Center
	Instruction on mental contrasting with implementation intentions	All	Y	Center

Based on these design considerations, DOL could consider pursuing center-level RCTs for these interventions, weighing the factors we described in Section II. Specifically, working closely with the sites will be crucial because of the higher sample sizes center-level RCTs require, and because these interventions generally require sustained involvement by staff over time.

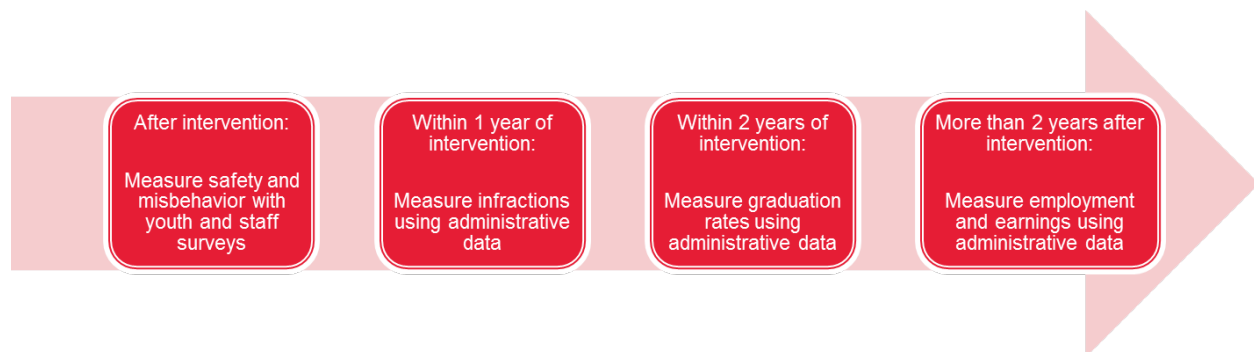
Alternative or supplement to an RCT. A matched comparison group design could serve as an alternative or a supplement to an RCT. A matched comparison group design would compare centers in the intervention group to other, similar centers that do not use the intervention (see Table III.2 for a potential list of matching variables using administrative Job Corps data). Strong candidates for matching variables would be center-level characteristics that correlate with the

perceived safety of study participants, such as the center’s baseline distribution of types of infractions and center size.

Measurement of outcomes. The effectiveness of the safety-related interventions could be measured using outcomes from both administrative and survey data. Administrative data could be used to examine short- and longer-term impacts on measures such as the graduation rate, criminal justice involvement, employment, and earnings. Primary survey-based measures could focus on intermediate outcomes such as perceptions by youth and staff of safety, appropriate behavior, and center climate. Primary outcomes based on administrative data could include the number and type of infractions recorded. Analyses of these outcomes should have a plan for interpreting differences in the levels of incidents reported by youth, staff, and in administrative data. Certain binary outcomes that are important to program staff and youth may not be suitable as key outcomes for evaluation because of very low or high incidence, which can lead to unstable impact estimates. For example, the rate of extremely violent events is an important descriptive measure, but such events are so rare that a few incidents could have outside influence on an impact estimate.

Exhibit IV.2 is a sample timeline for collecting data on the proposed primary outcomes. As with all of the evaluation design options we discuss in this report, DOL could consider making the later administrative data analyses optional if short-term impacts are not observed.

Exhibit IV.2. Measuring success for interventions to enhance the climate and safety of Job Corps centers



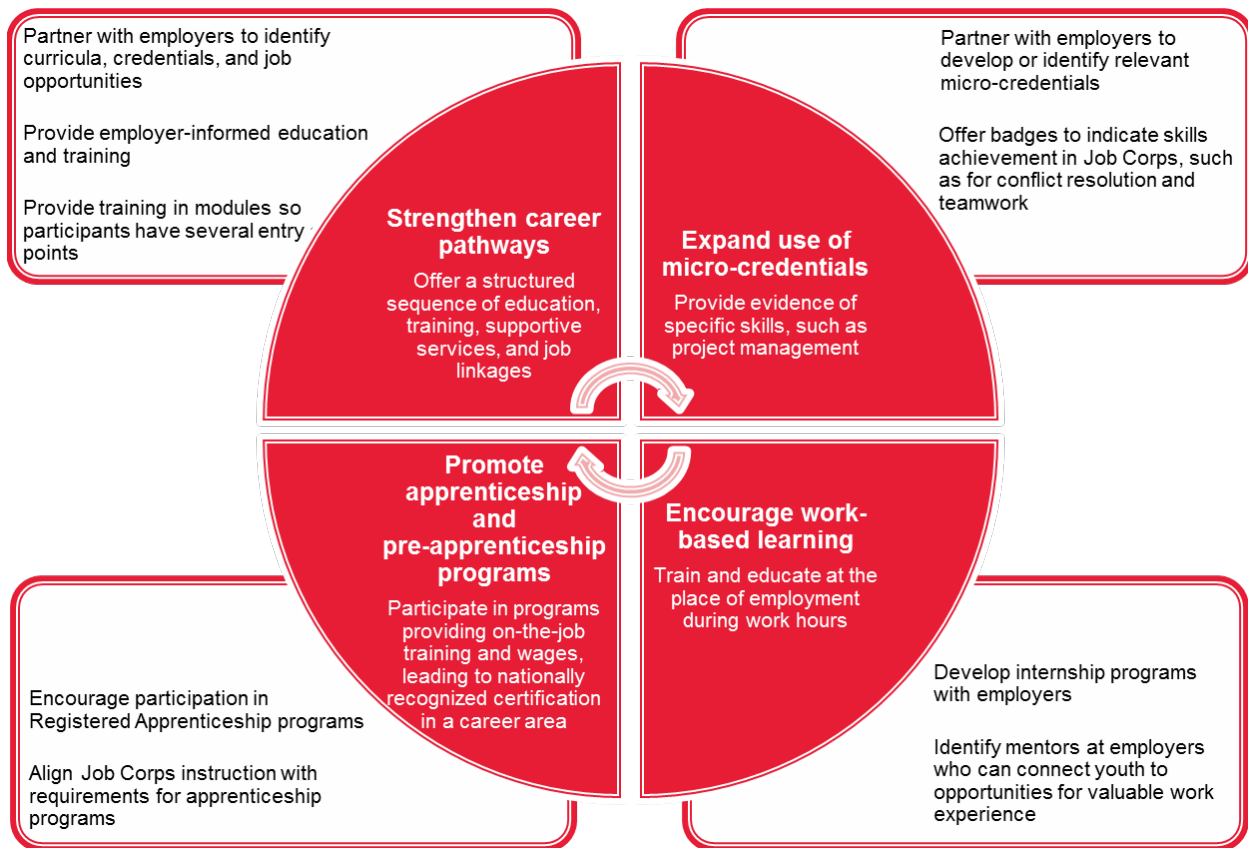
V. DESIGN OPTIONS FOR WAYS TO ENHANCE THE EMPLOYMENT SKILLS OF JOB CORPS GRADUATES

The Evidence scan identifies four approaches that could potentially be integrated more deeply into Job Corps' program model to enhance the employment skills of participants. These approaches, as with the existing model of Job Corps, aim to provide participants with skills that meet the rapidly evolving needs of businesses. We show the identified approaches and examples of interventions that use them in Exhibit V.1. Because Job Corps already integrates aspects of these approaches into its program model, some interventions can be viewed as intensifying existing career preparation services.

This section discusses design considerations for these interventions using Exhibit II.3 and Table II.1 as a guide. We provide less detail than in Section III to avoid repetition. We summarize key issues in Table V.1.

- **Implementation of intervention(s).** The example interventions build on collaborations between Job Corps staff and employers, between youth and employers, and between Job Corps and other programs that provide career training. Some interventions may be enhancements to existing practices at centers rather than new interventions, which would likely require input from employers and the development of new Job Corps-employer partnerships. Additional investigation would be necessary to identify which centers currently implement these practices, and the feasibility of developing enhancements within the time frame of the evaluation. For example, centers and their contractors might need varying levels of technical assistance to establish a credential program with local employers or to build new employer partnerships.
- **Benchmark/comparison condition.** The benchmark condition could be business as usual or another intervention. The Job Corps model already includes direct career preparation services and support services, so careful documentation of the comparison condition at each center will be important for interpreting the impact estimates.

Exhibit V.1. How can Job Corps enhance the employment skills and job readiness of graduates?



- Staff and youth to target.** The example interventions could include all staff and youth as potential participants, although participation in training programs and internships could be more limited in practice, depending on youth interest, youth skill sets, or limited capacity. Programs with limited capacity could consider targeting students based on having relevant career interests and skills. Job Corps could identify these students using the TABE or other assessments. Interventions that encourage staff to develop relationships with employers could also target specific Job Corps staff with relevant experience or in particular vocational trades.
- Unit of assignment.** The unit of assignment for the example interventions could be the center or the youth. Center-level RCTs minimize the potential for youth in the intervention group to influence the comparison group and the potential for the evaluation to change group dynamics in a way that is not consistent with the intervention's underlying theory. However, they also require much larger sample sizes. The weighing of these tradeoffs will depend on the particular intervention. For example, if DOL were to decide to test the effectiveness of interventions aimed at encouraging participation in Registered Apprenticeship programs, then Job Corps could randomly assign centers to either intensively promote or not promote Registered Apprenticeship to youth. Centers in the intervention group could have mass announcements, assemblies, open workshops, and other public reminders. However, if Job Corps were to pursue individual-level random assignment

for this intervention, it might be difficult for centers to maintain the desired contrast between the intervention and comparison groups, and this might reduce the potential intensity of the intervention. For instance, public, open reminders to enroll in Registered Apprenticeship programs might not be feasible. Group dynamics might also change at the center if youth learn that only a random subset of them are being targeted for a Registered Apprenticeship slot. The resulting impact estimates could then capture both the contamination bias and any influence of the changed group dynamics.

There are other examples, however, where individual-level random assignment would be preferred. For instance, if the intervention were to involve participation in an off-center internship program, random assignment at the individual level rather than at the center level could yield more precise impact estimates with little risk of contamination. Even though it is possible that the internship experiences of the intervention group could influence the outcomes of comparison group members in the same center, the risk seems small relative to the large precision and potential cost advantages of conducting random assignment within the center.

- **Feasibility of random assignment.** We see no clear hurdles to random assignment to rigorously test these enhancements. Demand by youth and center staff for these services would likely outpace supply (for example, for internships or mentors). As with the other interventions we previously discussed, the design phase of an evaluation would need to determine the extent to which centers use similar interventions, and which sites are most appropriate for an RCT testing the effectiveness of that intervention or a separate enhancement.
- **Sample size requirements.** As we discussed in Section II and illustrated in Table II.2, the sample size requirements will be much lower for youth-level RCTs than for center-level RCTs. Clearly, lower sample sizes will reduce study costs in terms of sample selection, intervention implementation, and data collection. However, youth-level randomization may not be feasible if the interventions are designed to affect the entire center (for example, if micro-credentials are used to enhance the center curriculum for all youth).
- **Data sources for outcomes.** The theory of change for these interventions supports intermediate impacts on credential attainment, work experience, and short-term earnings. Such outcomes are likely to be documented in program data during and after Job Corps enrollment, with minimal sample loss over time. Longer-term earnings data could be collected using administrative records data or surveys if the impact results on the intermediate outcomes are promising.

Table V.1. Key design considerations for example interventions to enhance the employment skills and job readiness of Job Corps graduates

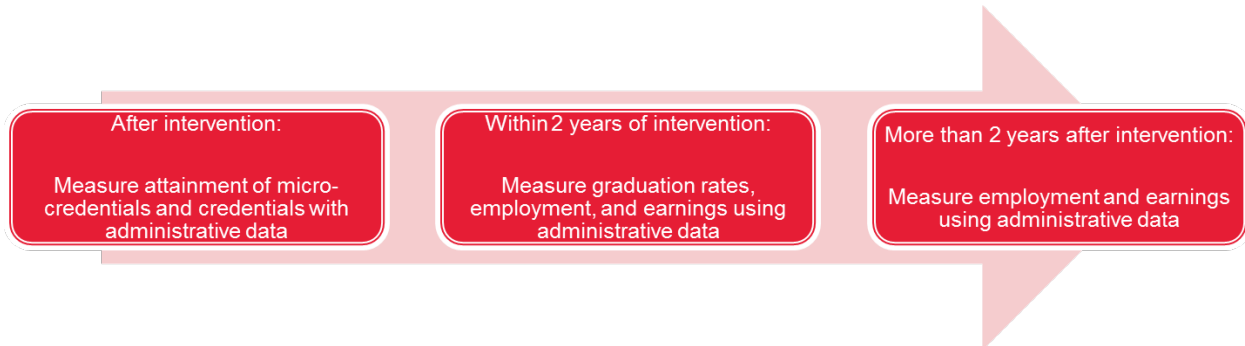
Theorized approach	Example intervention	Targeted staff or youth	Random assignment feasible?	Unit of assignment
Strengthen career pathways	Partner with employers to identify curricula, credentials, and job opportunities	All or based on interests and skills	Y	Youth or center
	Provide employer-informed education and training	All or based on interests and skills	Y	Youth or center
	Provide training in modules so participants have several entry and exit points	All or based on interests and skills	Y	Youth or center
Expand use of micro-credentials	Partner with employers to develop or identify relevant micro-credentials	All or based on staff experience	Y	Center
	Offer badges to indicate skills achievement in Job Corps, such as for conflict resolution and teamwork	All	Y	Youth or center
Encourage work-based learning	Develop internship programs with employers	All or based on interests and skills	Y	Youth or center
	Identify mentors at employers who can connect youth to opportunities for valuable work experience	All or based on interests and skills	Y	Youth or center
Promote apprenticeship and pre-apprenticeship programs	Encourage participation in Registered Apprenticeship programs	All or based on interests and skills	Y	Youth or center
	Registered Apprenticeship programs	All or based on interests and skills	Y	Youth or center

Based on these design considerations, youth- or center-level RCTs appear to be feasible, depending on the nature of the tested interventions. Relative to the other groups of interventions we discussed in previous sections, these have greater possibilities for individual-level random assignment due to lower risks of contamination. At the same time, the details of the implementation for some of the interventions might require center-wide changes (such as instructional changes and credentialing to align with industry standards), which might make it difficult to vary the interventions within a center. In addition, implementation of the interventions might take some time (for example, building program-employer partnerships), which would delay a rigorous intervention until centers can achieve implementation fidelity.

Alternative or supplement to an RCT. As with the other interventions we discussed earlier, quasi-experimental methods can be used to either supplement or replace an RCT if it is not feasible (for example, because not enough centers participate or many are already using the interventions to be tested). The matching variables in Table III.2 can be used to match centers or youth receiving the intervention to those who are not. In addition, an RD design could be used if Job Corps staff were to use a quantitative skills assessment tool to determine which youth would be good candidates for an internship or program with limited capacity. In that case, it may be possible to evaluate the impact of the intervention by comparing the outcomes of youth with assessment scores just above the cutoff to youth with scores just below the cutoff.

Measurement of outcomes. The primary evaluation outcomes can be obtained from administrative data. An evaluation could use administrative data to examine short- and longer-term impacts on outcomes such as credential attainment, employment, and earnings. Exhibit V.2 is a sample timeline for collecting data on the proposed primary outcomes. As with all of the evaluation design options we presented in this report, DOL could consider making the later administrative data analyses optional if short-term impacts are not observed.

Exhibit V.2. Measuring success for interventions to enhance the employment skills and job readiness of Job Corps graduates



This page has been left blank for double-sided copying.

VI. DESIGN OPTIONS FOR CHANGES TO GROUP DYNAMICS IN JOB CORPS CENTERS

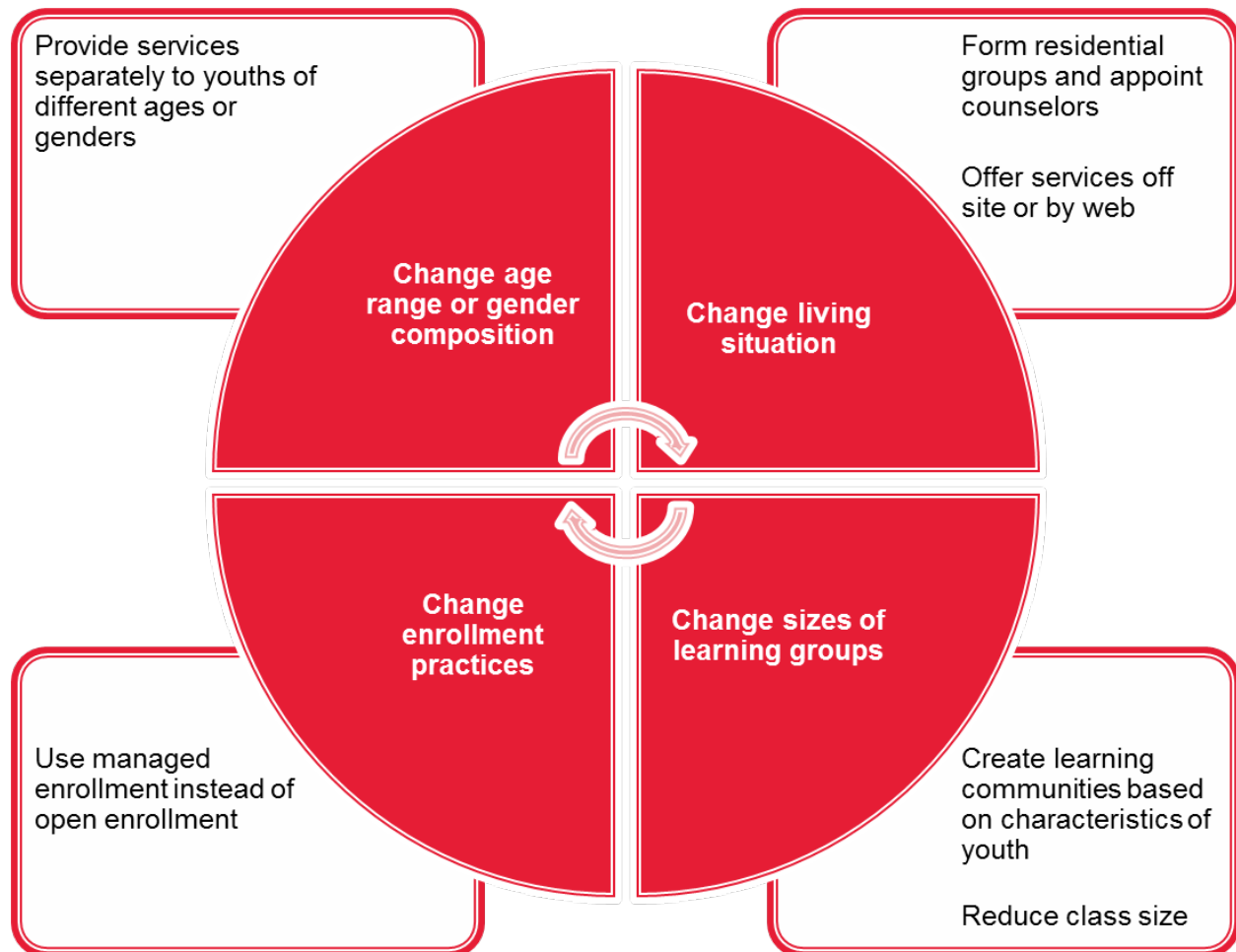
The Evidence scan identifies five features of the Job Corps program that contribute to group dynamics at centers: age of participants, gender of participants, whether youth live on site, group size, and enrollment practices. The scan summarizes the benefits and drawbacks of different group structures that Job Corps could consider using to improve participants' outcomes. We list some examples of interventions in Exhibit VI.1.

Next, we consider key design issues for these interventions using Exhibit II.3 and Table II.1 to guide the discussion. Table VI.1 provides a summary:

- **Implementation of intervention(s).** Rather than offering new services like interventions in previous sections, the example interventions here aim to make existing services more effective by changing the peer groups for Job Corps youth. Center staff could deliver these interventions by, for example, teaching core classes to younger and older youth separately, or by establishing a new center that focuses on a single target group. The DOL Cascades pilot is a recent example in which DOL established a new center that applies multiple interventions, including establishing cohort-based learning communities, using managed enrollment, and using a career pathways model.

The required structural changes to implement some of these interventions poses evaluation design challenges in terms of obtaining a sufficient number of centers that can be included in the intervention group within study resources.

Exhibit VI.1. How can Job Corps change group dynamics in centers to enhance program outcomes for youth?



- Benchmark/comparison condition.** The benchmark/comparison condition for these interventions could be business as usual (for example, providing combined services to all age groups) or another intervention. If individuals are assigned to the intervention and comparison groups within centers, careful documentation of how the research groups interact would be important to demonstrate that the comparison group's dynamics did not influence the intervention group's dynamics and vice versa.
- Staff and youth to target.** The example interventions could include a subset or all youth at a center as potential participants. Offering more services off site or by web might be targeted to youth suitable for the nonresidential program component who face difficulties in moving to their local Job Corps center, such as young mothers (some of whom Job Corps already allows to live off site) or youth with personal or family health issues. Job Corps could also consider linking youth and staff who share key characteristics. For example, staff who have relevant backgrounds could mentor residential groups formed by career interests, or if services are delivered separately to youths by gender, staff of the same gender could lead the classes.

- **Unit of assignment.** Because the aim of these interventions is to change group dynamics, the preferred unit of assignment is the center, but youth-level assignment may be possible in some situations. Youth-level assignment could be possible for interventions that can be delivered to a subset of the target youth, such as making some classes single-sex, expanding access to services by offering them off site or by web, or reducing class sizes. For instance, recruiters or center staff could randomly offer enrolled youth within a center the opportunity to participate in an online program for a portion of the Job Corps curriculum. Job Corps could randomize students to core academic classes of different sizes. However, assigning individuals to intervention or comparison groups within a center could influence group dynamics for all youth at the center. This would bias the impacts downward.
- **Feasibility of random assignment.** Center-level RCTs of interventions that require new centers would be difficult to conduct because of the time and financial resources involved. However, center- or youth-level RCTs of other interventions such as managed enrollment could be implemented.
- **Sample size requirements.** As Table II.2 illustrates, sample size requirements would be significantly higher for center-level RCTs than youth-level RCTs. The ongoing evaluation of DOL's Cascades pilot could help provide guidance on reasonable effect magnitudes and sample loss for these interventions in the Job Corps context. For interventions that require large structural changes or new centers, it might be infeasible to reach the sample sizes required to detect reasonable impacts with center-level RCTs. In situations where a youth-level RCT is not reasonable due to spillovers, DOL could consider randomizing applicants to centers because that will create similar groups of youth in the intervention and comparison centers. This is most likely to be feasible if the evaluation team can identify study centers that are in the same general areas and share Outreach and Admissions (OA) contractors. This will allow youth and OA contractors to continue relationships with centers in their regions as usual. A difficult problem with this design, however, is if the intervention group only contains a few centers, it would be difficult to disentangle the effects of the intervention from the effects of the centers that implement the intervention. For instance, suppose the intervention group consists of a single center. In that case, students could do better in that center in part because of the intervention and in part because the center might have a history of hiring better staff and working with more effective placement contractors than the comparison centers. Having as many centers in the intervention group as feasible helps alleviate this problem.
- **Data sources for outcomes.** The theory of change for these interventions supports intermediate impacts on retention and graduation. Such outcomes are likely to be documented in program data. Longer-term impacts can be obtained using UI or NDNH data or from a survey.

Table VI.1. Key design considerations for example interventions to change group dynamics in Job Corps centers

Theorized approach	Example intervention	Targeted staff or youth	Random assignment feasible?	Unit of assignment
Change age range or gender composition	Provide services separately to youths of different ages or genders	All or youth and staff who share characteristics	Y	Center
Change living situation	Form residential groups and appoint counselors	All	Y	Center
	Offer services off site or by web	All or youth facing challenges in moving into a center	Y	Center or youth
Change sizes of learning groups	Create learning communities based on characteristics of youth	All	Y	Center
	Reduce class size	All	Y	Center or youth
Change enrollment practices	Use managed enrollment instead of open enrollment	All	Y	Center

Alternative or supplement to an RCT. A matched comparison group design is an attractive option for this research area because analyses could be conducted with administrative data for a much larger sample. DOL could also consider pairing a small RCT in a pilot with a matched comparison group design that includes the RCT participants. Strong candidates for matching variables would be center- and individual-level characteristics that could predict academic and employment outcomes, such as the center’s performance measures and high school diploma attainment prior to center enrollment (see Table III.3 for additional candidates).

Measurement of outcomes. Administrative data could be used to examine short- and longer-term impacts on primary outcomes such as the program graduation rate, employment, and earnings. Exhibit VI.2 displays a sample timeline for collecting data on the proposed primary outcomes. As with all evaluation design options we discuss in this report, DOL could consider making the later administrative data analyses optional if short-term impacts are not observed.

Exhibit VI.2. Measuring success for interventions to change group dynamics in Job Corps centers

VII. CONCLUSIONS

This report focused on high-level design considerations for evaluating the impacts of a wide range of innovations that could be implemented in the Job Corps context based on key research questions identified from the separate Evidence scan report (Berk et al., 2018). Clearly, the first step in the research process is for DOL to choose areas of improvement in Job Corps and to select interventions that align with these areas. Candidate interventions could be identified based on key factors such as the outcomes that the interventions aim to improve, the strength of the evidence underlying the interventions, and the operational feasibility of implementation, including resources and the timeline.

The next step in the research process is to assess whether the timing is right for evaluation for the selected interventions. The intervention's implementation and context are key. First, the findings from an evaluation are more likely to be reliable and policy relevant if the intervention can be implemented with fidelity. Second, the evaluation is more likely to detect impacts if there is a sufficient contrast between the intervention and the services for comparison. Comparison services could be those currently offered by the program, or another intervention. A pilot phase prior to an evaluation can assist with both of these factors. During a pilot phase, DOL can identify and address potential implementation challenges, and DOL can assess whether there are refinements to the intervention and comparison services that could increase the contrast. Interventions with more complicated implementation, such as those requiring coordination across agencies or opening new centers, are most likely to benefit from a pilot phase prior to evaluation.

Clearly, the specific design for an evaluation will depend on the study context and should be carefully spelled out in study design documents. However, our analysis suggests several overarching evaluation design themes across the considered research questions. First, a randomized controlled trial (RCT)—the gold standard evaluation design—appears to be feasible for most innovations. In RCTs, individuals, groups of individuals, or entire centers are assigned randomly to either receive the intervention (or different interventions) or comparison services. This randomization process ensures balanced research groups in terms of their characteristics and service needs, yielding more rigorous estimates of program effects than other design options.

The Job Corps context is conducive to randomization for targeted innovations because the program is large, suggesting that more Job Corps participants could likely benefit from the innovations than could be served based on limited study resources. The feasibility of random assignment will ultimately need to be assessed separately for each evaluation, taking into account the details of the intervention. This investigation will involve documenting which related interventions are already used in Job Corps centers, the demand for intervention services, and implementation challenges for the intervention during the evaluation period—for example, whether large structural changes will be required and how long services will need to be provided to participants before outcomes can be measured.

If random assignment is not feasible or yields a design with insufficient sample sizes, the design effort should identify other potential alternatives to supplement or replace an RCT. A matched comparison group design is one option that would compare youth or centers in the intervention group to other, similar youth or centers not using the intervention. Matching is likely

to be feasible because of the rich program data collected by Job Corps in the Student Pay Allotment and Management Information System (SPAMIS). These data could also be linked to geographic data to help account for pre-existing differences between the intervention and comparison groups. Another potential design option is a regression discontinuity (RD) design where the intervention would be assigned based on a quantitative scoring rule, for example, a quantitative trauma assessment tool to determine which youth would be good candidates for trauma-based therapy based on a preset cutoff score.

A second common question for any RCT to test the success of the potential innovations is whether randomization should occur at the individual or center level. This critical decision will influence the nature of the contrast between the intervention and comparison group members and the sample sizes needed for the study to detect impacts. In RCTs where individuals are randomized, intervention and comparison group members may be able to interact with each other. If the intervention influences comparison group members through the individuals' interactions and changes group dynamics that underlie the intervention's theory, then the contrast between the groups could be lessened. These factors would bias the impact estimate towards zero. These issues would likely be less of a problem in designs where centers are randomized rather than individuals. However, center-level randomization requires much larger samples to achieve impact findings with the same level of precision. Thus, such RCTs could require selecting a large number of centers, which could be costly. This suggests that for cost reasons, it could be a better option to randomize individuals within centers even if there are moderate spillover effects. Weighing these tradeoffs will be a critical part of the design phase of the study.

The final consideration we found to be a common thread across interventions is the value of analyzing short-term outcomes to determine whether to examine long-term outcomes. Focusing initially on short-term outcomes along the causal pathway specified by a theory of change could yield policy-relevant impacts more quickly and cheaply than waiting until longer-term outcomes become available, and could require smaller targeted sample sizes. However, any effects on mediating outcomes will need to be interpreted carefully because they may not be as linked with later-term outcomes as the conceptual model might suggest or based on evidence from evaluations in other contexts. Nonetheless, beneficial impact findings on mediating outcomes could justify the costs of obtaining longer-term employment-related outcomes using administrative data sources such as the National Directory of New Hires data or survey data.

REFERENCES

- Clearinghouse for Labor Evaluation and Research. “Causal Evidence Guidelines, Version 2.1.” Washington, DC: Department of Labor, 2015.
- Millenky, Megan, Dan Bloom, and Colleen Dillon. “Making the Transition: Interim Results of the National Guard Youth ChalleNGe Evaluation.” Interim report. MDRC: New York, NY, 2010.
- Office of Inspector General-Office of Audit. “Job Corps Needs to Improve Enforcement and Oversight of Student Disciplinary Policies to Better Protect Students and Staff at Centers.” Report to the Employment and Training Administration. Report No. 26-15-001-03-370. Washington, DC: U.S. Department of Labor, February 27, 2015.
- Berk, Jillian, Linda Rosenberg, Lindsay Cattell, Johanna Lacoé, Lindsay Fox, Myley Dang, and Elizabeth Brown. “The External Review of Job Corps: An Evidence Scan Report.” Report submitted to the U.S. Department of Labor. Washington, DC: Mathematica Policy Research, March 2018.
- Schochet, Peter, and John Burghardt. “Do Job Corps Performance Measures Track Program Impacts?” *Journal of Policy Analysis and Management*, vol. 27, no. 3, 2008, pp. 556–576.
- Schochet, Peter, John Burghardt, and Steven Glazerman. “National Job Corps Study: The Impacts of Job Corps on Participants’ Employment and Related Outcomes.” Princeton, NJ: Mathematica Policy Research, 2001.
- Schochet, Peter, John Burghardt, and Sheena McConnell. “Does Job Corps Work? Impact Findings from the National Job Corps Study.” *American Economic Review*, vol. 98, no. 5, 2008, pp. 1864–1886.
- Schochet, Peter, John Burghardt, and Sheena McConnell. “National Job Corps Study and Longer-Term Follow-Up Study: Impact and Benefit-Cost Findings Using Survey and Summary Earnings Records Data.” Princeton, NJ: Mathematica Policy Research, 2006.
- Schochet, Peter. “Statistical Power for Random Assignment Evaluations of Education Programs.” *Journal of Educational and Behavioral Statistics*, vol. 33, no. 1, 2008, pp. 62–87.
- Schochet, Peter. “Statistical Power for Regression Discontinuity Designs in Education Evaluations.” *Journal of Educational and Behavioral Statistics*, vol. 34, no. 2, 2009, pp. 238–266.

This page has been left blank for double-sided copying.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

**Improving public well-being by conducting high quality,
objective research and data collection**

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■
SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

M50

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.