



# Local Evaluations of the Cohort 1 P3 Pilots: A Synthesis of Their Findings

## Final Report

**January 3, 2020**

Nan Maxwell and Armando Yañez

---

**Submitted to:**

U.S. Department of Labor  
Chief Evaluation Office  
200 Constitution Avenue, NW  
Washington, DC 20210  
Project Officer: Deborah Martierrez  
Contract Number: DOLQ129633249/DOL-  
OPS-15-U-00147

**Submitted by:**

Mathematica  
505 14th Street, Suite 800  
Oakland, CA 94612  
Telephone: (510) 830-3700  
Facsimile: (510) 830-3701  
Project Director: Linda Rosenberg  
Reference Number: 50152

**Disclaimer**

**This report was prepared for the U.S. Department of Labor (DOL) Chief Evaluation Office by Mathematica, under contract number DOLQ129633249/DOL-OPS-15-U-00147.**

**The views expressed are those of the authors and should not be attributed to DOL.**

**Mention of trade names, commercial products, or organizations does not imply endorsement of same by the U.S. Government.**

## Acknowledgments

The authors acknowledge several people who contributed to and supported this research. Guiding and overseeing the project, Christina Yancey and Deborah Martierrez in the Chief Evaluation Office at the U.S. Department of Labor provided advice and support that greatly enhanced our ability to execute the study and this report.

From Mathematica, Linda Rosenberg (the P3 Project Director) and Cay Bradley (the P3 Evaluation Technical Assistance Lead) provided leadership and guidance during all aspects of the study and feedback on report drafts. Jean Knab reviewed the report, offering comments that greatly enhanced its quality. Lindsay Cattell, Elias Hanno, and Robert Santillano supported the pilots and their local evaluators with technical assistance; Sharon Clark helped prepare the report; and Maura Butler provided editorial assistance.

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

## Contents

Abstract.....	ix
I. Introduction .....	1
A. Performance Partnership Pilots for Disconnected Youth (P3) and evaluations of their interventions.....	2
B. Synthesis of local evaluation reports .....	3
II. Interventions and Outcomes Studied .....	7
III. Interventions Achieving Expected Impacts .....	9
A. Level of rigor in the local evaluations.....	9
B. Interventions achieving expected education and employment impacts.....	10
C. Interventions achieving expected impacts in other domains .....	13
IV. Implementation Challenges and Successes .....	15
References .....	19
Appendix A Description of P3 Pilots .....	A.1
Appendix B Data Tables .....	B.1
Appendix C Determining Evidence Ratings.....	C.1

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

## Tables

III.1	Impacts on primary education and employment outcomes, by intervention.....	11
III.2	Achieved expected impacts on secondary outcomes, by intervention .....	12
B.1	Characteristics of pilot interventions .....	B.4
B.2	Characteristics of local evaluations.....	B.6
B.3	Establishing baseline equivalence .....	B.7
B.4	Defining and measuring outcomes .....	B.8
B.5	Clearly defining the intervention.....	B.10
B.6	Characteristics of RCTs .....	B.13
B.7	Baseline equivalence for analytic sample .....	B.15
B.8A	Significance of findings for interventions with high or moderate levels of rigor: Case management interventions .....	B.17
B.8B	Significance of findings for interventions with high or moderate levels of rigor: Other interventions .....	B.20

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



## Abstract

Youth often face complex, multifaceted challenges in transitioning to self-sufficient, productive adults and might need support in this process. Federal efforts to support youth are fragmented across programs and agencies, which has led to calls from state and local providers of youth services for a more efficient and integrated system to serve disconnected youth. As a result, the U.S. Congress authorized the Performance Partnership Pilots for Disconnected Youth (P3), first in the Consolidated Appropriations Act of 2014 and in subsequent annual appropriations. P3 allows its grantee organizations and their partners to waive regulations, such as spending and performance requirements, and to blend and braid money across authorized discretionary federal grant programs to better serve youth. In 2015, each of the nine initial grantees, also referred to as Cohort 1 pilots, included the opportunity for an independent, local evaluation to test the pilot interventions. In addition, the U.S. Department of Labor’s Chief Evaluation Office contracted with Mathematica and its partner Social Policy Research Associates to conduct a national evaluation of P3 to better understand the outcomes and system changes achieved.

This report—one component of the national evaluation—synthesizes findings from the local evaluations of the Cohort 1 pilots. This report assesses the extent to which the local evaluations established a causal impact between the studied intervention and participant outcomes and, for interventions that had such evidence, whether the evidence indicated the intervention had improved outcomes for youth. Research in this report answered three key questions: (1) What interventions and outcomes are the focus of the local evaluations? (2) What is the level of rigor in the local evaluations? and (3) Do the local evaluations find the expected impacts for their interventions with youth?

The synthesis showed that eight of the nine Cohort 1 pilots had local evaluations supporting causal evidence on how one or more aspects of their interventions affected education, employment, and other outcomes. Together, these evaluations covered six types of interventions: (1) case management (only), (2) case management and soft skill training, (3) case management and Workforce Innovation and Opportunity Act (WIOA) services, (4) leadership training, (5) the Teen Outreach Program® (TOP®), and (6) a two-generation education and training program for young parents and their children. All interventions were assessed by one local evaluation, with the exception of case management (only) which was assessed by three local evaluations.

The evaluations found that three of the six types of interventions demonstrated evidence of improving expected youth outcomes. Case management and WIOA services increased the probability of completing General Education Development (GED) tests and achieving readiness for college coursework. The two-generation education and training program showed evidence of increasing the probability that children attended a child care center and that their parents enrolled in the Temporary Assistance for Needy Families program. Case management (only) reduced school suspensions and increased course completion, return to high school, employability skills, employment, families’ participation in their children’s education, and service receipt from a partner. However, two of the three local evaluations examining case management (only) also found evidence of negative outcomes (reduced family member referrals to adult education and participation in career preparation and subsidized employment).

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

## I. Introduction

The federal effort to support disconnected youth,<sup>1</sup> who face complex, multifaceted challenges (Corcoran et al. 2012; Dion et al. 2013), is fragmented across programs and agencies. Such fragmentation can produce coordination problems for organizations. For example, youth might not receive a continuum of services because case managers are required to coordinate services across and within education, employment, and juvenile justice sectors and systems do not exist for such for communicating, coordinating, and monitoring across the sectors.

Because of coordination difficulties across program areas, improving education and employment outcomes of disconnected youth requires that programs innovate, implement evidence-based practices with fidelity, and better collaborate within and across program areas (Dion et al. 2013). The Performance Partnership Pilots for Disconnected Youth (P3) represents a coordinated effort within and across federal agencies to address the programmatic and administrative barriers that state, local, and tribal communities face in meeting the needs of disconnected youth. P3 grantees—with lead agencies as state, local, or tribal government entities, and their partners—were permitted to blend or braid money across federal discretionary programs, using approved waivers from regulations (such as spending and performance requirements).<sup>2</sup> Agencies authorized in the Consolidated Appropriations Act of 2014 to enter into such agreements included the Departments of Education, Labor, and Health and Human Services, along with the Corporation for National and Community Service and the Institute of Museum and Library Services. As authorized under that act, the federal partners selected applicants in 2015 (called Cohort 1 pilots) to implement outcomes-focused strategies to help young people develop the skills, knowledge, and behaviors needed to transition into and be self-sufficient in adulthood. (Appendix A describes the Cohort 1 pilots.) Congress has subsequently authorized up to 10 additional pilots in each subsequent fiscal year.

Each of the nine Cohort 1 pilots was awarded up to \$700,000 to support program implementation, partnerships, and local evaluations, and eight of the nine received supplemental funds ranging from \$48,000 to \$175,000 to further support these activities. The diverse set of Cohort 1 pilots was spread across the country (California, Florida, Illinois, Indiana, Kentucky, Louisiana, Oklahoma, Texas, and Washington) and across types of organizations (four were workforce development boards, three were local governments, one was a school system, and one was a tribal community). The settings in which pilots offered services also varied. For example, government agency locations included child welfare offices and YouthSource Centers funded by the Workforce Innovation and Opportunity Act (WIOA). Educational settings included middle and high schools, community colleges, and community learning centers. One pilot offered services at the public housing building where participating youth lived. Together, the pilots served a variety of low-income disconnected youth ages 14 to 24, including those in

---

<sup>1</sup> In this report, disconnected youth are defined as individuals ages 14 to 24 who are homeless, in foster care, involved in the justice system, or are neither employed nor enrolled in an educational institution (see <https://youth.gov/youth-topics/reconnecting-youth/performance-partnership-pilots/shared-goal-statement>). Such youth are also described as “at risk” (Fernandes-Alcantara 2014) and “opportunity youth” (White House Council for Community Solutions 2012).

<sup>2</sup> Agencies could waive any statutory, regulatory, or administrative requirement, as long as the waiver (1) was in keeping with important safeguards (for example, consistency with the statutory purposes of the program funding); (2) would not result in use of program funds in a way that denied or restricted individual eligibility for services or adversely affect the vulnerable populations who receive those services; and (3) adhered to nondiscrimination, wage and labor standards, or allocated funds to state and sub-state levels (Performance Partnership Pilots for Disconnected Youth 2017).

foster care, homeless youth, young parents, those involved in the justice system, unemployed youth, and those who have dropped out of school or are at risk of doing so (Rosenberg and Brown 2019).

Each Cohort 1 pilot contracted with a third party to conduct an evaluation as part of its grant (hereafter called a local evaluation). These local evaluations were structured to test a subset of the P3 pilots' targeted intervention activities and not the effectiveness of P3 as an authority providing flexibility to grantees with opportunities to use federal funds across programs.

To assess the P3 model, the Chief Evaluation Office at the Department of Labor (DOL), on behalf of its federal P3 partners, contracted with Mathematica and its partner, Social Policy Research Associates. This national evaluation was designed to build an understanding of the implementation and system changes that pilots achieved under P3, and to assess the outcomes of and impacts on youth participants across Cohort 1 pilots (see Box 1). This report synthesizes findings from the local evaluation reports to determine whether the research could establish a causal impact between the studied intervention and participant outcomes and, for interventions that had such evidence, whether the evidence indicated the intervention achieved the expected impacts. By compiling and synthesizing information about the interventions across local evaluations, we answered three research questions:

1. What interventions and outcomes are the focus of the local evaluations?
2. What is the level of rigor in the local evaluations?
3. Do the local evaluations find the expected impacts for their interventions with youth?

This section of the report provides background on the P3 local evaluations and the methodology used to synthesize their findings. The next two sections answer the research questions. Section II answers the first research question, and Section III answers the second and third questions. Section IV discusses the interventions' implementation challenges and successes.

## A. Performance Partnership Pilots for Disconnected Youth (P3) and evaluations of their interventions

The vision for P3 was to improve outcomes for youth by stimulating changes in the systems of agencies, programs, and funding streams, in ways that supported the implementation of innovative programs. Pilots were to use their P3 awards to engage in activities at both the systems and program levels. At the systems level, they were to establish partnerships to promote effective collaborations across their communities'

### Box 1 P3 national evaluation

The national evaluation included the following analyses:

1. An **implementation and systems analysis** examined the role and work of the federal, state, and local partners in changing systems and providing innovative services to youth. Data analyzed included two rounds of interviews with federal stakeholders and visits to Cohort 1 pilots in 2017 and 2018 and visits to the six pilots in 2019 authorized by the 2015 and 2016 Appropriations Acts.
2. An **outcomes analysis** described the youth who participated in Cohort 1 pilots, the services they received, and the employment and education outcomes they achieved. Data analyzed included administrative data collected from the Cohort 1 pilots.
3. A **synthesis of the local evaluations** identified interventions that could improve outcomes for youth by synthesizing the findings from the Cohort 1 local evaluation reports.

Results of these analyses will be reported via multiple papers, one of which is this report.

different organizations and programs serving youth. Improved system coordination and alignment would reduce the barriers that organizations face in jointly serving youth (for example, constraints created by different eligibility requirements) and provide coordinated services for disconnected youth. At the program level, pilots were to implement or expand innovative programs or services for youth and to recruit, engage, and retain youth to improve their employment, education, and other outcomes.

All Cohort 1 pilots engaged with a third-party evaluator to assess a subset of their P3 activities, specifically an intervention intended to improve youth outcomes (Appendix B, Table B.2). All but one evaluation used a causal design that included a comparison group.<sup>3</sup> Mathematica provided evaluation technical assistance (TA) to each of the Cohort 1 pilots and their evaluators to help them design and implement a high quality evaluation.<sup>4</sup> An evaluation TA liaison worked with each pilot throughout the evaluation. An initial assessment of each pilot identified the need for both group (for example, webinars and tools) and individualized assistance. Topics covered that are relevant to this synthesis included defining the target population; benchmarking the study design against relevant evidence standards, including the What Works Clearinghouse (WWC; 2017) and the Clearinghouse for Labor Evaluation and Research (CLEAR; Department of Labor 2019); developing a clear contrast between treatment and comparison conditions; ensuring baseline equivalence; and producing credible evaluation reports.

## B. Synthesis of local evaluation reports

The local evaluation reports served as the sole data source for our synthesis.<sup>5</sup> As a result, the description of each intervention in this report might differ from the information collected through interviews of P3 administrators, staff, and partners during the evaluation's implementation study (Rosenberg and Brown 2019). Such differences might arise because the intervention evaluated is a subset of the interventions provided or because external evaluators' perceptions of the intervention are different from those of program staff.

We developed and applied a five-step sequential process to answer the research questions. Each step represents a standard for proceeding to the next step. If the local evaluation did not meet the standard in one step, the intervention or outcome did not progress to the next step. The standards in each step were informed by those used by the federal research clearinghouses—WWC, CLEAR, and Home Visiting Evidence of Effectiveness (HomVEE) (Sama-Miller et al. 2018).<sup>6</sup>

Through this process, we identified (1) the local evaluations that might not be able to provide rigorous evidence that the intervention caused the expected outcomes and (2) the outcome domains in which the local evaluations did provide evidence that an intervention benefited youth. The consequence of this stepped process was that only a subset of the interventions and expected outcomes were included in the analysis.

---

<sup>3</sup> Cohort 1 pilots received additional points in the competitive review of their proposals for proposing a randomized controlled trial or quasi-experimental design evaluation conducted by a third-party evaluator.

<sup>4</sup> To ensure objectivity in conducting the synthesis, the Mathematica team providing evaluation TA did not conduct the synthesis. The evaluation TA lead did, however, review this report to ensure its accuracy.

<sup>5</sup> For inquiries about P3 and the local evaluation reports, please visit this webpage: <https://youth.gov/youth-topics/reconnecting-youth/performance-partnership-pilots>

<sup>6</sup> For details of the standards, see WWC (<https://ies.ed.gov/ncee/wwc/Handbooks>), CLEAR (<https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21>), and HomVEE ([https://www.acf.hhs.gov/sites/default/files/opre/homvee\\_executive\\_summary\\_2018\\_508.pdf](https://www.acf.hhs.gov/sites/default/files/opre/homvee_executive_summary_2018_508.pdf)).

Applying the standards in each step, a Mathematica researcher rated each evaluation and assessed whether the intervention achieved expected impacts. If information in the local evaluation report was not clear, the researcher asked the person who provided the evaluation TA to clarify. A senior researcher reviewed the researcher's documentation and indicated differences or concerns in the assigned ratings. The two reconciled any differences and arrived at a solution in line with the five-step process.

The five-step process included an assessment for:

1. **Study design.** If executed well, the design must be able to produce a causal impact. This step removed designs that were not a randomized controlled trial (RCT) or a quasi-experimental design (QED) evaluation.
2. **Clarity.** Clearly defined, measurable intervention and outcome(s). The description of an intervention must provide sufficient information about what is being evaluated, including an account of its activities and services and details about its core components. Such interventions were assessed as part of the synthesis as long as it had at least one clearly defined and measurable outcome. In addition, researchers must be able to properly attribute outcomes to the intervention. To synthesize a full range of outcomes that the interventions might have influenced, we assessed both primary and secondary outcomes. Primary outcomes were those derived from the primary research questions and were essential for conducting confirmatory tests of interventions' effectiveness. Secondary outcomes were derived from the secondary research questions. They were not considered confirmatory tests for evaluating interventions' effectiveness, but still considered important to the P3 pilots and local evaluators. Therefore, all outcomes in an evaluation were reviewed, with those eliminated from the synthesis that:
  - Were estimated for different *subgroups*. Even if the evaluation report considered subgroup impacts as a main finding, we considered findings for subgroups supplemental to the evaluation's findings for its full sample. Estimating an intervention's impact on the same outcomes for various groups increases the likelihood of finding statistically significant effects by chance (that is, false positives). To decrease the chance of reporting false positives in the synthesis, we only included outcomes estimated for the full sample.
  - Were defined as a *subset of a measure*. Estimating an impact for an outcome that is conditional on another outcome makes it impossible to determine which outcome the intervention is influencing. For example, gaining employment in a job that offers benefits would be excluded, because benefits are conditional on employment, and it would be impossible to determine whether the intervention impacted employment or access to benefits once employed. Clarity could be gained by estimating the impact on employment and, conditional on employment, estimating the impact of having a job that offers benefits.
  - *Combined outcomes*. Estimating an impact on an outcome that combines two distinct outcomes makes it impossible to determine what the intervention is impacting. For example, if the outcome is measured as attaining a high school diploma or completing vocational training, it captures two distinct outcomes. Clarity could be gained by estimating each outcome separately and then jointly.
  - Had *overalignment* with the studied intervention. If the outcome is actually part of the intervention—for example, enrolling in the training intervention—it is not a program impact but a measure of the dosage of the intervention.

**3. No confounding factors.** The study design must not allow factors other than the studied intervention to influence outcomes. If factors other than the intervention might have influenced outcomes and the effects could not be disentangled from the research design, the intervention was not included in the synthesis. For example, an intervention implemented at a single site by one staff person would not allow researchers to separate the impact of the intervention from the staff person or site environment.

**4. Baseline equivalence.** Members of the treatment and comparison groups must be similar before the intervention begins to ensure that the comparison group's experience presents a valid picture of what would have happened without the intervention. If differences in the characteristics or outcomes captured at baseline exist between the treatment and comparison group, those characteristics, and not the intervention, might have influenced impacts. The measures we

**Box 2**  
**Baseline equivalence measures**

- Age
- Gender
- Race/ethnicity
- Socioeconomic status
- Outcomes measured at baseline (if they can be collected)

used to establish baseline equivalence, listed in Box 2 (Appendix B, Table B.3 provides details), were informed by protocols developed and applied by the federal research clearinghouses.<sup>7</sup> We applied the CLEAR standards for establishing baseline equivalence on those measures: characteristics at baseline must not be statistically significant at the 5 percent level or must control for the characteristics with statistically significant differences when estimating impacts.

**5. Achieved expected impact.** We define an achieved expected impact as one in which the estimated impact of the intervention is statistically significant at the 5 percent level and is in the desirable direction (for example, the evaluation found the intervention increased student attendance). We relied on the local evaluation reports and their theories of change to identify the impacts that the intervention expected to achieve. Estimated impacts that were not significant at the 5 percent level or had significance but in the opposite direction of the expected impact, we consider as not achieving the expected impact.

The studies of interventions that could support a causal impact (Step 1) answered the first research question, What interventions and outcomes are the focus of the local evaluations? Specifically, we identified interventions with clarity in their description and measurable outcomes (Step 2) with respect to the outcomes that the intervention might be able to affect (Step 3).

---

<sup>7</sup> We used the CLEAR reentry and WWC dropout prevention protocols to identify baseline characteristics for establishing equivalence between the treatment and comparison groups because they examine populations and outcomes similar to the pilots. Similarly, because the HomVEE evidence guidelines provide rules for two-generation models, we used it to identify baseline characteristics for the one pilot with that model.

Using Step 4, we answered the second research question, What is the level of rigor in the local evaluations? We developed and applied evidence ratings (Box 3) that depend, in part, on whether the design is an RCT or QED (see Appendix C), with the highest rating reserved for RCTs with low attrition and no reassignment (that is, virtually all study participants were included in the analysis and did not switch from the treatment or control group to which they were assigned).

Using Step 5, we answered the third research question, Do the local evaluations find the expected impacts for their interventions with youth? Within each of the three outcome domains—education, employment, and other—we examined whether each estimated impact of the intervention was expected. Identifying interventions that achieved expected impacts within each outcome domain allowed us to assess the kinds of outcomes influenced by the interventions.

As is standard in this type of work, our approach to synthesizing findings across local evaluation reports contains several limitations. Most importantly, we base our analysis on only the information provided in the reports, and we could not independently confirm their findings. Nor could we assess fidelity in the intervention’s implementation. In addition, some of the local evaluations might have captured some of the system-level effects of P3 as part of the intervention, but we do not have information to determine whether and how often that occurred.

### **Box 3** **Ratings for rigor**

- **High rating:** Strong evidence exists that the estimated impacts are attributable to the intervention.
- **Moderate rating:** Evidence exists that the estimated impacts are attributable at least in part to the intervention.
- **Low rating:** Estimated impacts may not be attributable to the intervention; other factors might have contributed.



## II. Interventions and Outcomes Studied

This section of the report answers the first research question—What interventions and outcomes are the focus of the local evaluations?—using the first three steps of the five-step process. We eliminated from this discussion—and the remainder of the synthesis—the intervention assessed by a pre-post evaluation without a comparison group, because its design did not provide confidence that the intervention caused the results. (As Box 4 summarizes, this is the only stage in which we eliminated an intervention from analysis.) All eight of the evaluations with a causal design using a comparison group clearly defined their intervention in the local evaluation report and all allowed for an assessment of the intervention without confounding factors.<sup>8</sup> Five of the eight interventions included some form of case management. Appendix B provides a full listing of the interventions (Table B.5). Of the eight evaluations with a causal design, five demonstrated that the studied intervention achieved at least one of its expected impacts.

All evaluations studied interventions offered only to the treatment group, although the control or comparison group could participate in other programs or services provided by the pilot or other agencies. Five of the eight evaluations studied interventions that offered only one additional service to the treatment group relative to the comparison group: three offered case management services (only), one offered a leadership training, and one offered the Teen Outreach Program® (TOP®) curriculum. Three evaluations studied interventions with more than one service offered only to the treatment group: two offered case management services augmented with either soft-skills training or access to WIOA services, and one offered a two-generation education and training program with three components—mentoring, subsidized employment, and the Parents as Teachers (PAT) curriculum.

All eight evaluations had some outcomes meeting our criteria for inclusion and some outcomes that did not. We eliminated subgroup outcomes for two evaluations: one looked at outcomes by grade (38 outcomes) and the other by youth who were in foster care, on probation, or youth in public housing (18 outcomes). We eliminated combined outcomes for two evaluations (attainment of employment/enrolled in postsecondary education [three outcomes]) and an overaligned outcome in one evaluation (employment for youth while receiving subsidized employment [1 outcome]).

### Box 4

#### Results of the five-step process

1. Design supports casual research: One intervention with a pre-post design without a comparison group was eliminated
2. Clarity in descriptions of studied intervention and measurement: None eliminated
3. No confounding factors: None eliminated
4. Baseline equivalence: None eliminated
5. Achieved expected impacts: Three of the six interventions analyzed did not achieve expected impacts

---

<sup>8</sup> One local evaluation used two comparison groups: one historical and one contemporaneous. The historical comparison group examined the impacts between a comparison group receiving services from 2013 to 2015 and a treatment group that received services between 2016 and 2017, which made it impossible to determine how much of the observed impact was due to the studied intervention and how much was due to economic conditions. Because of this confounding factor, we include only the research for the contemporaneous comparison group in the synthesis.

The interventions were designed to affect education, employment, and other outcomes. Box 5 summarizes the studied interventions and some of the outcomes included in the local evaluation reports. (Tables B.4 and B.5 in Appendix B provide a comprehensive list of outcomes and description of interventions.) Of the three evaluations assessing the impact of case management (only), two expected it to improve education outcomes such as reducing absences and suspensions and one by increasing out-of-school youth returns to secondary education. Two of three of the evaluations expected it to increase employability skills and employment attainment, and one expected it to increase the number of WIOA and support services received. When case management was combined with soft-skills training, it was expected to increase education outcomes (such as high school/General Education Development [GED] completion and decreased suspensions and unexcused absences); when it was combined with WIOA services, it was expected to increase both education (such as attainment of a high school diploma or GED) and employment outcomes (such as attainment of unsubsidized employment).

The other three interventions were also expected to improve multiple outcomes. Leadership and TOP® training were expected to increase education outcomes (such as reduced risk for dropping out and loss of credit, and increased high school/GED completion). In addition, leadership training was expected to increase cultural knowledge. The two-generation education and training program was expected to increase positive child development, employment for the parent, and enrollment in public benefit programs for families.<sup>9</sup>

## **Box 5** **Interventions and outcomes studied in causal designs**

### **Interventions**

- Case management (only) (three evaluations)
- Case management and soft-skills training
- Case management and WIOA services
- Leadership training
- TOP®
- Two-generation education and training

### **Select outcomes**

#### **Education**

- High school/GED completion
- Out-of-school youth returns to secondary education
- Suspensions decreased
- Unexcused absences decreased
- Loss of school credit decreased

#### **Employment**

- Employment attainment
- Employability skills increased
- Internship/paid work experience completion
- Job-readiness training completion
- WIOA service receipt

#### **Other**

- Positive child development
- Cultural knowledge
- Public benefit enrollment

---

<sup>9</sup> Employment outcomes were overaligned with the studied intervention, subsidized employment, and not assessed.

### III. Interventions Achieving Expected Impacts

This section answers the second and third research questions—What is the level of rigor in the local evaluations? and Do the local evaluations find the expected impacts for their interventions with youth?—using the last two steps of the five-step process. We answer the second question separately for education and employment outcomes and other outcomes.

#### A. Level of rigor in the local evaluations

All eight interventions with evaluations using causal research designs with comparison groups had a high or moderate rating for rigor using the DOL CLEAR standards.<sup>10</sup> Three interventions were assessed as an RCT and, therefore, received a high rating (Appendix B, Table B.6). Five had a moderate rating, since they were assessed as either a QED or an RCT that needed to establish baseline equivalence of its treatment and comparison groups (Appendix B, Table B.7). Although three of the five did not establish baseline equivalence for their analytic samples, all included the proper controls in their regression models to adjust for differences measured at baseline and, as a result, met the CLEAR criteria for baseline equivalence. The local evaluations using a QED used propensity score matching to create comparison groups similar to the treatment group, used propensity score weighting to establish baseline equivalence, or included the outcome measures specified in our criteria related to age, gender, race/ethnicity, and socioeconomic status when estimating impacts.

Of the 81 outcomes examined, each of which had to establish baseline equivalence separately, 29 had impacts estimated with a high level of rigor, 52 were rated as moderate rigor, and none was rated as low rigor. Although the studies using a QED established baseline equivalence or controlled for differences in their analyses, three of the evaluations noted some additional limitations that resulted from their study design. For example, two studies cited unobserved factors, such as differences in motivation between youth who chose to participate in the P3 pilot program and those who chose not to participate, and another cited possible positive assignment bias to treatment and comparison services as limitations that might have affected study outcomes.<sup>11</sup> Two of the evaluations using a QED did not report any limitations related to their study design.

---

<sup>10</sup> Mathematica provided pilots and their evaluators with evaluation TA geared to the CLEAR evidence guidelines. Had we applied the WWC or HomVEE standards to review the local evaluations, only four would have a high or moderate rating, because they did not meet the rules for establishing baseline equivalence under those rules.

<sup>11</sup> Case managers likely assigned a greater number of promising youth to the treatment condition in one pilot. The intervention was designed for case managers to assess youth and then set and sequence services needed to increase reengagement in high school or GED completion and better connect youth to employment services. Youth assessed as work-ready could enroll in the intervention (treatment group), whereas those who were not work-ready were required to complete their GED and receive support services (comparison group). The possible positive assignment to treatment and comparison services was not considered a confounding factor, because the sequencing of services was a component of the intervention; however, the employment findings might be biased in favor of the treatment group.

## B. Interventions achieving expected education and employment impacts

Two of the five interventions evaluated with moderate or high levels of rigor had evidence that they achieved at least one of their expected primary or secondary impacts in the education or employment domain: case management (only) and case management and WIOA services.<sup>12</sup> Importantly, not all of the research that examined these interventions demonstrated that they achieved the expected impacts; some findings were null or even in the opposite direction and therefore did not achieve the expected impact. (Appendix B, Table B.8 provides details on magnitude and significance of all primary and secondary findings.) Three interventions—leadership training, case management and soft-skills training, and TOP®—did not achieve any expected primary or secondary impacts in the education and employment domains.

During the evaluation TA, local evaluators were advised to distinguish between primary and secondary research questions. The primary outcomes identified in the primary research questions could provide evidence that the interventions achieved their expected impacts, such as increasing participants' return to high school education and their employability skills. The impacts on interventions' primary education and employment outcomes are in Table III.1. Some of the secondary outcomes identified in secondary research questions aligned with education or employment and shed light on how the interventions might have influenced whether pilot participants met a state-mandated online course graduation requirement, improved their grade point average to a 2.0 or higher, or completed a paid internship. In total, the local evaluations reported that the interventions achieved the expected impact on seven secondary education and employment outcomes (Table III.2). (See Table B.8 for the list including secondary outcomes.)

---

<sup>12</sup> The two-generation intervention evaluated an employment outcome, but the outcome was overlapped with the intervention. Therefore, this step did not assess any education or employment outcomes for this intervention.

**Table III.1. Impacts on primary education and employment outcomes, by intervention**

Intervention	Outcome	Finding (time point)
<b>Education</b>		
Case management (only), Pilot 1	Suspensions	– (1 year); ✓ (2 years)
Case management (only), Pilot 1	Unexcused absences	0 (1 year and 2 years)
Case management (only), Pilot 2	Returned to secondary education	✓ (within 1 year)
Case management (only), Pilot 3	Number of absences	0 (1 year)
Case management (only), Pilot 3	Number of suspensions	0 (1 year)
Case management (only), Pilot 3	Ever suspended	0 (1 year)
Case management and soft-skills training, Pilot 4	Attendance	0 (1 year)
Case management and soft-skills training, Pilot 4	Grade promotion	0 (1 year)
Case management and soft-skills training, Pilot 4	Passed state-mandated test	0 (1 year)
Case management and soft-skills training, Pilot 4	Completed high school	0 (1 year)
Case management and WIOA services, Pilot 5	Obtained high school diploma or equivalent	0 (9 months)
Leadership training, Pilot 6	Cultural knowledge	0 (1 month)
Leadership training, Pilot 6	Motivation for school	0 (1 month)
Leadership training, Pilot 6	Decreased risk for dropping out of school	0 (1 month)
Leadership training, Pilot 6	Improved school behavior	0 (1 month)
TOP®, Pilot 7	Passed English course	0 (1 month)
TOP®, Pilot 7	Absences	0 (1 year)
<b>Employment</b>		
Case management (only), Pilot 1	Increased employability skills test score	✓ (1 year and 2 years)
Case management (only), Pilot 2	Participated in career preparation	– (within 1 year)
Case management (only), Pilot 2	Participated in subsidized employment	– (within 1 year)
Case management (only), Pilot 2	Employed at end of program	✓ (within 1 year)
Case management (only), Pilot 3	Received WIOA service	0 (1 year)
Case management and WIOA services, Pilot 5	Attained unsubsidized employment	0 (9 months)

Notes: A checkmark (✓) indicates that the local evaluation found that the intervention achieved expected impacts at the 5 percent significance level. A 0 indicates that impacts were not significant at the 5 percent level, and a minus sign (–) indicates that the impacts were in the opposite direction at the 5 percent significance level. The time point in parentheses specifies how long after program start the outcome was assessed.

TOP® = Teen Outreach Program®; WIOA = Workforce Innovation and Opportunity Act.

**Table III.2. Achieved expected impacts on secondary outcomes, by intervention**

Intervention	Outcome	Finding (time point)
<b>Education</b>		
Case management (only), Pilot 1	Met state-mandated online course graduation requirement	✓ (2 years)
Case management and WIOA services, Pilot 5	Attained GPA of 2.0	✓ (9 months)
Case management and WIOA services, Pilot 5	Completed two GED tests	✓ (9 months)
Case management and WIOA services, Pilot 5	Attained college-course readiness	✓ (9 months)
<b>Employment</b>		
Case management and WIOA services, Pilot 5	Completed job-readiness training	✓ (9 months)
Case management and WIOA services, Pilot 5	Completed career education class	✓ (9 months)
Case management and WIOA services, Pilot 5	Paid internship completion	✓ (9 months)

Notes: A checkmark (✓) indicates that the local evaluation found that the intervention achieved expected impacts at the 5 percent significance level. The time point in parentheses specifies how long after program start the outcome was assessed.

GED = General Education Development; GPA = grade point average; WIOA = Workforce Innovation and Opportunity Act.

**Case management (only).** Two of the three local evaluation reports assessing case management (only) found that the intervention achieved at least one of its expected impacts at the 5 percent significance level; the third did not achieve any expected impacts.

- In the first case management (only) pilot, a case manager worked one on one with youth to develop individualized service plans to help guide their selection of other available services. Within the education domain, the intervention achieved two of its expected impacts. Although this intervention ultimately achieved its expected primary impact on reduced school suspensions at the two-year follow-up (difference in means = -0.10), one-year impacts were actually the opposite of what was predicted (difference in means = 0.09). The intervention achieved its expected secondary impact on increased completion of an online course (a state-mandated graduation requirement) (difference in means = 0.07) two years following random assignment.<sup>13</sup> Within the employment domain, the intervention achieved its expected primary impacts, demonstrating an increase in test scores for employability skills one year (difference in means = 0.12) and two years (difference in means = 0.13) after random assignment.
- In the second case management (only) pilot, case managers helped youth return to school and reported an increased probability that youth returned to school within the first follow-up year (difference in means = 0.21). The intervention also achieved its expected primary impact on increasing employment at the end of the program (difference in means = 0.20), although youth

<sup>13</sup> Although not in the education domain, the same intervention found that family members of youth receiving case management services were *less* likely to receive referrals to adult education.

receiving the intervention were also less likely to receive career-preparation skills training (job shadowing, occupational skills training, pre-apprenticeship, or apprenticeship) (difference in means =  $-0.07$ ) and subsidized employment services (difference in means =  $-0.21$ ).

**Case management and WIOA services.** This intervention did not achieve its expected impacts for its primary outcomes: obtain high school diploma or equivalent and unsubsidized employment. However, evidence showed that it achieved its expected secondary impacts on education and employment outcomes. Within the education domain, it increased probability that out-of-school youth took two GED tests (irrespective of whether they passed) (difference in means =  $0.29$ ) and attained college-course readiness (difference in means =  $0.04$ ), and that those who enrolled in college attained a GPA of 2.0 (difference in means =  $0.09$ ). Within the employment domain, it achieved the expected secondary impact on increasing completion of job-readiness training (difference in means =  $0.27$ ), career education classes (difference in means =  $0.11$ ), and paid internships (difference in means =  $0.21$ ).<sup>14</sup>

### C. Interventions achieving expected impacts in other domains

Two interventions achieved expected outcomes at the 5 percent significance level in other domains, including the two-generation education and training program for young parents and their children and the case management (only) intervention.

**The two-generation program for young parents and their children.** This intervention achieved expected impacts on two primary outcomes: increasing the probability that children attended their child care program more regularly one month after enrollment (difference in means =  $6.07$ ) and that enrollment in the Temporary Assistance for Needy Families (TANF) program increased among parents 11 months after enrollment (difference in means =  $0.11$ ). It did not achieve expected impacts on its other primary outcomes (see Appendix B, Table B.8).

**Case management (only).** One intervention had evidence that it achieved the expected impact on two secondary outcomes: increasing family participation in school events at one and two years after random assignment (difference in means =  $0.22$  and  $0.09$ , respectively). It also had evidence of achieving the expected impact on youth being more likely to participate in program services available to both treatment and comparison group members one year after random assignment (difference in means =  $0.10$ ). Another evaluation of a case management (only) intervention achieved expected secondary impacts on increasing rates of youth receiving services from a partner one year after program entrance (difference in means =  $0.15$ ).

---

<sup>14</sup> Positive impacts on outcomes in the employment domain might have emerged because of positive assignment into the treatment group (see Footnote 8).

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



## IV. Implementation Challenges and Successes

Our synthesis of the research and findings in the eight rigorous local evaluation reports provided insights into which interventions had rigorous evidence demonstrating expected impacts. It indicated that across four of the eight local evaluation reports, the studied interventions achieved expected impacts on 18 primary and secondary outcome measures at the 5 percent significance level. Across all eight reports, the studied interventions did not achieve expected impacts on 63 outcomes, with 4 outcome measures having impacts in the opposite direction and 59 outcome measures having no impact.<sup>15</sup> In this section, we use information from the local evaluation reports to describe factors in the implementation that challenged and facilitated the interventions' ability to realize expected impacts.

The local evaluation reports described challenges in implementation. Because five of the eight local evaluation reports assessed some form of the case management intervention, we discuss implementation difficulties for that intervention separately from the others. For in-school youth at risk of dropping out, case managers assessed youths' needs, developed an individual service plan that included both academic and career goals as well as strategies to achieve those goals, helped youth identify services that could assist them in achieving their goals, and provided youths' families with assistance to address educational and resource needs. The local evaluators noted that challenges arose, however, as case managers' ability to provide services varied—not all case managers were bilingual (and some students spoke only Spanish) and schools did not always allow them to pull youth from class, decreasing their interaction with youth. For out-of-school youth, local evaluation reports suggested that youth appreciated the support case managers provided and being treated like adults, but challenges arose in connecting youth to services needed. One report discussed the perceived challenges that arose from the youths' environments, including transience and problems at home or with family. Another discussed potential challenges in terms of staff training: case managers were trained to reconnect youth to secondary education and were less familiar with postsecondary training opportunities. The study authors suggested that this factor might have increased youths' return to secondary education and employment but reduced the probability that they accessed career preparation and employment services. One potentially critical implementation barrier—within the spirit of P3—was noted in the local evaluation report that did not achieve expected impacts for the case management (only) intervention: case managers reported to the evaluators that they often felt they were fighting to place youth in appropriate services outside their jurisdiction rather than collaborating across jurisdictions.

Information in the local evaluation reports assessing interventions other than case management highlighted several implementation challenges that might have prevented the studied intervention from achieving expected impacts, including:

- Inconsistency in providing key intervention components for the first cohort of participants
- High staff turnover rates that produced lapses in which no services were provided

---

<sup>15</sup> Outcomes are counts summed for each grantee and then across grantees. For example, if two grantees measured absences, we counted it twice, because the association was for different interventions. Similarly, if a grantee measured absences at one- and two-year follow-ups, we counted it twice, because the association was for different analytic samples. Had we excluded the case management and WIOA services intervention from the analysis due to the positive assignment, the five interventions across seven local evaluation reports would have achieved expected impacts on 12 primary and secondary outcome measures and would not have achieved expected impacts on 60 outcome measures, with 4 outcome measures in the opposite direction and 56 outcome measures with no impact.

- A natural disaster that made it difficult to implement the intervention
- Conflicts between the timing of the program services and other after-school activities

In addition to these implementation challenges, the local evaluation reports discussed research challenges that might have prevented them from being a true assessment of the studied intervention. Such challenges included:

- Insufficient time to observe longer-term outcomes. According to evaluation reports, five of the local evaluations had theories of change that specified longer-term outcomes than were not measurable during the study period. For example, two of the pilots sought to increase postsecondary degree attainment.
- The comparison group received many of the same services as the treatment group or was exposed to the studied intervention. For example, four of the evaluations' counterfactual conditions excluded one or two components of the pilot intervention, resulting in both groups accessing similar education and job training services, but the comparison group in three of the four evaluations did not have access to the P3 pilot case manager, and the comparison group in the other evaluation did not have access to the TOP® curriculum (see Appendix Table B.5). Additionally, three of the reports noted some contamination in which comparison group members were exposed to the intervention condition.
- Insufficient sample, and thus power, to detect impacts. Three of the reports acknowledged that their small sample sizes were a limitation. For example, one reported requiring a minimum sample size of 90 youth in its power analysis but only had an analytic sample size of 64 participants.<sup>16</sup>
- One of the studies reported inconsistencies in its data entry, which the authors suggested could have affected observed impacts.

Despite these challenges, the local evaluation reports presented evidence that two interventions achieved expected impacts in the education and employment domains, and two of the reports discussed how the interventions' implementation contributed to these successful outcomes. A local evaluation report of a case management (only) intervention described how access to students' records and strong relationships with various alternative school systems contributed to the pilot's success in reconnecting youth to school. The same report noted that the case managers often presented the prospect of a job as an incentive to participate in other pilot activities. The case managers perceived that participants were more likely to enroll in alternative education and complete their secondary degree if allowed to work part-time.

The local evaluation report that demonstrated the effectiveness of case management and WIOA services included qualitative findings from participant interviews. Those participants noted that the support from and consistent contact with their case workers helped them overcome challenges and provided them with the support needed to reengage in their schooling. They also noted the benefits of employment services: receiving information about job fairs, assistance with job searches, exploring career options, and getting hands-on experience in their careers of interest. In particular, participants cited the guidance received during resume preparation workshops and training programs as instrumental in helping them to increase their confidence to apply for and obtain employment.

The focus of P3 was on outcomes-based strategies to help youth develop the knowledge, skills, and behaviors needed to transition into and be self-sufficient in adulthood. The knowledge gleaned from the

---

<sup>16</sup> Sample sizes ranged from 50 to 1,370 (for both treatment and comparison groups).

local evaluation reports suggests that the pilots have made progress. Still, the local evaluations were not designed to assess the extent to which pilots changed systems to help youth overcome their challenges in transitioning to adulthood. The national evaluation's other components are assessing pilots' ability to fully realize the P3 authority, for example, by coordinating funding facilitated by waivers to change systems across the education, employment, and other domains.

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

## References

- Corcoran, M., F. Hanleybrown, A. Steinberg, and K. Tallant. "Collective Impact for Opportunity Youth." 2012. Available at <http://www.opensocietyfoundations.org/sites/default/files/collective-impact-opportunity-youth20120919.pdf>. Accessed February 8, 2015.
- Dion, R., M.C. Bradley, A. Gothro, M. Bardos, J. Lansing, M. Stagner, and A. Dworsky. "Advancing the Self-Sufficiency and Well-Being of At-Risk Youth: A Conceptual Framework." OPRE Report # 2013-13. Washington, DC: U.S. Department of Health and Human Services, Office of Planning, Research and Evaluation, Administration for Children and Families, 2013.
- Department of Labor. "CLEAR Causal Evidence Guidelines, Version 2.1." 2019. <https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21>. Accessed February 11, 2019.
- Fernandes-Alcantara, A. "Vulnerable Youth: Background and Policies." CRS Report 7-5700, 2014. Available at [http://digitalcommons.ilr.cornell.edu/key\\_workplace/932](http://digitalcommons.ilr.cornell.edu/key_workplace/932). Accessed February 8, 2015.
- Performance Partnership Pilots for Disconnected Youth. "First Annual Report to Congress." July 2017. Available at <https://youth.gov/sites/default/files/P3-Report-to-Congress-508.pdf>. Accessed April 12, 2019.
- Rosenberg, Linda, and Elizabeth Brown. "Early Experiences of the Performance Partnership Pilots for Disconnected Youth (P3): Cohort 21 Pilots." Princeton, NJ: Mathematica, July 2019.
- Sama-Miller, Emily, Lauren Akers, Andrea Mraz-Esposito, Marykate Zukiewicz, Sarah Avellar, Diane Paulsell, and Patricia Del Grosso. "Home Visiting Evidence of Effectiveness Review." Washington DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018. Available at [https://www.acf.hhs.gov/sites/default/files/opre/homvee\\_executive\\_summary\\_2018\\_508.pdf](https://www.acf.hhs.gov/sites/default/files/opre/homvee_executive_summary_2018_508.pdf). Accessed February 11, 2019.
- What Works Clearinghouse. "What Works Clearinghouse Standards Handbook, Version 4.0." Washington, DC: Department of Education, Institute for Education Statistics, 2017. Available at [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_standards\\_handbook\\_v4.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf). Accessed February 11, 2019.
- White House Council for Community Solutions. "Final Report: Community Solutions for Opportunity Youth." Washington, DC: White House Council for Community Solutions, 2012. Available at [http://www.serve.gov/sites/default/files/ctools/12\\_0604whccs\\_finalreport.pdf](http://www.serve.gov/sites/default/files/ctools/12_0604whccs_finalreport.pdf). Accessed February 8, 2015.

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

## **Appendix A**

### **Description of P3 Pilots**

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



Descriptions of the complete Cohort 1 pilot services offered, as described in the local evaluation reports, included:

- **Baton Rouge, Louisiana.** Connected at-risk youth in alternative schools to intensive case management with service coordination across education, employment, and juvenile justice sectors. Additional services included career coaching, professional learning communities, and internship and industry certification opportunities.
- **Broward County, Florida.** Provided tiered case management services to in-school youth informed by comprehensive youth assessment with touch points in school, out of school, and during summer. Provided a holistic continuum of care coordination focused on social-cognitive development, interpersonal relationships, and internal self-efficacy.
- **Chicago, Illinois.** Provided a two-generation education and training program that served both young parents and their children. Services to young parents included mentorship, peer support, home visiting, subsidized employment placement, and training in child care centers to improve parenting skills and self-sufficiency. Services to children included Head Start, child socializations, and home visits.
- **Indianapolis, Indiana.** Provided case management services with weekly one-on-one mentoring and core services through a “coordination team” that monitored progress toward individual goals and made service recommendations to health, employment, and job training providers.
- **Los Angeles, California.** Provided services to reconnect youth to education with support from case managers. Monthly regional meetings were held for case review and coordination among programs and community-based partner organizations.
- **The State of Oklahoma.** Provided specialized case management services to assist foster care youth with education completion, transportation assistance, aging out of care, and housing assistance.
- **Seattle, Washington.** Case managers sequenced education and employment services at the appropriate time to increase re-engagement in high school or General Education Development (GED) services and connections to Workforce Innovation and Opportunity Act (WIOA) employment services.
- **Southeastern Kentucky.** Delivered the Teen Outreach Program® (TOP®) to at-risk youth in three high schools. TOP® components include group discussions, positive adult guidance and support, and community service-based learning. Topics include healthy relationships, communication, goal setting, decision making, values clarifications, adolescent development, and sexuality.
- **Ysleta del Sur Pueblo.** Provided a leadership-based curriculum incorporating cultural knowledge to teach tribal youth about their history, their language, and unique services they can access. Case managers provided additional individualized services with monthly meetings to track progress and connect youth to additional services as needed.

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

## **Appendix B**

### **Data Tables**

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

This appendix contains detailed tables that provide supporting information for statements made in the text. The local evaluation reports provide the information contained in the tables. Tables are presented in the order in which the text references them. Tables B.1 through B.3 support Section I; Tables B.4 and B.5 support Section II; and Tables B.6 through B.8 support Section III.

We used the following acronyms in the tables:

CLEAR	Clearinghouse for Labor Evaluation and Research
CNCS	Corporation for National and Community Services
DOL	U.S. Department of Labor
FAFSA	Free Application for Federal Student Aid
GED	General Education Development
HHS	U.S. Department of Health and Human Services
IMLS	Institute of Museum and Library Services
ISP	individual service plan
n.a.	not applicable
P3	Performance Partnership Pilots for Disconnected Youth
PAT	Parents as Teachers
QED	quasi-experimental design
RCT	randomized controlled trial
TANF	Temporary Assistance for Needy Families
TOP®	Teen Outreach Program®
SNAP	Supplemental Nutrition Assistance Program
WIC	Women, Infants and Children
WIA	Workforce Investment Act
WIOA	Workforce Innovation and Opportunity Act
WWC	What Works Clearinghouse

**Table B.1. Characteristics of pilot interventions**

Characteristic	Percentage (number) of evaluations
<b>Target populations</b>	
Disconnected youth	100% (9)
At-risk of dropout	33% (3)
Foster care youth	11% (1)
Teen parents	11% (1)
Tribal youth	11% (1)
Youth in public housing	11% (1)
<b>Domains in which services were offered</b>	
Education	100% (9)
Employment	78% (7)
Social development/well-being	56% (5)
Service access	44% (4)
<b>P3 pilot service offered</b>	
Case management	88% (8)
Work experience opportunities	67% (6)
Tutoring and homework assistance	44% (4)
WIOA program services	44% (4)
Education exploration	33% (3)
Employability skills training	33% (3)
Individual service plan	33% (3)
Academic assessment	22% (2)
Career assessment	22% (2)
Career exploration	22% (2)
Career planning	22% (2)
Financial literacy	22% (2)
GED completion support	22% (2)
Mentoring	22% (2)
Referrals to community services	22% (2)
TOP® curriculum	22% (2)
Academic support services	11% (1)
Career academics	11% (1)
Character education	11% (1)
College Bound training	11% (1)
Credit recovery	11% (1)
Emergency housing	11% (1)
End of course exam support	11% (1)
Entrepreneurial training	11% (1)
Families and Schools Together (FAST)	11% (1)
Head Start (youth's child)	11% (1)

Table B.1 (continued)

Characteristic	Percentage (number) of evaluations
Healthy meals at each site	11% (1)
Home visits	11% (1)
Individualized learning	11% (1)
Leadership program	11% (1)
Mental health assessment	11% (1)
Mental/behavioral health services	11% (1)
Parent engagement (youth's parent)	11% (1)
Peer mentoring	11% (1)
Physical fitness and nutrition education	11% (1)
Project-based learning	11% (1)
Supportive services	11% (1)
Transition services	11% (1)
Youth in government program	11% (1)
<b>Setting in which programs were implemented</b>	
Middle/high school	33% (3)
Workforce center	33% (3)
Child welfare offices	11% (1)
Community college	11% (1)
Head Start sites	11% (1)
Public housing site	11% (1)

Source: P3 program logic models in local evaluation reports. Actual service activities and studied interventions may vary from what logic model reports.

Note: Table is based on nine evaluations.

**Table B.2. Characteristics of local evaluations**

Characteristic	Percentage (number) of evaluations
<b>Evaluation design</b>	
QED	55% (5)
RCT	33% (3)
Pre-post design	11% (1)
<b>Counterfactual (one evaluation had two counterfactuals; the pre-post design has no counterfactual)</b>	
Business as usual	33% (3)
All but ISP and case management	22% (2)
All but program curriculum	11% (1)
All but WIOA	11% (1)
Alternative education program	11% (1)
Service delivery in the past (historical comparison group)	11% (1)
<b>Outcome domains examined</b>	
Education	88% (8)
Employment	44% (4)
Other	
Social development	33% (3)
Service access	22% (2)
<b>Intervention components</b>	
Case management	86% (7)
Education reconnection or completion	50% (4)
Connect to employment/internships	36% (3)
Social/cognitive development	36% (3)
Career coaching	13% (1)
Home visiting	13% (1)
Housing	13% (1)
Transition services	13% (1)

Source: P3 program local evaluation reports.

Note: Table is based on nine evaluations.



**Table B.3. Establishing baseline equivalence**

Baseline category	Accepted measures
Age	Average age
Gender	All categories included in the evaluation report
Race/ethnicity	All categories included in the evaluation report
Socioeconomic status	One economic well-being measure—income, earnings, poverty levels based on federal thresholds, free and reduced-price lunch status, being from a single-parent family, parent’s education, immigrant or English learning status, teen parent, special education or disability status, public benefit receipt (TANF, SNAP, Medicaid), education or employment of a household member
Outcomes measured at baseline	Must be included if such measures could be collected. For studies estimating academic outcomes, baseline equivalence is required for at least one of the following: standardized test scores, whether behind in grade level, frequency of behavior or discipline incidents, rate of school attendance, grade point average.

Source: Adapted from WWC (2017, Dropout intervention protocol); CLEAR (DOL 2019, Reentry protocol); Sama-Miller et al. (2018, Evidence guidelines).

**Table B.4. Defining and measuring outcomes**

Intervention	Measurable outcomes
<b>Case management</b>	
Only, Pilot 1	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• High school/GED completion</li> <li>• Post-secondary enrollment</li> <li>• Unexcused absences</li> <li>• Suspensions</li> <li>• Career technical education endorsement on diploma</li> <li>• Reading courses passed</li> <li>• State-mandated online course graduation requirement completed</li> <li>• FAFSA completed</li> </ul> <p><b>Employment</b></p> <ul style="list-style-type: none"> <li>• Employed after high school</li> <li>• Employability skills test score</li> <li>• Summer youth employment</li> </ul> <p><b>Other</b></p> <ul style="list-style-type: none"> <li>• Program attendance</li> <li>• Housing stability</li> <li>• Juvenile justice involvement</li> <li>• Family participation in school event</li> <li>• Family referral to adult education</li> </ul>
Only, Pilot 2	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Returned to secondary education</li> <li>• Post-secondary education or training completion</li> <li>• Degree or certification completion</li> </ul> <p><b>Employment</b></p> <ul style="list-style-type: none"> <li>• Skills training participation</li> <li>• Subsidized employment participation</li> <li>• Employed at end of program</li> </ul>
Only, Pilot 3	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Absences</li> <li>• Number of suspensions and ever suspended</li> </ul> <p><b>Employment</b></p> <ul style="list-style-type: none"> <li>• WIOA services received</li> </ul> <p><b>Other</b></p> <ul style="list-style-type: none"> <li>• Services received from partner</li> </ul>
And soft-skills training, Pilot 4	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Attendance</li> <li>• Completing school</li> <li>• Grade promotion</li> <li>• Passing state-mandated test</li> </ul>

Table B.4 (continued)

Intervention	Measurable outcomes
And WIOA employment and education services, Pilot 5	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Attaining college course readiness</li> <li>• Attaining high school diploma or equivalent</li> <li>• Completing two general education development tests</li> <li>• High school skills test level gain</li> <li>• Attaining grade point average of 2.0</li> </ul> <p><b>Employment</b></p> <ul style="list-style-type: none"> <li>• Attaining unsubsidized employment</li> <li>• Job-readiness training completion</li> <li>• Career education class completion</li> <li>• Paid internship completion</li> </ul>
<b>Other</b>	
Leadership training, Pilot 6	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Cultural knowledge assessment</li> <li>• Decreased in number of risk factors for dropping out of school</li> <li>• Loss of school credit</li> <li>• Motivation for school questionnaire</li> <li>• Number of referrals for poor behavior</li> <li>• Grade point average</li> </ul>
TOP®, Pilot 7	<p><b>Education</b></p> <ul style="list-style-type: none"> <li>• Absences</li> <li>• Passing English course</li> </ul>
Two-generation education and training program: (mentoring, PAT, and subsidized employment), Pilot 8	<p><b>Other</b></p> <ul style="list-style-type: none"> <li>• Child development (literacy scores, cognitive scores, social-emotional scores, Head Start attendance)</li> <li>• Public benefit receipt (WIC, TANF, SNAP)</li> </ul>

Source: P3 program local evaluation reports.

**Table B.5. Clearly defining the intervention**

	Definition of intervention	Description of intervention assessed	Counterfactual	Adequate description	No confounding factors
<b>Case Management</b>					
Only, Pilots 1-3	Case management. Individual or small group meeting(s) with an employment specialist or counselor who helps to assess clients' needs and address barriers. May include coaching and developing an individualized plan.	Case managers engaged with youth to develop and complete an ISP, which typically included academic and post-secondary career goals and strategies to achieve those goals. The case managers helped their assigned youth identify the services offered by after-school programs, WIOA, summer employment opportunities, and other community organizations that would assist them in achieving their ISP goals and facilitated their enrollment and participation in after-school programming. (Pilot 1)	Offered same out-of-school time services but without access to a case manager or completion of an ISP	✓	✓
Only, Pilots 1-3	Case management. Individual or small group meeting(s) with an employment specialist or counselor who helps to assess clients' needs and address barriers. May include coaching and developing an individualized plan.	A counselor was placed at each WIOA program site to assess educational needs and help return youth to school. (Pilot 2)	Business as usual services at WIOA sites without a counselor	✓	✓
Only, Pilots 1-3	Case management. Individual or small group meeting(s) with an employment specialist or counselor who helps to assess clients' needs and address barriers. May include coaching and developing an individualized plan.	Caseworkers engaged with youth in monthly face-to-face meetings that focused on accomplishing education and employment goals identified in their independent living and transition plans. (Pilot 3)	Business as usual child welfare services without a specialized child welfare caseworker	✓	✓
And soft-skills training, Pilot 4	Soft-skills training. Training in so-called "soft" skills, such as punctuality, manners, professional dress, how to interact with colleagues, how to handle conflict.	Assigned case managers helped youth complete an ISP tailored to meet needs such as academic, career, and social. Career coaches helped connect youth to internship and industry certification opportunities.  Learning community sessions provided by career coaches included soft-skills training.	Alternative education program	✓	✓

Table B.5 (continued)

	Definition of intervention	Description of intervention assessed	Counterfactual	Adequate description	No confounding factors
And WIOA services, Pilot 5	WIOA employment and education services. Services provided at centers operated under the WIOA and/or WIA.	Case managers decide ordering of services based on an assessment and relationship with the youth. Services include enrolling in GED completion services for up to 60 days before determining enrollment in employment and enhanced case management services. Youth in the treatment group received additional employment and education services funded through WIOA.	Contemporaneous comparison group: basic case management and GED completion support services	✓	✓
And WIOA services, Pilot 5	WIOA employment and education services. Services provided at centers operated under the WIOA and/or WIA.	Case managers decide ordering of services based on an assessment and relationship with the youth. Services include enrolling in GED completion services for up to 60 days before determining enrollment in employment and enhanced case management services. Youth in the treatment group received additional employment and education services funded through WIOA.	Historical comparison group: served from July 2013 to June 2015; received basic case management, GED completion support services, and WIOA program services	✓	0
<b>Other</b>					
Leadership training, Pilot 6	Any education training program not based in a school setting (definition for education).	A curriculum of nation-building theory as well as history, language, tribal government was taught.	Assessment and ISP, wraparound services, and referrals	✓	✓
TOP®, Pilot 7	Promotes the positive development of adolescents through curriculum-guided, interactive group discussions; positive adult guidance and community support; and community service learning.	TOP® curriculum was administered as part of a school-year program in addition to other P3 programming.	Tutoring, mentoring, and academic counseling	✓	✓

Table B.5 (continued)

	Definition of intervention	Description of intervention assessed	Counterfactual	Adequate description	No confounding factors
Two-generation education and training program, Pilot 8	<p>Mentoring. Programs that match mentors with youth or adults who may benefit from an ongoing relationship with a responsible adult.</p> <p>PAT. A home visiting model that promotes the early development, learning, and health of young children by supporting parents and caregivers.</p> <p>Subsidized employment. Programs create or support work opportunities for individuals who would not otherwise be employed. They may provide short-term work and income opportunities, and/or improve employability by offering training, work supports, and transition services to support participants as they move into unsubsidized placements.</p>	<p>Mentors work with participants to set education goals and support them in developing job skills. Mentors also serve as work-site liaisons and lead group sessions.</p> <p>Mentors provided two 90-minute home visits per month using the PAT curriculum to cover topics related to self-advocacy, child development, and parenting.</p> <p>Mothers in the program received subsidized part-time job placements as Head Start literacy coaches.</p>	Only standard Head Start services, including education, health, dental, mental health, nutrition, and family-engagement components	✓	✓

Note: Table does not include the pre-post study. A description was determined to be adequate if it provided sufficient detail to clarify the nature of the intervention/treatment being assessed. A checkmark (✓) indicates that the description was adequate or no confounding factor was identified, and a 0 indicates that a confounding factor existed. Definition is taken from CLEAR (DOL 2019) except for TOP® and PAT, which were taken from the model developer because they were not described in CLEAR.

**Table B.6. Characteristics of RCTs**

Intervention studied	Outcomes studied	Low attrition (Year 1)	Low attrition (Year 2)	No reassignment	Baseline equivalence in analytic sample	Assessment of rigor
<b>Education</b>						
Case management (only), Pilot 1	Suspensions, unexcused absences, reading courses passed, state-mandated online course graduation requirement completed	✓	n.a.	✓	✓	High
Case management (only), Pilot 1	Family referral to adult education	—	n.a.	✓	✓	Moderate
Case management (only), Pilot 1	Suspensions, unexcused absences, reading courses passed, state-mandated online course graduation requirement completed, FAFSA completed, adult education referral	✓	✓	✓	✓	High
Case management and soft-skills training, Pilot 4	Attendance	✓	n.a.	✓	✓	High
Case management and soft-skills training, Pilot 4	Grade promotion, passing a state-mandated test	✓	n.a.	✓	✓	High
Case management and soft-skills training, Pilot 4	Completing school	—	n.a.	✓	✓	Moderate
Leadership training, Pilot 6	Cultural knowledge assessment, motivation for school, decrease in number of risk factors for dropping out of school, number of referrals for poor behavior, loss of school credit, grade point average	✓	n.a.	✓	✓	High
<b>Employment</b>						
Case management (only), Pilot 1	Summer youth employment, employability skills test score	✓	✓	✓	✓	High
	Summer youth employment	n.a.	—	✓	✓	Moderate

Table B.6 (continued)

Intervention studied	Outcomes studied	Low attrition (Year 1)	Low attrition (Year 2)	No reassignment	Baseline equivalence in analytic sample	Assessment of rigor
<b>Other</b>						
Case management (only), Pilot 1	Program attendance, housing stability, juvenile justice involvement, family participation in school event	✓	✓	✓	✓	High

Notes: Numbers for case management (only) indicate different interventions (see also Table B.7). A checkmark (✓) indicates that the condition was met; a minus sign (–) indicates it was not.

n.a. = not applicable; the research did not examine outcomes in these time periods.



**Table B.7. Baseline equivalence for analytic sample**

Intervention studied	Outcomes studied	Baseline equivalence on characteristics	Baseline equivalence on outcome measures	Controls for baseline equivalence outcomes
<b>Education</b>				
Case management (only), Pilot 2*	Returned to secondary education	✓	n.a.	n.a.
Case management (only), Pilot 3*	Absences, number of suspensions, ever suspended	✓	✓	✓
TOP@*, Pilot 7	Passing English course, absences	✓	✓	✓
TOP@*, Pilot 7	Passing English course, absences	✓	✓	✓
Case management and WIOA services, Pilot 5	Attaining high school diploma or equivalent, high school skills test level gain, completing two GED development tests, attaining college course readiness, attaining GPA of 2.0	✓	n.a.	n.a.
<b>Employment</b>				
Case management (only), Pilot 2*	Skills training participation, subsidized employment participation, employed at end of program	✓	n.a.	n.a.
Case management (only), Pilot 3*	WIOA services received	✓	n.a.	n.a.
Case management and WIOA services, Pilot 5	Attaining unsubsidized employment, job-readiness training completion, career education class completion, paid internship completion	✓	n.a.	n.a.

Table B.7 (continued)

Intervention studied	Outcomes studied	Baseline equivalence on characteristics	Baseline equivalence on outcome measures	Controls for baseline equivalence outcomes
<b>Other</b>				
Case management (only), Pilot 3*	Services received from partner	✓	n.a.	n.a.
Two-generation education and training program*, Pilot 8	Cohort 1: Child development (literacy scores, cognitive scores, social- emotional scores, and Head Start attendance); public benefit receipt for parents (WIC, TANF, SNAP)	✓	✓	✓
Two-generation education and training program*, Pilot 8	Cohort 2: Child development (literacy scores, cognitive scores, social-emotional scores); public benefit receipt for parents (WIC, TANF, SNAP)	✓	✓	✓

Notes: Numbers for case management (only) indicate different interventions (see also Table B.6). A checkmark (✓) indicates that the condition was met; a minus sign (–) indicates it was not.

n.a. = not applicable. For example, if an outcome measure such as returning to secondary education was not assessable at baseline, the local evaluation report did not have to establish baseline equivalence for this outcome measure.

\*Satisfied baseline equivalence by controlling for statistically significant differences in analysis.

**Table B.8A. Significance of findings for interventions with high or moderate levels of rigor: Case management interventions**

Intervention	Outcome	Period after program enrollment	Rigor	Findings	Analysis	Difference in means
<b>Education</b>						
Case management, Pilot 1	Suspensions	1 year	High	—	Regression coefficient	0.09
Case management, Pilot 1	Suspensions	2 years	High	✓	Regression coefficient	-0.10
Case management, Pilot 1	Unexcused absences	1 year; 2 years	High	0	Regression coefficient	-0.03; -0.04
Case management, Pilot 1	State-mandated online course graduation requirement completed	1 year	High	0	Regression coefficient	0.1
Case management, Pilot 1	State-mandated online course graduation requirement completed	2 years	High	✓	Regression coefficient	0.07
Case management, Pilot 1	Family referred to adult education	2 years	High	—	Regression coefficient	-0.09
Case management, Pilot 1	Reading classes passed	1 year; 2 years	High	0	Regression coefficient	-0.02; -0.05
Case management, Pilot 1	Family referred to adult education	1 year	Moderate	0	Regression coefficient	0.02
Case management, Pilot 1	Family referred to adult education	2 years	High	0	Regression coefficient	0.00
Case management, Pilot 2	Returned to secondary education	Within 1 year	Moderate	✓	Regression adjusted means	0.21
Case management, Pilot 3	Number of absences	1 year	Moderate	0	Raw means	0.31
	Number of suspensions	1 year	Moderate	0	Raw means	-0.59
Case management, Pilot 3	Ever suspended	1 year	Moderate	0	Raw means	-0.07
Case management and soft-skills training, Pilot 4	Attendance	1 year	High	0	Regression adjusted means	-1.63
Case management and soft-skills training, Pilot 4	Grade promotion	1 year	High	0	Regression adjusted means	0.01
Case management and soft-skills training, Pilot 4	Passing state-mandated test	1 year	High	0	Regression adjusted means	0.09

Table B.8 (continued)

Intervention	Outcome	Period after program enrollment	Rigor	Findings	Analysis	Difference in means
Case management and soft-skills training, Pilot 4	Completing high school	1 year	Moderate	0	Raw means	0.00
Case management and soft-skills training, Pilot 4	Attendance	1 year	Moderate	0	Raw means	-1.63
Case management and WIOA services, Pilot 5	Attaining high school diploma or equivalent	9 months	Moderate	0	Weighted ANCOVA model	0.31
Case management and WIOA services, Pilot 5	Attaining GPA of 2.0	9 months	Moderate	✓	Weighted ANCOVA model	0.09
Case management and WIOA services, Pilot 5	Completing two GED development tests	9 months	Moderate	✓	Weighted ANCOVA model	0.29
Case management and WIOA services, Pilot 5	Attaining college course readiness	9 months	Moderate	✓	Weighted ANCOVA model	0.04
Case management and WIOA services, Pilot 5	High school skills test level gain	9 months	Moderate	0	Weighted ANCOVA model	-0.02
<b>Employment</b>						
Case management, Pilot 1	Employability skills test score	1 year and 2 years	High	✓	Regression coefficient	0.12; 0.13
Case management, Pilot 1	Summer youth employment	1 year	High	0	Regression coefficient	0.12
Case management, Pilot 1	Summer youth employment	2 years	Moderate	0	Regression coefficient	-0.09
Case management, Pilot 2	Career preparation participation	Within 1 year	Moderate	—	Regression adjusted means	-0.07
Case management, Pilot 2	Subsidized employment participation	Within 1 year	Moderate	—	Regression adjusted means	-0.21
Case management, Pilot 2	Employed at end of program	Within 1 year	Moderate	✓	Regression adjusted means	0.20
Case management, Pilot 3	WIOA service receipt	1 year	Moderate	0	Raw means	0.02
Case management and WIOA services, Pilot 5	Attaining unsubsidized employment	9 months	Moderate	0	Weighted ANCOVA model	0.20

Table B.8 (continued)

Intervention	Outcome	Period after program enrollment	Rigor	Findings	Analysis	Difference in means
Case management and WIOA services, Pilot 5	Job-readiness training completion	9 months	Moderate	✓	Weighted ANCOVA model	0.27
Case management and WIOA services, Pilot 5	Career education class completion	9 months	Moderate	✓	Weighted ANCOVA model	0.11
Case management and WIOA services, Pilot 5	Paid internship completion	9 months	Moderate	✓	Weighted ANCOVA model	0.21
<b>Other</b>						
Case management, Pilot 1	Family participation in school event	1 year and 2 years	High	✓	Regression coefficient	0.22; 0.09
Case management, Pilot 1	Program attendance	1 year	High	✓	Regression coefficient	0.10
Case management, Pilot 1	Stable housing	1 year and 2 years	High	0	Regression coefficient	0.02; -0.01
Case management, Pilot 1	Juvenile justice involvement	1 year and 2 years	High	0	Regression coefficient	-0.02; -0.02
Case management, Pilot 3	Services received from partner	1 year	Moderate	✓	Raw means	0.15

**Table B.8B. Significance of findings for interventions with high or moderate levels of rigor: Other interventions**

Intervention	Outcome	Period after program enrollment	Rigor	Findings	Analysis	Difference in means
<b>Education</b>						
Leadership training, Pilot 6	Cultural knowledge	1 month	High	0	OLS regression	0.27
Leadership training, Pilot 6	Motivation for school	1 month	High	0	OLS regression	-0.25
Leadership training, Pilot 6	Decreased risk for dropping out of school	1 month	High	0	OLS regression	-0.62
Leadership training, Pilot 6	School behavior	1 month	High	0	OLS regression	0.05
Leadership training, Pilot 6	Loss of school credit	1 month	High	0	OLS regression	0.33
Leadership training, Pilot 6	Grade point average	1 month	High	0	OLS regression	0.33
TOP®, Pilot 7, Cohort 1	Passing English course	1 year	Moderate	0	HLM	-0.03
TOP®, Pilot 7, Cohort 1	Absences	1 year	Moderate	0	HLM	1.05
TOP®, Pilot 7, Cohort 2	Passing English course	1 year	Moderate	0	HLM	-0.002
TOP®, Pilot 7, Cohort 2	Absences	1 year	Moderate	0	HLM	-2.96

Table B.8 (continued)

Intervention	Outcome	Period after program enrollment	Rigor	Findings	Analysis	Difference in means
<b>Other</b>						
Two-generation education and training program, Pilot 8, Cohort 1	Child literacy scores	3 months	Moderate	0	Regression adjusted means	0.78
Two-generation education and training program, Pilot 8, Cohort 1	Child literacy scores	9 months	Moderate	0	Regression adjusted means	2.62
Two-generation education and training program, Pilot 8, Cohort 1	Child literacy scores	1 year	Moderate	0	Regression adjusted means	-1.29
Two-generation education and training program, Pilot 8, Cohort 1	Child's cognitive scores	3, 9, and 12 months	Moderate	0	Regression adjusted means	0.67; 1.61; -0.81
Two-generation education and training program, Pilot 8, Cohort 1	Child's social-emotional scores	3, 9, and 12 months	Moderate	0	Regression adjusted means	0.60; 1.17; 0.21
Two-generation education and training program, Pilot 8, Cohort 1	Child Head Start attendance	1 month	Moderate	✓	Regression adjusted means	6.07
Two-generation education and training program, Pilot 8, Cohort 1	Child Head Start attendance	5 months	Moderate	0	Regression adjusted means	-1.66
Two-generation education and training program, Pilot 8, Cohort 1	Child Head Start attendance	7 months	Moderate	0	Regression adjusted means	6.29
Two-generation education and training program, Pilot 8, Cohort 1	Public benefit receipt–WIC	11 months	Moderate	0	Regression adjusted means	0.10
Two-generation education and training program, Pilot 8, Cohort 1	Public benefit receipt–TANF	11 months	Moderate	✓	Regression adjusted means	0.11
Two-generation education and training program, Pilot 8, Cohort 1	Public benefit receipt–SNAP	11 months	Moderate	0	Regression adjusted means	0.03
Two-generation education and training program, Pilot 8, Cohort 2	Child literacy scores	3, 9, and 12 months	Moderate	0	Regression adjusted means	1.64; -0.57; 1.50
Two-generation education and training program, Pilot 8, Cohort 2	Child cognitive scores	3, 9, and 12 months	Moderate	0	Regression adjusted means	3.52; 1.68; 3.88
Two-generation education and training program, Pilot 8, Cohort 2	Child social-emotional scores	3, 9, and 12 months	Moderate	0	Regression adjusted means	3.00; 2.77; 2.95
Two-generation education and training program, Pilot 8, Cohort 2	Child Head Start attendance	5 and 7 months	Moderate	0	Regression adjusted means	2.57; 6.27

Table B.8 (*continued*)

Notes: Italics represent secondary outcomes. A checkmark (✓) indicates that the local evaluation found that the intervention achieved expected impacts at the 5 percent significance level. A 0 indicates that impacts were not significant at the 5 percent level, and a minus sign (–) indicates that the impacts were in the opposite direction at the 5 percent significance level.



## **Appendix C**

### **Determining Evidence Ratings**

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

To answer the second research question—What is the level of rigor in the local evaluations?—we constructed an evidence rating that was informed by those used by the federal What Works Clearinghouse (WWC 2017), Clearinghouse for Labor Evaluation and Research (Department of Labor 2019), and Home Visiting Evidence of Effectiveness (HomVEE) (Sama-Miller 2018). The three-tiered evidence ratings describe the ability of a study to produce an internally valid estimate of an intervention’s impact.

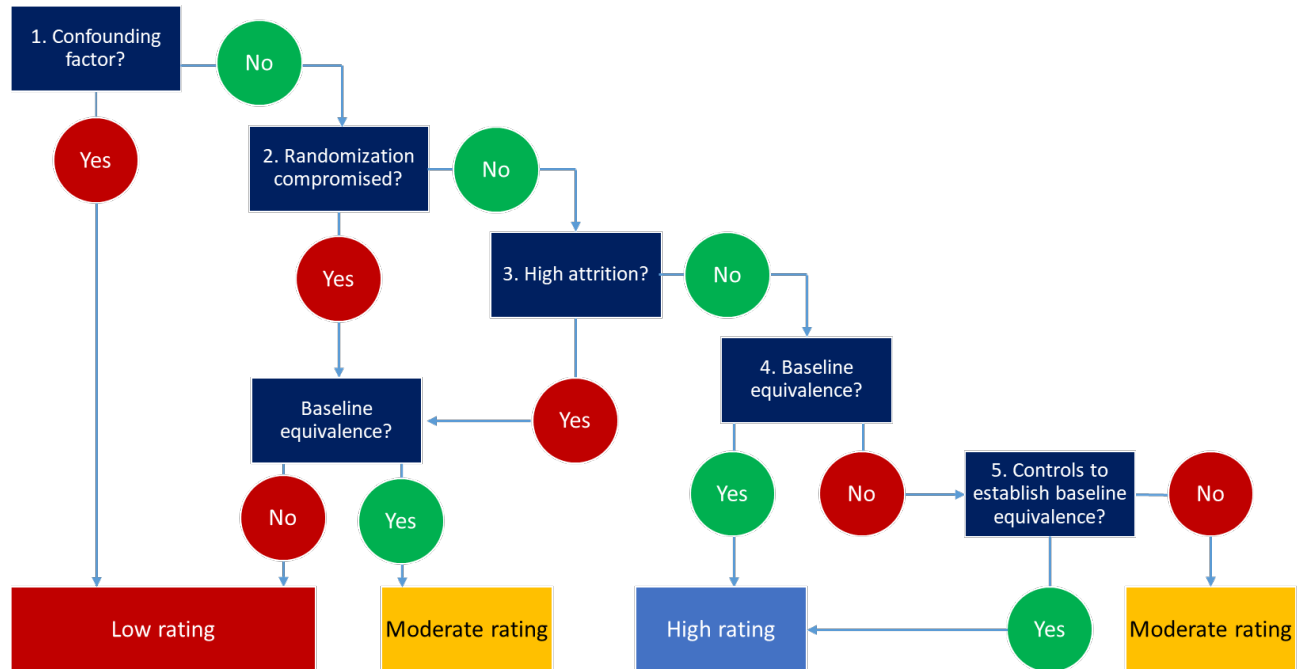
- A high rating means that strong evidence exists that the effects estimated in the local evaluation are attributable to the intervention examined.
- A moderate rating means that evidence exists that the effects estimated in the local evaluation are likely attributable at least in part to the intervention examined. However, other factors that were not accounted for in the study may also have contributed.
- A low rating means that we cannot determine whether the effects estimated in the local evaluation are attributable to the intervention examined, because other factors are likely to have contributed to the impact. A low rating does not imply that the study’s results are not useful for some purposes, but they should be interpreted with caution.

The determination of the evidence rating a local evaluation received depends, in part, on whether the design is a randomized controlled trial (RCT) or quasi-experimental design (QED).

RCTs received one of three evidence ratings. An RCT rated as high had low attrition and no issues with randomization. Low attrition meant that few people in the treatment or comparison group left the study, or the differences in rates between the groups was small, and no reassignment meant that study participants were not switched from the study condition (treatment or comparison) to which they were assigned at baseline. An RCT rated as moderate had issues with randomization or high attrition but demonstrated baseline equivalence. An RCT rated as low had a confounding factor or did not demonstrate baseline equivalence and needed to do so.

Figure C.1 summarizes the steps we took to determine the evidence rating for an RCT. As the figure shows, RCTs are eligible for the highest evidence rating if the research meets the stringent standards needed to unambiguously establish causality. Steps 1 and 2 apply to the study as a whole, and Steps 3 through 5 apply to each outcome separately, such that each outcome may receive a different rating. Any RCT with a confounding factor automatically receives a low rating, because the presence of a confounding factor makes it impossible to separate the effect of the intervention from the effects of other factors. If no confounding factor is present, the RCT must ensure that randomization was not compromised. If randomization is compromised, the presumed baseline equivalence from randomization might be eliminated. Compromised randomization could involve movement between the treatment and the comparison groups or could involve adding new participants after random assignment. An RCT with compromised random assignment would be reviewed as a QED and would be eligible for only a moderate evidence rating. The same is true for an RCT with a high level of attrition from the study, because it undermines the assumption that members of the treatment and comparison groups are equivalent. We applied the conservative attrition rules specified by the WWC standards to calculate overall attrition rates for each outcome measure. Lastly, those RCTs that meet all other criteria but fail to establish baseline equivalence would be rated moderate.

**Figure C.1. Flowchart for rating RCT designs**

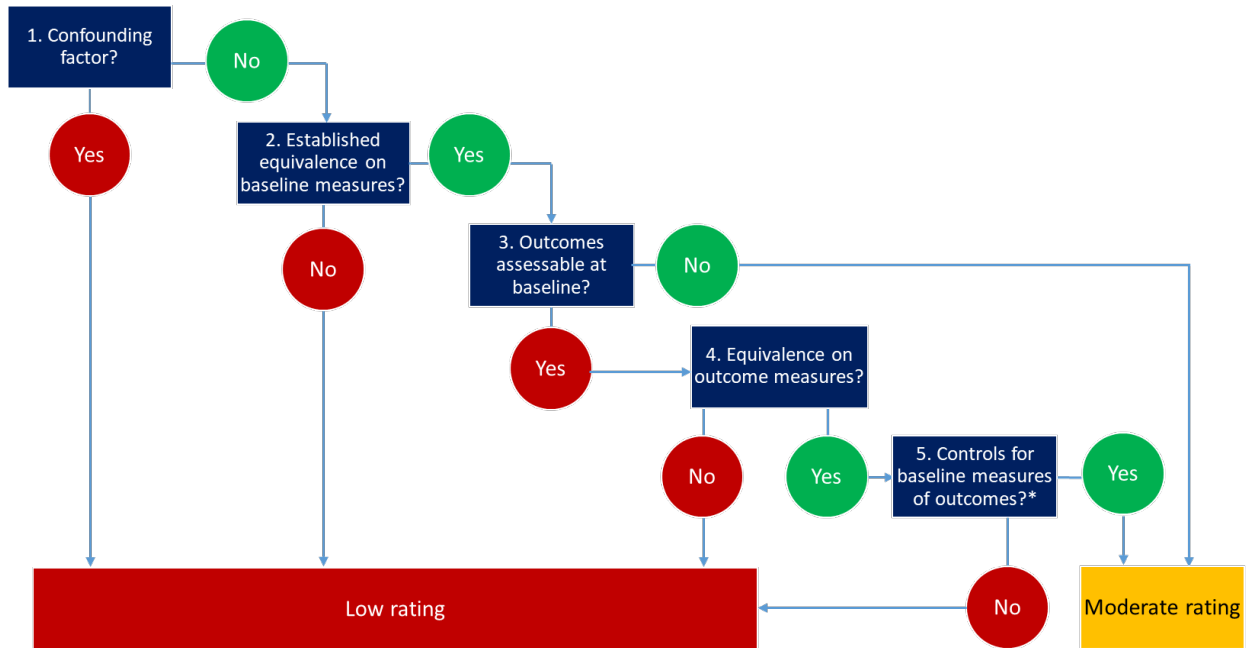


Source: Adapted from Sama-Miller et al. (2018).

QEDs received one of two evidence ratings. A QED rated as moderate has either (1) baseline equivalence by demonstrating no statistically significant differences between the treatment and control groups for characteristics shown in Appendix B, Table B.3 and on outcome measures included in Appendix B, Table B.4, or (2) controlled for differences in baseline measures in their analysis. A QED rated as low has a confounding factor or did not establish baseline equivalence.

Local evaluations implementing a QED or an RCT with a high level of attrition or compromised random assignment are eligible for a moderate evidence rating. Figure C.2 summarizes the steps we took to determine whether the QED-reviewed studies received a moderate or low rating. The presence of a confounding factor means the study received a low rating. If a study has no confounding factors, we verified baseline equivalence for measures shown in Appendix B, Table B.3. The evaluation must establish baseline equivalence on outcome measures by demonstrating that there are no statistically significant differences in outcomes of interest (listed in Appendix B, Table B.4) that were collected at baseline (if available) or by controlling for any differences in those measures in their analysis. Without baseline equivalence or controls for differences in outcomes at baseline (if available), the study received a low rating.

Figure C.2. Flowchart for rating QED studies



Source: Adapted from Sama-Miller et al. (2018).

\*If feasible to collect at baseline.

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

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

---

**Mathematica**

Princeton, NJ • Ann Arbor, MI • Cambridge, MA  
Chicago, IL • Oakland, CA • Seattle, WA  
Tucson, AZ • Woodlawn, MD • Washington, DC

**EDI Global, a Mathematica Company**

Bukoba, Tanzania • High Wycombe, United Kingdom

**Mathematica**

Progress Together

[mathematica.org](https://www.mathematica.org)