

## Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

### INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at <https://eric.ed.gov/submit/> and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

### GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID

Publication/Completion Date—(if *In Press*, enter year accepted or completed)

Check type of content being submitted and complete one of the following in the box below:

- If article: Name of journal, volume, and issue number if available
- If paper: Name of conference, date of conference, and place of conference
- If book chapter: Title of book, page range, publisher name and location
- If book: Publisher name and location
- If dissertation: Name of institution, type of degree, and department granting degree

DOI or URL to published work (if available)

**Acknowledgement of Funding**— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

“This work was supported by U.S. Department of Education [Office name]   
through [Grant number]  to Institution] . The opinions expressed are  
those of the authors and do not represent views of the [Office name]   
or the U.S. Department of Education.

# Using Iterative Experimentation to Accelerate Program Improvement: A Case Example

Evaluation Review  
2022, Vol. 46(5) 469–516  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0193841X20923199  
journals.sagepub.com/home/erx



Rebecca A. Maynard<sup>1</sup> , Rebecca N. Baelen<sup>1</sup>,  
David Fein<sup>2</sup>, and Phomdaen Souvanna<sup>3,4</sup>

## Abstract

This article offers a case example of how experimental evaluation methods can be coupled with principles of design-based implementation research (DBIR), improvement science (IS), and rapid-cycle evaluation (RCE) methods to provide relatively quick, low-cost, credible assessments of strategies designed to improve programs, policies, or practices. This article demonstrates the feasibility and benefits of blending DBIR, IS, and RCE practices with embedded randomized controlled trials (RCTs) to improve the pace and efficiency of program improvement. This article describes a two-cycle experimental test of staff-designed strategies for improving a workforce development program. Youth enrolled in Year Up's Professional Training Corps (PTC) programs were

---

<sup>1</sup> University of Pennsylvania, Philadelphia, PA, USA

<sup>2</sup> Social and Economic Policy, Abt Associates, Rockville, MD, USA

<sup>3</sup> Abt Associates, Rockville, MD, USA

<sup>4</sup> Brandeis University, Waltham, MA, USA

## Corresponding Author:

Rebecca A. Maynard, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104, USA.

Email: rmaynard@upenn.edu

randomly assigned to “improvement strategies” designed to boost academic success and persistence through the 6-month learning and development (L&D) phase of the program, when participants spend most of their program-related time in courses offered by partner colleges. The study sample includes 317 youth from three PTC program sites. The primary outcome measures are completion of the program’s L&D phase and continued college enrollment beyond the L&D phase. The improvement strategies designed and tested during the study increased program retention through L&D by nearly 10 percentage points and increased college persistence following L&D by 13 percentage points. Blending DBIR, IS, and RCE principles with a multi-cycle RCT generated highly credible estimates of the efficacy of the tested improvement strategies within a relatively short period of time (18 months) at modest cost and with reportedly low burden for program staff.

**Keywords**

social experiment, randomized controlled trial, design-based implementation research, improvement science, career preparation, youth development

**Introduction and Framing**

The new normal for public policy in the United States and, increasingly, in other parts of the world is reliance on evidence to support decisions about which programs, policies, and practices (hereafter referred to as programs) to support in order to improve target outcomes and/or lower costs of achieving program goals. Under ideal conditions, supporting evidence on the expected effectiveness of programs under consideration will come from randomized controlled trials (RCTs) or other evaluation designs that produce credible impact estimates. However, program developers and operators rarely have access to or experience acquiring such evidence to support their decision-making. Credible evidence of program effectiveness (when it exists) is often derived from moderate- to large-scale experimental evaluations, often initiated by teams of external evaluators years before the study’s findings are released.

This article draws on a recent case example of embedding experimental evaluations of staff-developed program improvement strategies into routines of practice. It illustrates how traditional experimental approaches to program

evaluation can be blended with principles of design-based implementation research (DBIR), improvement science (IS), and rapid-cycle evaluation (RCE) strategies to accelerate the pace and precision of developing and refining interventions to improve program outcomes (see Table 1).

### *Traditional Program Evaluation*

By the turn of this century, there was a robust, multidisciplinary evaluation community in the United States experienced in conducting large-scale, experimental-design program evaluations. Evaluators had a 35-year history of social experimentation to guide their evaluation designs, beginning with the National Health Insurance Experiment, a suite of Negative Income Tax Experiments, and the Housing Allowance Demand Experiments conducted in the 1960s and early 1970s (Gueron & Rolston, 2013; Kennedy, 1980; Munnell, 1987; Newhouse, 1993). Today, rhetoric throughout all branches of the federal government, in many state and local government agencies, and among philanthropic organizations is calling for rigorous evidence to support policy and program funding decisions (Haskins & Margolis, 2014; Maynard et al., 2016).

The emphasis on evidence has propelled the widespread use of RCTs and various quasi-experimental methods to evaluate new programs and assess changes to existing ones. Until recently, most social experiments and other major program evaluations have focused on programs and program variations designed by evaluators, program developers, or a federal or state governing agency. The evaluations have generally focused on assessing “mature” programs, and their primary goal has been to judge whether continued program funding or funding for a program enhancement is justified. Toward this end, these evaluations often involve multiple study sites, large samples and long-term follow-up. Often, they are costly and result in study findings that released years after the evaluation began and, not infrequently, after interest in them has waned.

### *Program- and Practice-Centered Evaluation Methods*

For some time, it has been common in business and medicine to conduct experimental tests of existing practices versus variations of existing practices to determine their potential to improve outcomes and/or reduce costs (Manzi, 2012). Commonly, these experimental tests entail large-scale, low-cost, and rapid-cycle testing of tweaks to routine practices. For example, they may test alternative approaches to product placement in retail stores,

**Table 1.** Comparison of Approaches to Evaluation Design and Implementation Used in the Case Example.

Primary Actors/Sources of Evidence	Case Example: Mini-Experiment: "IS With a Twist"	Traditional Program Evaluation	Design-Based Implementation Research	Improvement Science (IS)	Rapid-Cycle Evaluation
Identification of research goals and questions					
Evaluator and/or funder		X			
Practitioner/program lead	X		X		
Evaluator and practitioner	X			X	X
Design of strategies for testing					
Evaluator and/or funder		X			
Practitioner/program lead	X			Primary	
Evaluator and practitioner			X	Secondary	X
Implementation of strategies for testing					
Usual program staff	X	X		X	X
Specially trained and supported staff		X	X		
Monitoring and data collection					
Evaluator		Extensive	Extensive	Extensive	Light
Program staff	Moderate to Light	Moderate	Extensive	Extensive	Moderate to Light
Evidence of effectiveness					
Professional judgments		Secondary	Primary	Primary	
Systematic observation and documentation	Secondary	Secondary	Primary	Primary	Secondary
Surveys and focus groups	Secondary	Secondary			
Experimental testing	Primary	Primary			Primary

Note. For further description of the various evaluation methods, see Appendix Table A1.

nudges to reduce transmission of infections in hospitals, or variations in content or modes of disseminating marketing and product safety information.

Until recently, program implementers in the fields of education and social welfare have taken on primary responsibility for program improvement efforts, relying mostly on professional judgments to guide decisions about practice adoption, revision, or abandonment. Typically, efforts to evaluate program-initiated improvement strategies have relied on methods such as DBIR and IS—both of which are adaptations of the Deming’s Continuous Quality Improvement (CQI) approach (Zuchowski et al., 2019). Both rely primarily on qualitative approaches to assess participant needs, explore the logic behind proposed improvement strategies, and measure participant responses to universal implementation of the changes. DBIR and IS evaluations generally do not use experimental designs for estimating participants’ response to improvement strategies.

### *Rapid-Cycle Experimental Evaluation of Improvement Strategies*

In the past few years, there have been some interesting examples of practitioner-led, evaluator-supported efforts to accelerate the pace and lower the cost of program improvement by embedding experimental tests of promising strategies into routines of practice. This approach relies on empirical evidence rather than professional judgment for assessing the efficacy of tested strategies. Like traditional program evaluations, these rapid-cycle experimental evaluations generally rely on systematic observations of strategies being implemented and tested to supplement the experimental test. Several of these practitioner-led and evaluator-supported efforts have sought to alter immediately observable outcomes such as program application, homework completion, and attendance (Christie et al., 2017; Cody & Asher, 2014; Daily et al., 2018; Dechausay & Anzelone, 2016; Ross et al., 2018).

This article provides a case example that blends DBIR and IS methodologies with a series of embedded experiments designed to test program developed improvement strategies intended to boost academic success and program retention of disadvantaged youth enrolled in Year Up’s yearlong Professional Training Corps programs. The study focuses on the first 6 months of the program, termed the learning and development (L&D) phase, when youth are enrolled full time in college courses—some basic skills courses and others technical courses aligned with career training tracks, such as information technology (IT) or cyber security.

The improvement strategies tested in the evaluation were developed collaboratively by the evaluation team and PTC program staff in about three months. They were then implemented with a random subset of participants in the next cohort of participants enrolled in the program (first cycle of testing). Based on experiences in the first testing cycle, the strategies were quickly tweaked following cycle 1 before they were used with the next cohort of participants enrolled (second cycle of testing). The full evaluation results from the study were shared with program staff within a few months of the time the second cohort of participants was scheduled to complete L&D.

Below, we provide background for the case example and discuss the motivation and approach to developing the strategies to be tested. Next, we describe the study design, the sample, and the data. We conclude by presenting the study findings and discussing their implications and limitations.

## **Background for the Case Example**

This case example is an improvement study that was part of a larger evaluation of Year Up's PTC program. The PTC is a yearlong career development program for disadvantaged 18- to 24-year-old high school graduates. During the first 6-months of the program, termed the L&D phase, participants engage full time in basic and technical skills courses offered through college partners. They also participate in a professional skills (pro-skills) course and other career support activities provided by Year Up staff. Youth who successfully complete L&D advance to internship, where they work 4 days a week in employer-sponsored internships and spend 1-day a week with Year Up staff building pro-skills and preparing for a full-time schedule of career-focused employment, which is sometimes paired with continued college coursework. Throughout the program, youth are assigned to learning communities (LCs) of about 20 youth where they receive mentoring from Year Up staff and opportunities to engage in myriad skill development and support activities with peers in their same career training track.

The PTC is an adaptation of Year Up's original program model, which targets the same population of youth but serves them in stand-alone programs where Year Up staff provide all the basic and technical skills training in addition to the pro-skills training, internships, and LC supports. PTC programs partner with area technical colleges in the hopes of achieving participant impacts comparable to the core program (Fein & Hamadyk, 2018; Roder & Elliott, 2014), but at substantially lower costs—primarily by relying on the colleges to provide the basic and technical skills training. Because the PTC programs are located on the college campus rather than in

stand-alone offices, participants also can tap into various support services offered by the college.

Like Year Up's core program, the PTC has high expectations for youth but also offers high levels of support. The theory of change underpinning both program models is grounded in a robust body of research, which indicates the importance of acquiring both professional and occupational skills as well as the necessity of supports and feedback to help youth find and succeed in career-oriented jobs (Belfield & Bailey, 2011; Jacobson et al., 2005; Jacobson & Mokher, 2009; Jepsen et al., 2009; Kane & Rouse, 1995; Marcotte, 2010). Year Up's high-expectation and high-support model builds off evidence that strong monitoring and support allows for the development of basic and soft skills necessary for succeeding in college coursework and internships (Bailey et al., 2015; Bloom & Miller, 2018; Casazza & Silverman, 2013; Hagedorn & Kuznetsova, 2016; Trekson, 2016; Weiss et al., 2015; Weissman et al., 2012).

The program aims to improve the labor market success of participants in three ways. Youth acquire professional skills through a pro-skills course as well as through daily modeling of and feedback on professional workplace behaviors—enforced through Year Up's code of conduct (the “participant contract”). Participants develop occupation specific competencies through college courses and internships experiences (Jaggars, 2011; Perin, 2011), and youth receive youth “high support—high feedback” through active coaching and support services provided by program staff and through engagement in learning communities. Participants receive weekly feedback and support in the context of a designated learning community of peers and Year Up staff (Sommo et al., 2012; Visser & Stern, 2015).

Since the inception of PTC in 2011, Year Up leadership has been keen to conduct an independent evaluation of the PTC that, ultimately, examines its overall impacts. However, its priority early on was to identify and address implementation challenges that arose with the new model, and this was the motivation behind the improvement study highlighted in this article.

## **Motivation and Approach to the Improvement Study**

In 2014, Year Up received funding from the Social Innovation Fund (SIF)/GreenLight Fund to develop and evaluate one of the PTC programs located in the mid-Atlantic region (Site 1) using a RCT. An early finding of this evaluation was a concerning low rate of retention during L&D—the first 6-months of the 12-month program. Participants who failed to complete L&D were not eligible to advance to internships for the second 6 months,

which was viewed as a critical to prepare participants for jobs with advancement potential. As a result, Year Up and the evaluation team prioritized conducting an improvement study designed to examine patterns and sources of attrition as well as to test promising strategies for boosting academic success and retention during L&D.

Recognizing that concerns about academic performance and persistence during L&D were issues for most of the PTC programs, Year Up National and the evaluation team included two additional PTC programs as part of the improvement study to develop and test the efficacy of academically focused improvement strategies. Importantly, the improvement strategies were to be developed largely by local program staff through a process with similarities to those used in DBIR and IS.

One of the added programs is in the Northeast (Site 2) and the third is in the Southeast (Site 3).<sup>1</sup> The decision to include three programs in the study provided a larger, more diverse staff to plan the improvement strategies for testing, as well as a larger sample and more diverse contexts for testing them.

The Year Up National staff established a working group across the three sites to coordinate planning the improvement strategies and to work with the evaluation team to plan approaches to implementation that would maximize the credibility and usefulness of the study findings. Early conversations with program staff revealed that much of participant attrition occurred near the end of L&D and often was directly tied to participants failing their college courses. A contributing factor was that, unlike in its stand-alone programs, in the PTC programs, Year Up staff lacked control and oversight of the college courses. This meant that PTC staff frequently learned of participants' difficulties in their courses after it was too late to help them recover.

Consistent with Year Up's emphasis on creating a program model that was financially sustainable and its commitment to rapid improvement, the improvement strategies tested needed to (1) be minimally burdensome to implement, (2) require little additional cost to implement, (3) require limited time to implement and test (one or two cycles of operation), and (4) hold high promise for improving academic performance and persistence. For practical reasons, they also needed to be ones that lent themselves to rigorous evaluation of their effectiveness.

The evaluation team worked with Year Up National staff and staff at the three study sites to develop and test the improvement strategies. Program staff felt strongly that the strategies should emphasize early identification of the participants who were facing academic challenges and quick responses to the specific challenges faced. Within these broad parameters, the evaluation staff facilitated "brainstorming" sessions with program staff to guide

their development of promising strategies tailored to local circumstances. Specifically, local program staff were empowered to revisit site-specific strategies they had considered and/or tried previously as well as to explore new strategies or promising refinements to existing approaches. The evaluation team largely listened but occasionally shared novel evidence-based strategies aligned with site-specific needs. They also aided the programs by documenting decisions in ways that were useful for evaluation and programmatic purposes.

Throughout this process, the evaluation team drew on principles and practices of DBIR and IS for identifying and designing approaches for implementing the improvement strategies to be tested (Bryk et al., 2011; Bryk et al., 2015; Means & Penuel, 2005; Proger et al., 2017). The team also used rigorous experimental principles and practices to inform methods for testing the effectiveness of those strategies (Orr, 1999). At the same time, the team was mindful of the imperative for rapid testing of strategies and acting on the results of early experiences to further refine the strategies (Cody & Asher, 2014). Importantly, the evaluation team ensured its objectivity in conducting the impact evaluation by having a fire-wall between the analyst who had access to the administrative data used in the impact analysis and the team who worked with the PTC program staff on the design, implementation and monitoring of the improvement strategies.

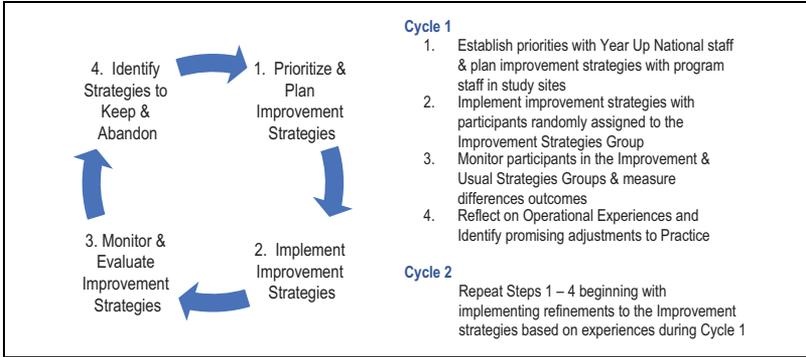
The work proceeded in three stages: (1) roughly three months for planning the improvement strategies to be tested; (2) a year to implement and test the strategies over two cycles of L&D, with opportunities for modification between cycles based on early implementation experiences; and (3) roughly 6 months for rigorously assessing the effectiveness of the strategies for improving L&D retention (Table 2). The implementation and testing of the improvement strategies in stage two followed methods used in “Plan-Do-Study-Act” (PDSA) cycles with each cycle of participant cohorts (Bryk et al., 2015; Tichnor-Wagner et al., 2017). Program staff made strategic adjustments to the improvement strategies between Step 4 of Cycle 1 and Step 1 of Cycle 2 (Figure 1). After completing two cycles of testing, the evaluation team analyzed the outcomes data as well as monitoring data gathered throughout the two cycles of testing and reported the findings back to both the participating PTC programs to Year Up National staff.

### *Stage 1: Developing the Improvement Strategies*

Staff at the three study sites worked closely with Year Up National and the evaluation team over several months to identify and prepare to implement the

**Table 2.** Improvement Study Time Line.

Stage 1: Plan	Stage 2: Rollout Strategies in Testable Ways	Stage 3: Evaluate and Communicate
Identify Sites and Strategies	Improvement Cycle 1	Improvement Cycle 2
(July to December 2016)	(January to June 2017)	(July to December 2017)
<ol style="list-style-type: none"> <li>1. Brainstorm strategies</li> <li>2. Plan rollout</li> <li>3. Train staff in design of strategies</li> <li>4. Prepare for intake and random assignment</li> </ol>	<ol style="list-style-type: none"> <li>1. Enroll January 2018 cohort</li> <li>2. Randomly assign participants to treatment condition</li> <li>3. Deliver learning and development (L&amp;D) based on assigned group</li> <li>4. Conduct regular check-ins</li> <li>5. Review and refine improvement strategies</li> </ol>	<ol style="list-style-type: none"> <li>1. Estimate impacts</li> <li>2. Document strategies used</li> <li>3. Interview site staff and participants</li> <li>4. Collect artifacts used for improvement strategies</li> <li>5. Share findings</li> </ol>
		(January to June 2018)



**Figure 1.** The improvement cycles.

improvement strategies for monitoring and supporting participants during L&D. The process entailed three steps. The first step involved the evaluation team engaging local site staff in reflective brainstorming exercises to identify participants’ challenges with academics and uncovering gaps or weaknesses in programming that potentially could address those challenges.

The contributions of local site staff were essential to this process, as staff possess “on-the-ground” knowledge and expertise related to academic challenges commonly faced by participants, know the usual strategies for monitoring academic performance and identifying challenges, and are aware of the difficulties they faced monitoring and responding to support needs. During “brainstorming” sessions, site staff were asked to describe participant challenges with a focus on academics and prompted to share established and common practices used to address them. They also were prompted to talk about their “wish list” of other strategies that might be helpful.

Staff noted a variety of “usual” strategies and supports offered by their PTC program and the local college partner (e.g., weekly stipends, advising/coaching by college and PTC staff, and general support services provided by the PTC program, referral organizations, and the partner college). For example, the brainstorming exercises revealed that under Year Up’s usual practices, there was no clear protocol between the PTC programs and their college partners for monitoring and supporting participants’ academic engagement and performance. They noted that, as a result, Year Up staff and college partners’ responses to participants’ needs varied in timeliness, substance, and quality. Inconsistencies in these areas were exacerbated by college instructors’ varied levels of familiarity with Year Up’s “Participant Contract” (i.e., their code of conduct and point system for managing the

consequences of violations) and the strategies used to monitor and support participants. Efforts to develop effective monitoring and support strategies with college partners were reportedly challenging because college staff were generally unaware of the partnership between the college and Year Up and because of difficulties associated with establishing and enforcing protocols between the two organizations.

Brainstorming sessions with program staff surfaced four related weaknesses in the programs' usual approach to monitoring and supporting participants academically. First, there were gaps and lags in communication between the point-person at the partner college and the instructors that prevented PTC staff from receiving real-time information on participants' academic progress and/or their struggles with coursework. Second, there was a lack of focused attention by PTC staff on academics during coaching time and no accountability of coaches for discussing academic issues with their coachees. Third, there were neither systems nor routines for managing academic information on participants throughout their time in the program, which inhibited collaborative approaches to address and/or track academic challenges. Fourth, there were weaknesses in the systems and supports available to help participants address academic challenges and hold them accountable for using available supports (either within the PTC program or at the partner college).

The second step of the process for developing the improvement strategies for testing involved establishing criteria to respond to the above-mentioned weaknesses in programming and build on untapped resources at the PTC programs and partner colleges. Selected strategies needed to be concrete, modest-to-no-cost improvements to current practices. The goal was for the chosen strategies to be ones that could improve academic success through one of three means: (1) improve access by PTC coaches to participant performance data (e.g., through reaching out to instructors to obtain grade and attendance information), (2) surface and make better use of information on academic performance during coaching, and (3) improve access to and use currently untapped resources and supports to assist participants experiencing academic challenges.

Upon establishing the areas of weakness in programming and the criteria for selecting improvement strategies, Year Up National staff, the evaluation team, and staff at each of the three study sites chose the specific strategies to be tested in their programs. Within the broad parameters outlined above, the improvement strategies adopted could and did vary across programs in large part due to differences in perceived needs and opportunities for improvement (Table 3). However, four of the improvement strategies to be tested in Cycle 1 were common to all three programs.

**Table 3.** Usual and Improvement Strategies Tested (by Site and Cycle).

Improvement Strategy	Cycle 1			Cycle 2		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
<i>Communication and information sharing between Year Up (YU) and college/instructors</i>						
<i>Usual strategy:</i> All instructors invited to meet with YU Staff for Orientation; staff conduct periodic classroom observations to monitor attendance, professional dress, and so on						
<i>Improvement strategy</i>						
<i>Initial plan:</i> Develop and implement prescriptive monitoring protocol to track assignment completion and academic performance indicators, tailor outreach to instructors at beginning of term, and host instructor luncheons	✓	✓	✓	✓	✓	✓
<i>Implementation</i>	✓	✓		✓	✓	
Created shared spreadsheet containing feedback from instructors, sharing of grades varied based on instructor				✓	✓	
Tailored outreach to instructors at beginning of term				✓		
Hosted instructor luncheon				✓	✓	
<i>Modifications:</i> Incorporate information from coaching sessions into shared spreadsheets, lower expectation of collecting grades from instructors				✓	✓	✓

(continued)

**Table 3.** (continued)

Improvement Strategy	Cycle 1			Cycle 2		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
<i>Systems and routines for managing participants' academic information</i>						
<i>Usual approach:</i> Information shared between instructors, YU staff, and coaches (those coaching students in Usual Strategies Group) includes data on class attendance, professional dress, and typical behaviors in YU contract; no intentional discussion of academic issues in Learning Community (LC) meetings or enactment of academic support plans						
<i>Improvement strategies</i>						
<i>Initial plan:</i> Review academic history info from admissions process; YU program lead or academic coordinator shares academic info collected through prescriptive monitoring protocol with coaches of participants in Improvement Strategies Group and follows up with 1:1 session with coaches as needed	✓	✓	✓			
<i>Implementation</i>						
Site managers routinely reached out to instructors to collect academic info and shared info with coaches of participants in Improvement Strategies Group		✓			✓	
Academic coordinator designated to reach out to instructors and others at the college, share info with coaches of participants in Improvement Strategies Group, and flagged "students of note" in a spreadsheet	✓			✓		
<i>Modifications:</i> Academic issues surfaced in coaching are discussed in LC meetings, participants of concern are flagged, and academic support plans put in place				✓	✓	✓

(continued)

**Table 3.** (continued)

Improvement Strategy	Cycle 1			Cycle 2		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
<i>Attention to academics during coaching</i>						
<i>Usual strategies:</i> All coaches provide coaching as usual with focus on YU topic of the week and professional skill development; participants receive introduction and orientation to learning and development portfolio project (no academic focus)						
<i>Improvement strategies</i>						
<i>Initial plan:</i> Increase focus on academics during coaching; participants introduced to expanded portfolio assignment that includes focus on academics (coaches to check in and advise)	✓	✓	✓	✓	✓	✓
<i>Implementation</i>						
Increased focus on academics during coaching, informal flagging of participants' academic issues	✓	✓	✓			
Used one-page guide to facilitate academically focused conversations in coaching		✓	✓	✓		
Had participants pull-up grades during coaching from college learning management system (LMS)		✓	✓	✓		
Expanded portfolio assignment to include focus on academics (coaches checked in and advised participants)			✓			
<i>Modifications:</i> Coach training on Weekly Academic Coaching Notes Sheet and the Binder, coaches routinely flag students of concern and have participants pull-up grades from college LMS, and reduced expectation of including academic component in portfolio assignment				✓	✓	✓

(continued)

**Table 3.** (continued)

Improvement Strategy	Cycle 1			Cycle 2		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
<i>Additional supports</i>						
<i>Usual strategies: Participants encouraged to seek out tutoring should they need it, reminded of online tutoring</i>						
<i>Improvement strategies</i>						
<i>Initial plan: Increase referrals to existing college tutoring and support services as well as online tutoring</i>	✓	✓	✓	✓	✓	✓
Create textbook library		✓			✓	
Hire academic coordinator to meet regularly with instructors and collect data on participants' academic performance and engagement, share info with coaches of participants in Improvement Strategies Group	✓			✓		
Order Wi-Fi hot spots for participants in Improvement Strategies Group to use			✓			✓
<i>Implementation: Increased referrals to existing college tutoring and support services as well as online tutoring</i>	✓	✓	✓	✓	✓	✓
Created textbook library		✓			✓	
Hired academic coordinator	✓			✓		
<i>Modifications: Provide additional coaching to encourage participants in Improvement Strategies Group to take advantage of available academic and nonacademic supports; YU program staff to develop a compendium of available supports</i>				✓	✓	✓

Staff in each site developed a prescriptive protocol for monitoring participant attendance at all L&D-related activities (including college classes), completion of course assignments (e.g., homework), and interim performance indicators (e.g., quiz and test grades, feedback from instructors). The guiding principle behind this improvement strategy was to develop centralized processes for routinely collecting information from instructors on participants' academic engagement and performance using procedures tailored to the local environment. To aid in this monitoring, coaches (and in the case of one site, a designated staff member) were charged with establishing mechanisms they judged promising for routinely accessing participants' grades.

Second, site staff were committed to collecting information on students' academic histories before they entered the program and sharing this information with the relevant parties (e.g., participants' coaches). In addition, the Year Up Program lead or manager was to share information collected through the prescriptive monitoring protocol with coaches and other staff working with participants receiving the improvement strategies. Information was to be shared through a combination of verbal exchanges and shared worksheets. One-one-one follow-up sessions between the program lead or manager and coaches were to be carried out as needed.

Third, all sites were to encourage coaches working with participants receiving the improvement strategies to step up their focus on academics during their coaching sessions. During the first testing cycle, coaches were advised to begin by creating their own strategies for increasing emphasis on academics and, over time, to share the experiences (successful and not) with other coaches working with participants receiving the improvement strategies. As usual, the coaches working with participants assigned to receive the improvement strategies were told to check in and advise their coachees on their portfolio project—a capstone project of the PTC—and encourage their coachees to incorporate an academic focus in their portfolios.

Fourth, in all three sites, coaches working with participants receiving the improvement strategies were to routinely encourage the participants to take advantage of academic resources and supports available through their college partner (e.g., tutoring). To enhance their academic monitoring and support efforts, each program was given modest discretionary funds. In Cycle 1, one program decided to leverage these funds to create a textbook library for participants receiving the improvement strategies. Another planned to tap the discretionary funds to purchase mobile Wi-Fi hot spots as a means of improving participants' home internet access. The third site used its discretionary funds to hire an academic coordinator to aggressively monitor academic performance of participants in their college courses.

Most notably, this half-time staff member was expected to collect information on all participants, but to share with staff only information for those participants randomly assigned to the improvement strategies group.

### **Study Design: Implementing and Testing Strategies (Stages 2 and 3)**

The embedded experiment was designed to address two primary research questions: (1) What is the impact of the improvement strategies on the likelihood that program participants will successfully complete the program's 6-month L&D phase? And (2) what is the impact of the improvement strategies on the likelihood participants will be enrolled in college the month following completion of L&D? To answer these questions, youth who enrolled in the program during the two cycles of testing were randomly assigned—using individual randomization within site and career training track—to coaching groups that continued with the usual strategies for academic monitoring and supports or to coaching groups that adopted the improvement strategies. In most instances, coaches were paired and then randomized to work with only participants using either the usual or the improvement strategies for academic monitoring and support. However, in one site during Cycle 1, all coaches served a mix of students in the usual strategies and the improvement Strategies Groups, but with varying levels of information being passed on to them about the academic performance and challenges students were facing to inform their coaching. For Cycle 1, the coaches in this site also were trained to follow the relevant strategies for responding to identified challenges (i.e., the usual or improvement strategies).

### ***Sample Size and Allocation***

We aimed to enroll a sample of 300 youth, half of whom would be randomly assigned to the group that would receive the usual program strategies for academic monitoring and supports and the other half of whom would be assigned to the group that would receive the improvement strategies designed to boost L&D completion and continued enrollment in college. Randomization to the Usual or Improvement Strategies Groups would ensure comparable groups of youth in each treatment condition. By enrolling 300 participants in the study sample and assigning equal numbers to each strategy group, we estimated that we would have 80% power to detect meaningful size impacts (i.e., 12.6 and 7.2 percentage points for binary outcomes with control group [i.e., Usual Strategy Group] means of .7 and

.9, respectively). If the sample was evenly distributed across sites (i.e., 100 participants in each site), we estimated that we would have 80% power to detect impacts of 19.2 and 11.0 percentage points for binary outcomes with control group means of .7 and .9, respectively.

### *Data and Measures*

Data for the study came from multiple sources. Information on the backgrounds of participants in the study sample and their program performance were obtained from Year Up's administrative data system, and data on college enrollment came from the National Student Clearinghouse.<sup>2</sup> These data sources were complemented by participant surveys, coach surveys (Cycle 2 only), individual interviews with site staff, group interviews with participants, observations of coaching sessions and LC meetings, and biweekly monitoring calls with site staff.

The primary outcome measures, L&D completion and college enrollment in the month following L&D, are from Year Up's administrative data and data from the National Student Clearinghouse, respectively. The outcome data from both sources are available for the full study sample and are considered quite reliable.

We used data from a modified version of the Pre-Internship Survey Year Up routinely administers to participants as they complete L&D and prepare for internships to explore secondary impact questions and mechanisms of change. The modified Pre-Internship Survey included questions about academic preparedness, academic challenges during L&D, and academic supports received during L&D. Seventy-eight (78) percent of all sample members completed the survey. However, the response rate was higher for the Improvement Strategies Group than for the Usual Strategies Group (82% and 74%, respectively).

The evaluation team conducted bi-weekly monitoring calls with Program Managers and Site Directors at each program to learn about the experiences of participants and coaches in the two groups. These calls also allowed for maintaining open communication between local site staff and the evaluation team and for documenting notable events that could affect program experiences and/or outcomes for participants.

Near the end of the study, coaches in Cycle 2 were asked to complete a survey about their approaches to and experiences coaching program participants. This survey was completed by at least one coach from 27 of 32 coaching groups. The study team also periodically visited sites to observe coaching sessions and LC meetings, as well as to conduct group interviews with participants and individual interviews with staff.

## *Characteristics of the Study Sample*

A total of 317 participants were enrolled in the study sample over two enrollment cycles. Thirty-five (35) percent enrolled in Site 1, 25% enrolled in Site 2, and 40% enrolled in Site 3 (Table 4). Participants were recruited following Year Up’s usual procedures, with the exception that all applicants were informed about the study and were required to agree to participate should they enroll in Year Up. Within each site, those who consented to participate in the study were randomized (using a blocked individual randomization design) to receive either the usual or improvement strategies for academic monitoring and supports. Blocking was by training track. In the few cases where there were siblings, they were randomized to the same condition or to different conditions at the request of the site director.

In all three sites, participants in the study sample broadly mirrored the characteristics of youth targeted by Year Up. Roughly half of the participants in the sample were female, about 70% self-identified as Black or African American, about 15% self-identified as Hispanic or Latino, and about 13% self-identified as White or another race/ethnic group (Table 5). At enrollment, about 45% of participants in the sample were under age 20, 40% were aged 20–22, and about 15% were aged 23 or older.

Participants in the sample exhibited a range of readiness for the program, as measured by number of “risk factors,” such as unstable housing, primary responsibility for a child, and exposure to violence and/or trauma. About 25% exhibited no risk/readiness factors, whereas about 20% had six or more. About 40% of the participants in the sample had no prior college experience, while more than 30% had a year or more of college.

Because of the random assignment of participants to the Improvement and Usual Strategies Groups, there were only small-to-modest differences in participant characteristics between groups and none of the differences was sufficiently large to be statistically significant at the .10 level on a two-tailed test. In contrast, there were statistically significant differences in the characteristics of participants who enrolled in the first and second intake cycles of the study (Maynard et al., 2018). A higher proportion of the participants enrolled in the second cycle were under age 20 (57% vs. 31%) and significantly fewer had prior college experience (34% vs. 41%). These differences are likely due to the timing of recruitment for the two groups. The former group was enrolled at the end of the calendar year, whereas the latter group was recruited over the summer, close to the time when many eligible individuals had just completed high school.

**Table 4.** Sample Size by Treatment Condition, Program, and Enrollment Cycle.

Program	Cycle 1			Cycle 2			Total Enrolled		
	Usual Strategies Group	Improvement Strategies Group	Total	Usual Strategies Group	Improvement Strategies Group	Total	Usual Strategies Group	Improvement Strategies Group	Total
Site 1	27	27	54	30	26	56	57	53	110
Site 2	20	19	39	22	18	40	42	37	79
Site 3	28	30	58	36	34	70	64	64	128
Total	75	76	151	88	78	166	163	154	317

Source: Year Up Salesforce data.

Note. Accepted applicants who attended orientation were randomized to the Improvement Strategies Group or to the Usual Strategies Group. No participants switched condition during the study.

**Table 5.** Characteristics of the Study Sample at Enrollment, Total and by Program.

Participant Characteristic	Total	Program			p Value
		Site 1	Site 2	Site 3	
Gender = Female (%)	49.5	41.8	55.7	52.3	.1210
Race-ethnicity (%)					<.0001***
Black or African American	71.6	85.5	48.1	74.2	
Hispanic or Latino	14.8	10.0	36.7	5.5	
White or another race	13.5	4.5	15.2	20.3	
Age (%)					.0429**
Under 20	44.1	34.5	46.8	50.8	
20–22	40.4	45.5	44.3	33.6	
23 or older	15.5	20.0	8.9	15.6	
Number of risk/readiness concerns (%)					<.0001***
0–1	26.5	28.0	13.9	33.1	
2–3	35.7	32.9	50.6	28.8	
4–5	16.6	15.9	26.6	11.0	
6+	21.2	23.2	8.9	27.1	
Prior college (any; %)					.0013***
0 Years	41.4	55.5	26.6	38.3	
<1 Year	26.8	17.3	36.7	28.9	
1+ Years	31.9	27.3	36.7	32.8	
Prior college (Full-time; %)					.0018***
0 Years	41.4	55.5	26.6	38.3	
<1 Year	37.2	28.2	44.3	40.6	
1+ Years	21.5	16.4	29.1	21.1	
Sample size	317	110	79	128	

Source: Year Up program data.

Note. The study sample consists of youth who were first enrolled in Year Up with the January 2018 or July 2018 cohort at the three programs participating in the study. Data on the number of risk factors is missing for 12% of the study sample.

\*Statistically significant at the .10 level. \*\*Statistically significant at the .05 level. \*\*\*Statistically significant at the .001 level—on two-tailed tests.

### Methods of Estimation

All estimates were generated using STATA Version 15.0. Because individuals in the study sample were randomly assigned to the Improvement or Usual Strategies Groups, simple difference in means tests produce unbiased estimates of the impacts of the improvement strategies for academic monitoring and supports. However, for most analyses, we used multivariate regression models that controlled for demographic and background characteristics,

site indicators, and controls for blocking factors used in the randomization (e.g., site, training track, and sibling status) to improve the precision of estimates and to control for random differences between the two treatment groups (Bloom & Sommo, 2005; Orr, 1999; Raudenbush & Bryk, 2002).

There was no missing outcome data for the primary research questions. However, there was a 21% nonresponse on the supplemental survey (Modified Pre-Internship Survey) and, more concerning, a sizable difference in the response rate for participants in the Usual and Improvement Strategies Groups (see Maynard et al., 2018). To mitigate response bias in the descriptive analysis of those data, nonresponse weights were created and used in the analysis (see Maynard et al., 2018, for a description of the process for generating the weights).

## **Implementation of Tested Strategies**

Throughout the study period, the evaluation team carried out low-stakes, low-burden routine monitoring of the usual and improvement strategies implemented. This allowed for documentation of the treatment contrast, ensured the integrity of the evaluation, and supported program decisions on midcourse corrections (between the first and second testing cycles). This information, along with feedback from a focus group conducted with coaches who worked with participants in the Improvement Strategies Group and structured conversations with site leadership, led to decisions about modifications in the improvement strategies to be implemented for Cycle 2 (see Table 3). It also seeded an effort by the Year Up National team to work with local program staff to pull together and organize an “Academic Coaching Binder” (henceforth referred to as “the Binder”). The Binder included a variety of materials, some newly developed by coaches using the improvement strategies, while others were preexisting but repurposed to focus on academics.

### ***Enhancements or Changes to Tested Strategies During Cycle 1***

Some of the tested strategies were altered or enhanced during implementation in Cycle 1. For example, the original plan for the improvement strategies involved implementing centralized processes to routinely collect information from instructors using procedures tailored to the local environment. Strategies included gaining direct access to the partner colleges’ learning management system (LMS) and/or having a staff person

responsible for liaising with college faculty and the registrar to capture information on a regular basis (e.g., the academic coordinator in one site).

Midway through Cycle 1, the three programs converged on a common strategy that involved participants in the Improvement Strategies Group routinely pulling up their course information (e.g., upcoming assignments, syllabi, and grades) on the college LMS during weekly individual coaching sessions. In addition, biweekly monitoring calls with program staff revealed that all sites desired a tool for coaches working with participants in the Improvement Strategies Group to be able to track academic issues that arose during their coaching sessions. As a result, midway through Cycle 1, Year Up National staff and the evaluation team collaborated to create a one-page academic coaching guide for coaches working with youth in the Improvement Strategy Group.

### *Adjustments to Strategies Tested in Cycle 2*

Although some enhancements and changes were made to the tested strategies during Cycle 1, others were made in response to what was learned through focus groups conducted at the end of Cycle 1 with program leads from all three sites and coaches working with participants in the Improvement Strategies Group at one site. The focus group conducted with coaches allowed for candid discussion about the relative merits of using the one-page academic coaching guide, as well as for talking through issues that arose while using the guide.

Coaches viewed the guide as a powerful tool for identifying participants who were struggling with their courses, helping them uncover root causes of the challenges, allowing them to flag student concerns for other staff members (especially support services staff), and providing a way to document planned and completed actions to address challenges. That said, some coaches reported that aspects of the guide felt formulaic and that it would benefit from more space for recording information and guidance on supporting students facing academic challenges (e.g., referrals and follow-up actions).

This focus group, along with informal feedback from coaches at the other two sites, sparked a collaboration among the evaluation team, Year Up National team, and local program staff to create the Binder. Examples of resources in the Binder are tips for engaging with instructors, tips for troubleshooting academic challenges, and academically focused warm-up questions for coaching sessions.<sup>3</sup> Prior to the start of Cycle 2, all coaches in

the Improvement Strategies Group attended a formal training on how they could use the Binder.

One tool in the Binder that was widely used in Cycle 2 was the Weekly Academic Coaching Notes Sheet (commonly referred to as the Notes Sheet). This is an updated version of the one-page coaching guide used in Cycle 1. The Notes Sheet includes guidance and action steps that coaches can recommend to participants when trying to address academic issues that surface. It provides expanded space for tracking academic information as well as a space to document students' academic goals for the week. The Notes Sheets serve as a running record of behaviors coaches track, challenges they identify, supports and guidance they offer, and outcomes they observe or seek to achieve.

The expectation that coaches will update their Notes Sheets regularly intentionally nudges them to ask coachees to pull up their grades and assignments in the college LMS each week and to engage them in conversations about their academic work and status. Based on feedback following Cycle 2, the structure of the Notes Sheets reportedly helped coaches monitor participants' performance and hold them accountable for following up on actions necessary for staying or getting back on track with their courses (e.g., by accessing tutoring, seeking help from the instructor, buying or borrowing the text book, completing delinquent assignments).

In Cycle 2, coaches working with participants in the Improvement Strategies Group began sharing their completed Notes Sheets with their program manager and peer coaches during weekly LC meetings. This information sharing prompted more frequent communication among staff about participants' academic performance and support needs as well as available resources that had been helpful to other students.

Coaches working with participants in the Improvement Strategies Group reported being more likely than their counterparts working with participants in the Usual Strategies Group to refer their coachees to academic supports at the local college partner. However, participant feedback on the value of the resources available through the partner colleges was mixed.

Several of the improvement strategies used in Cycle 1 were either abandoned or de-emphasized going into Cycle 2 due to implementation challenges or findings that the strategies were not especially useful. For instance, all three sites lowered their expectation that they could rely on communications with instructors to obtain consistent feedback and insight into participants' academic performance. Coaches found it easier and more productive to have participants' share access to their LMS site during

coaching sessions than trying to secure the information from college instructors through email correspondence or direct contact.

Only one site hosted an instructor orientation luncheon in Cycle 1 to establish rapport with instructors and open lines of communication. They also hosted a similar luncheon in Cycle 2. The other two sites were unable to make this event happen in Cycle 1. Lastly, there had been a desire to make the final portfolio assignment completed by all Year Up students more academically focused for participants in the Improvement Strategies Group—enforced by their coaches. However, staff at all three sites deemed this a low priority.

## **Estimated Impacts of the Improvement Strategies**

### *Primary Outcomes*

Over two cycles of testing, the improvement strategies for academic monitoring and supports resulted in substantially higher L&D completion rates and rates of continued college enrollment into the month after L&D. Participants in the Improvement Strategies Group were 9.6 percentage points more likely than their counterparts in the Usual Strategies Group to complete L&D (79% vs. 69%,  $p$  value = .05; Table 6, upper panel). Consistent with the changes in the improvement strategies for Cycle 2, the estimated impact for Cycle 2 participants is considerably larger than that for Cycle 1 (14.3 vs. 4.3 percentage points, respectively) and it is highly statistically significant for Cycle 2, but not Cycle 1 ( $p$  values = .036 and .540 for Cycles 2 and 1, respectively).

Notably, the estimated differences are positive, though not statistically significant, for each of the three sites. Estimates range from a 6.7 percentage point gain in the L&D completion rate for participants in the Improvement Strategies Group as compared with their counterparts in the Usual Strategies Group in Site 2 over the two cycles of testing. Participants in the Improvement Strategies Groups in Sites 1 and 2 were over 11 percentage points more likely than their counterparts in the Usual Strategies Group to complete L&D. Moreover, the estimated impacts of the improvement strategies for Cycle 2 were larger than those for Cycle 1 in each of the three sites (see Maynard et al., 2018). In fact, the site with the largest estimated impact of the improvement strategies in Cycle 2 showed no evidence of impact of the improvement strategies in Cycle 1.

Participants in the Improvement Strategies Group also were more likely than their Usual Strategies Group counterparts to continue with college

**Table 6.** Estimated Impacts of the Improvement Strategies for Academic Monitoring and Supports on L&D Completion Rates.

Sample	Sample Group Means (%)		Difference in Means		5% Confidence Interval (%)
	Improvement Strategies	Usual Strategies	% Point	p Value <sup>a</sup>	
<b>Completion of L&amp;D</b>					
<b>Total sample</b>	<b>78.9</b>	<b>69.3</b>	<b>9.6*</b>	<b>.052</b>	<b>[1.5, 17.6]</b>
Cohort/cycle					
January 1, 2018	77.6	73.3	4.3	.540	[-7.2, 15.7]
July 2, 2018	80.2	65.9	14.3**	.036	[3.2, 25.4]
Program					
Site 1	77.9	66.7	11.2	.182	[-2.6, 25.1]
Site 2	83.1	71.4	11.7	.204	[-3.4, 26.8]
Site 3	77.0	70.3	6.7	.397	[-6.3, 19.8]
<b>Enrollment in college in Month-7 following enrollment</b>					
<b>Total sample</b>	<b>67.24</b>	<b>53.99</b>	<b>13.25**</b>	<b>.003</b>	<b>[6.00, 20.50]</b>
Cohort/cycle					
Cycle 1					
January 1, 2018	47.20	41.33	5.87	.334	[-4.10, 15.84]
July 2, 2018	84.76	64.77	19.99**	.002	[9.51, 30.47]

(continued)

**Table 6.** (continued)

Sample	Sample Group Means (%)		Difference in Means		5% Confidence Interval (%) [Lower Bound, Upper Bound]
	Improvement Strategies	Usual Strategies	% Point	p Value <sup>a</sup>	
Enrollment in college in Month-7 following enrollment					
Program					
Site 1	42.02	26.32	15.70**	.031	[3.77, 27.63]
Site 2	99.61	92.86	6.75	.196	[-1.82, 15.32]
Site 3	68.40	53.13	15.27*	.076	[1.17, 29.37]
Sample size	154	163	317		

Source: Data are from the Year Up program data system.

Note. These data were weighted to account for the blocking of participants prior to randomization. The estimates of the mean difference between the treatment groups are based on regression models that included covariates for participant characteristics at the time of program enrollment. Means for the Usual Strategies Group are unadjusted means. Full regression estimates are available from the first author upon request. L&D = learning and development.

The bold-face type is used for findings related to the primary research questions.

<sup>a</sup>p Values for tests of differences in impacts by enrollment cohort/cycle and by office are listed above those for the individual impact estimates.

\*Statistically significant at the .10 level. \*\*Statistically significant at the .05 level. \*\*\*Statistically significant at the .001 level—on two-tailed tests.

coursework after completing L&D. Data from the National Student Clearinghouse indicate that, across the three sites, 67% of those in the Improvement Strategies Group compared with 54% of their counterparts in the Usual Strategies Group ( $p$  value = .003) were enrolled in college during the first month following the end of their L&D cycle (Table 6, lower panel). As with impacts on L&D completion rates, the estimated impacts on college enrollment were especially large (20 percentage points,  $p$  value = .002) for participants in the Improvement Strategies Group who enrolled in Cycle 2. The estimated impact for Cycle 1 is both small (5.9 percentage points) and not statistically significant ( $p$  value = .334).

### *Secondary Outcomes*

Consistent with higher rates of persistence through L&D, participants in the Improvement Strategies Group were enrolled in college an average of 0.5 months longer during the observation period than were their counterparts in the Usual Strategies Group (5.6 vs. 5.1 months;  $p$  value = .004; Table 7, top panel). The size of the estimated impacts on months of college enrollment is roughly similar across all three sites (ranging from .43 to .53 months), with statistically significant estimates for Sites 1 and 2 ( $p$  values = .072 and .028, respectively).

Year Up's participant contract, which sets expectations for behavior and specifies a code of conduct, is an important strategy used to reinforce Year Up's high expectations. It was not clear how the improvement strategies adopted for this study would affect the incidence of contract infractions—that is, behaviors in violation of the contract—which, in turn, affect participants' weekly stipend. One could hypothesize that improvement strategies, which promote more and closer oversight of participants' activities, would result in greater conformance with program expectations and, thus, lower recorded behavioral infractions. On the other hand, it is possible that higher levels of monitoring would increase opportunities for staff to observe code violations that, in turn, reduce weekly stipends and could lower program retention.

The improvement strategies tested did not significantly alter the number of recorded infractions for participants during L&D (Table 7, middle panel). Those in the Improvement Strategies Group had an average of 7.0 infractions, and those in the Usual Strategies Group had an average of 6.6 infractions. The only notable subgroup difference was for Site 3, where those in the Improvement Strategies Group had an average of 1.5 more recorded

**Table 7.** Estimated Impacts of the Improvement Strategies for Academic Monitoring and Supports on Behavioral Outcomes.

Sample	Group Means		Difference in Means		95% Confidence Interval
	Improvement Strategies	Usual Strategies	Difference	p Value <sup>a</sup>	
<b>Months of college enrollment during the 7-Months following Enrollment in the Professional Training Corps Program</b>					
<b>Overall</b>	<b>5.59</b>	<b>5.10</b>	<b>0.49***</b>	<b>.004</b>	<b>[0.21, 0.77]</b>
Cohort/cycle					
January 1, 2018	5.00	4.80	0.20	.439	[-0.21, 0.61]
July 2, 2018	6.12	5.35	0.77**	.001	[0.39, 1.15]
Program					
Site 1	4.74	4.24	0.50	.177	[-0.11, 1.11]
Site 2	6.95	6.52	0.43*	.072	[0.04, 0.82]
Site 3	5.46	4.93	0.53**	.028	[0.14, 0.92]
<b>Number of contract infractions during learning and development (L&amp;D)</b>					
<b>Overall</b>	<b>7.03</b>	<b>6.56</b>	<b>0.47</b>	<b>.414</b>	<b>[-0.47, 1.41]</b>
Cohort/cycle					
January 1, 2018	7.42	6.27	1.15	.204	[-0.35, 2.65]
July 2, 2018	6.64	6.81	-0.17	.816	[-1.37, 1.03]
Program					
Site 1	4.78	5.37	-0.59	.450	[-1.87, 0.69]
Site 2	8.52	8.26	0.26	.834	[-1.80, 2.32]
Site 3	8.05	6.50	1.55*	.087	[0.07, 3.03]
<b>Number of contract points remaining at the end of L&amp;D</b>					
<b>Overall</b>	<b>195.64</b>	<b>166.9</b>	<b>28.74*</b>	<b>.070</b>	<b>[2.80, 54.68]</b>
Cohort/cycle					
January 1, 2018	178.24	187.47	-9.23	.696	[-48.00, 29.54]

(continued)

**Table 7.** (continued)

Sample	Group Means		Difference in Means		95% Confidence Interval [Lower Bound, Upper Bound]
	Improvement Strategies	Usual Strategies	Difference	p Value <sup>a</sup>	
July 2, 2018 Program	212.76	149.38	63.38***	.003	[28.51, 98.25]
Site 1	179.55	137.37	42.18*	.085	[1.99, 82.37]
Site 2	198.39	143.45	54.94*	.051	[8.83, 101.05]
Site 3	209.51	208.59	0.92	.974	[-45.22, 47.06]
Sample size	154	163	317		

Source: Data are from the Year Up program data system.

The bold-face type is used for findings related to the primary research questions.

Note. These data were weighted to account for the blocking of participants prior to randomization. The estimates of the mean difference between the treatment groups are based on regression models that included covariates for participant characteristics at the time of program enrollment. Means for the Usual Strategies Group are unadjusted means. Full regression estimates are available from the first author upon request. The maximum contract points participants could have at the end of L&D is 360 (150 initial points, plus up to 10 points per week over 21 weeks of L&D with no infractions).

<sup>a</sup> p Values for tests of differences in impacts by enrollment cohort/cycle and by office are listed above those for the individual impact estimates.

\*Statistically significant at the .10 level. \*\*Statistically significant at the .05 level. \*\*\*Statistically significant at the =.001 level—on two-tailed tests.

contract infractions than did their counterparts in the Usual Strategies Group ( $p$  value = .09).

In contrast to findings for contract infractions, participants in the Improvement Strategies Group had an average of 29 more contract points (a positive outcome) at the end of L&D than did their counterparts in the Usual Strategies Group ( $p$  value = .07; Table 7, Panel 3). However, this difference is entirely concentrated among participants enrolled in Cycle 2. The Improvement Strategies Group in Cycle 2 had an average of 63 (30%) more points at the end of L&D than did their counterparts in the Usual Strategies group ( $p$  value = .003). Moreover, the impacts are generated from positive impacts on contract point balances for the Improvement Strategies Groups at Sites 1 and 2. Participants in both the Improvement and Usual Strategies Groups in Site 3 had similar, relatively high contract point balances at the end of L&D (about 209 of 360 possible points—150 initial points, plus up to 10 points per week over 21 weeks of L&D with no infractions).

### *Mechanisms of Change*

The study included two subanalyses designed to identify factors that potentially explain observed impacts. One is an examination of participants' reported program experiences, particularly as they relate to engagement in college courses and experiences with instructors, coaches, and peers. The other is an examination of the contrast in actions carried out by coaches during group and individual coaching with an eye toward identifying patterns associated with the implementation of the improvement strategies and how that may have contributed to the observed impacts on the primary outcomes of interest.

Participants in both groups reported generally similar program experiences. Based on data from the modified Pre-Internship Survey administered to participants by Year Up near the end of L&D (or, for participants who terminated early, administered within days of their departure), there were no notable differences in how participants in the Improvement and Usual Strategies Groups viewed their program experiences. Participants in both groups reported generally similar levels of school and work-related activities (see Maynard et al., 2018, for details). Participants in the two groups reported taking similar numbers of courses (an average of about 4.5 per semester), and experiencing similar levels of course difficulty (moderate, on average).

Those in the Improvement Strategies Group reported spending slightly more time on homework (an average of 11 vs. 9.7 hr a week) and completing an average of 0.3 more courses (4.2 vs. 3.9). Both groups reported

working about 12 hours a week, on average, and had similar views on the extent to which work adversely affected their academics (moderately, on average). However, none of these differences was statistically significant at conventional levels (Table 8).

When asked whether they had encountered a range of challenges in their most difficult college course (i.e., attending class regularly, keeping up with assignments, doing well on tests and assignments, understanding the course), the responses for the two groups were generally quite similar. Both groups reported having had moderate success getting timely feedback and having received moderate levels of support when they did encounter difficulties. This support most often came from PTC staff, other participants in the course, or college instructors. Neither group reported frequent use of tutors.

Despite quite distinct differences in the tactics used by coaches working with participants in the Improvement and Usual Strategies Groups (see further discussion below), participants in both groups reported having high levels of communication with their coaches (an average of 3.7–3.8 on a scale of 1–5, where 1 = *never* and 5 = *daily*). This suggests that a high level of communication with one's coach does not necessarily translate into meaningful coaching for academics.

Participants in the two groups generally reported that feedback and support on coursework—provided by college instructors, PTC program staff, and others in their courses—was generally good to excellent and timely. However, participants in the Improvement Strategies Group rated the quality of the support from PTC staff significantly higher than did those in the Usual Strategies Group ( $p$  value = .042). Participants in both groups reported relatively low frequencies of receiving feedback and support from tutors at the college or online (average score of 1.3–1.6 on a 3-point scale).

Overall, participants in both the Improvement and Usual Strategies Groups reported having had good experiences in the program and reported being very likely to recommend the program to others (average scores of 8.4 and 8.5 on a scale of 0–10 for the Improvement and Usual Strategies Groups, respectively). With a few exceptions, participants in both groups reported having received similar types, intensities, and qualities of supports to address challenges and similar overall levels of satisfaction with the program.

Coaches were envisioned as the primary agents for facilitating improvements in participant outcomes. Consistent with this expectation, their responses to the coach surveys indicate that coaches working with participants in the Improvement Strategy Groups spent their time differently during their coaching sessions than did their counterparts working with participants in the Usual Strategies group. Moreover, these differences in

**Table 8.** Participants' Reported Views of Their Learning Experiences During Learning and Development (L&D).

Measure	Group Means		Difference in Means		Confidence Interval [Lower Bound, Upper Bound]
	Improvement Strategies	Usual Strategies	Difference	p Value	
<b>Academics</b>					
Courses during L&D	4.61	4.51	0.10	.6790	[-0.37, 0.57]
Difficulty of courses (1 = very difficult, 3 = not difficult)	2.23	2.16	0.07	.3657	[-0.08, 0.22]
Hours per week on homework	11.05	9.71	1.34	.3033	[-1.22, 3.90]
Number of courses completed	4.24	3.90	0.34	.2587	[-0.26, 0.95]
<b>Employment</b>					
Hours per week worked for pay	12.10	12.86	-0.76	.7391	[-5.25, 3.73]
Degree work adversely affected academics (1 = not at all, 3 = a great deal)	2.20	2.16	0.04	.7523	[-0.21, 0.30]
Sample size	134	114	248		

Source: These data are from an online survey administered to participants. Participants who completed the L&D phase of the program received the survey very near the completion of L&D. Those who left the program before completing L&D were asked to complete the survey shortly after they left the program. It was an augmented version of the Pre-Internship Survey Year Up that participants complete prior to entering their internship positions.

Note. Overall, 78% of participant in the study sample completed the survey—84% of those in the Improvement Strategies Group and 72% of those in the Usual Strategies Group.

\*Statistically significant at the .10 level. \*\*Statistically significant at the = .05 level. \*\*\*Statistically significant at the .001 level on two-tailed tests.

time use were consistent with the improvement strategies being tested. For example, higher percentages of those working with participants in the Improvement Strategies Group, as compared with their counterparts working with participants in the Usual Strategies Group, reported using coaching time to address academic, social, personal, and work-related issues. Only slightly lower percentages reported using coaching time to address Year Up-related issues (see Maynard et al., 2018, for details).

Coaches working with the Improvement Strategies Group reported having substantially greater awareness of the academic challenges their coachees were facing during L&D compared to coaches working with participants in the Usual Strategies Group. For example, even though participants were randomly assigned to the Improvement or Usual Strategies Groups and, thus, statistically similar in their readiness for the program, coaches using the improvement strategies were more likely than their counterparts working with usual strategies to report that academics were a major challenge for their coachees (15% vs. 3%). Conversely, coaches using usual strategies were more likely to report that academics were “not at all” an issue for their coachees (37% vs. 23%).

The most notable difference in reported behavior by coaches working with the two groups of participants is the amount of time coaches reported spending discussing academic activities, performance, and support needs of participants. When asked about their most recent coaching session, coaches working with the Improvement Strategies Group reported having discussed academics with participants in approximately 43% of coaching sessions. In contrast, coaches working with the Usual Strategies group reported discussing academics in only 11% of their most recent coaching sessions.

As mentioned above, this greater focus on academic issues among coaches working with the Improvement Strategies Group did not come at the expense of discussing other issues (e.g., personal, social, work). Specifically, coaches in the Improvement Strategies Group reported greater frequency of discussing personal issues than did their counterparts who were working with the Usual Strategies Group (41% vs. 22% of coaching sessions). The same was true for the frequency of discussing social issues (also 41% vs. 22% of coaching sessions) and work-related issues (35% vs. 25% of coaching sessions).

Notably, coaches working with the Improvement Strategies Group were 3 times more likely than their counterparts working with the Usual Strategies Group to report having referred participants to tutoring (43% vs. 14%). However, both coaches and participants reported that the available tutoring tended to be poorly aligned with participants’ needs and not always easily accessible.

## Discussion

This case example illustrates the power of blending tools of DBIR, IS, RCE and traditional RCT evaluation methods strategically to quickly design, implement, refine, and rigorously test the effectiveness of new and different program strategies. The bundle of strategies that programs designed and initially tested in Cycle 1 showed promise for improving program retention through and college persistence beyond L&D. The gains in Cycle 1 were insufficient to achieve program sustainability. However, the close observation and evaluation of those experiences informed refinements in the strategies that appear to have strengthened their effectiveness in Cycle 2.

The rapid-cycle experimental impact findings provide highly credible estimates of those effects across site and over time, while the companion qualitative research that was designed around DBIR and IS principles was powerful for guiding the design of and refinements of the improvement strategies. The one caveat is that we are not able to fully control for the possibility that some of the difference in the magnitude of impacts between cycles 1 and 2 could be attributable to the differences noted above in the characteristics of the youth enrolled in the fall and spring enrollment cycles. However, it seems unlikely that those differences would dominate the contribution of the refinements to the improvement strategies.

In the end, the improvement strategies tested over two enrollment cycles resulted in sizable improvements in the academic persistence of participants. By the end of only two cycles of testing, all three programs had improved their retention rates substantially using strategies that were relatively easy to implement and adopt at low-to-no cost. All three programs identified and implemented strategies that allowed them to routinize access to information about participant engagement and performance in their college courses. They could use that information to engage with participants experiencing difficulties to identify contributing factors and come up with feasible approaches to address them. Notably, however, the strategy changes that led to the observed improvements varied across sites in large part due to differences in local contexts and associated opportunities and challenges.

Following the completion of the study, coaches in the study sites began having all participants routinely pull-up their grades in the college LMS during coaching sessions. Coaches also began routinizing processes to encourage and facilitate conversations among program staff about participants' academic progress. Routinely accessing the college LMS during coaching sessions changed the coaching conversations by bringing academics to the foreground. As a result, there was greater accountability and

follow-through by both participants and coaches for how coaching time was used and what coaches and coachees were expected to do between sessions.

In Cycle 2 of the study, coaches in the Improvement Strategies Group were trained on the use of the Binder, which, along with the Notes Sheet, helped coaches working with participants in the Improvement Strategies Group establish routines for coaching around academics and equipped them with tools for both surfacing issues and working with participants to address them. In the words of one coach: “The coaching binder [ . . . ] was so awesome . . . because [instead of asking] ‘Okay so how was your week-end’ . . . [we focus on] how can we [help you] academically?”

In follow-up interviews with program staff at Site 1, we learned that staff have continued working to improve tracking student engagement and performance in their college courses and to provide quicker and better responses to help those showing signs of difficulty. They also are continuing their efforts to improve information sharing among staff within the PTC and between PTC and college staff. Reportedly, it is becoming common for issues that arise during coaching sessions to be shared in some form during LC meetings and with other program staff, which fosters shared efforts to monitor and support academic needs of all participants.

Reportedly, troubleshooting academic issues students encounter now is viewed as more of a shared responsibility of the PTC programs and their college partners than was the case prior to this study. For example, in one site, there now are multiple avenues through which information about academic challenges is shared. There are biweekly meetings with the college academic advisor, the program’s associate director of programs and the program managers to discuss participants’ academic challenges and offer suggested sources of remedy where warranted. Further, academic advisors now contact program managers (generally by email), as needed, to share information about students in need of support based on information obtained through college instructors. This sometimes leads to follow-up conversations and/or meetings with individual participants.

Although more systems for communication and collaboration exist than prior to the study, program staff report that there is still room for improvement. Too often, academic issues are not communicated to PTC staff until it is too late for intervention. In addition, the academic resources available at the local colleges are limited, and participants frequently report having difficulty accessing them or that they are unhelpful. However, it is not always clear whether the available services are not useful or if students are not adept at taking full advantage of them. For example, in the case of tutoring services, it seems likely that some of the services may be helpful

but difficult to access and others may be poorly aligned with participants' needs. But, it also is likely that some participants lack the initiative and/or requisite skills to engage with tutors productively. PTC staff are continuously striving to create convenient spaces that encourage participants to spend more time studying at the college, as this allows greater access to available academic support services and provides more opportunities for contact with and feedback from the program.

### *Lessons for Future Improvement Studies*

In this case example, we were able to facilitate the efforts of programs to prioritize and embark on serious, evidence supported improvement efforts. In doing so, we reached beyond traditional experimental program evaluation methods to capitalize on the strengths of DBIR and IS methods for improving program design and implementation. In this, we acted as facilitators to accelerate efforts of program staff to identify priority strategies for testing and to plan for effective implementation of those strategies in a testable manner. We used the tactics of RCE methods for maintaining a focus on priority, proximal outcomes, and reliable and accessible data sources. The following are key lessons from our experience:

1. It can be helpful for evaluators to facilitate brainstorming by program staff working to identify the priority targets for improvement. Such brainstorming should be wide-ranging and include issues of feasibility, promise of improving outcomes, and contingencies.
2. It is helpful to standardize improvement goals across site, but to expect variability in what program staff deem most promising for achieving those goals. In our case, the sites agreed that their priority operational goals were to improve timely access to information on participants' academic performance and to have coaches and staff prepared to help participants who were falling short of expectations in their college courses. However, sites differed with respect to how they could most effectively achieve these operational goals. Allowing program staff to tailor the specifics of their improvement strategies was important in securing staff buy-in. It also likely led to the strategies tested being more effective than had we adopted a uniform protocol for all sites.
3. It was critical to have a nonburdensome strategy for enrolling participants in the study sample and randomly assigning them to treatment condition. We inserted the informed consent process into the

usual program enrollment process. Then, the evaluation team worked with the National Year Up evaluation staff to conduct the randomization of participants immediately preceding program orientation.

4. Other aspects of the evaluation also need to be nonburdensome for staff. We were careful to impose minimal burden on the program staff throughout the evaluation. Furthermore, the evaluation was nonjudgmental; we were collectively exploring the benefits of altering the strategies sites used for monitoring and supporting participants in their college courses.
5. Findings from the study were shared with Year Up National and site staff in a timely, easy-access, and nonjudgmental manner. After Cycle 1, the evaluation team packaged qualitative feedback from the sites in easily digestible formats (i.e., power points, one-page briefs, and simplified meeting notes) to allow them to make quick, informed decisions about modifications for Cycle 2. Results of the study were presented to staff within a few months of end of Cycle 2 of testing, again using easy-access formats. We prepared a brief pre-read that was emailed to all stakeholders in advance of webinars with each site and with the national office. This was followed up by a more detailed, but still easy-access post-read designed for online reading.

### *Study Limitations*

This was a relatively short-term, low-budget evaluation with a narrow focus. It focused exclusively on the impacts of the tested improvement strategies, not on overall impacts of the program. The study sample was relatively small (317) and spread across two cycles of testing improvement strategies that were intentionally modified between cycles.

The study was designed to test whole bundles of strategies, not the individual components. For example, it was not designed to estimate the impact of guaranteed access to textbooks, as distinct from the impact of other improvement strategies included in the bundle tested. Relatedly, we cannot be certain how much of the stronger impacts in for Cycle 2 is attributable to differences in the composition of the participating youth in the study sample or to the refinements in the improvement strategies tested. Finally, the limited nature of the implementation monitoring means that we do not have especially rich detail on precisely where, when, how, and how well various aspects of the program were implemented and why.

All of these limitations arise from deliberate choices made by the evaluation team to prioritize a quick, low-burden, credible test of promising improvement strategies designed by program staff over a more traditional summative impact evaluation—one that would generate credible evidence on the overall effectiveness of a program (e.g., on ultimate employment and earnings goals) that was already known to need major improvements to meet benchmarks for success. Such a study would answer different questions and take longer and require larger samples and more resources. Both types of studies are valuable; however, at the time we settled on the study design, all parties to the decision agreed that it would be prudent to prioritize the improvement study over the summative evaluation. Notably, we are in the process of completing a summative evaluation of the PTC, the results of which will reflect some of the findings of this improvement study.

## Appendix

**Table A1.** Taxonomy of Approaches to Program Evaluation.

### Traditional Program and Policy Evaluation

Commonly, traditional evaluations are described as using “social research methods to systematically investigate the effectiveness of social intervention programs.

They draw on the techniques and concepts of social science disciplines and is intended to be useful for improving programs and informing social action aimed at ameliorating social problems” (Rossi et al., 2018). Since the turn of the century, there has been an increasing evidence on the use of experimental evaluations to generate unbiased estimates of the expected impacts of practices, policies, or programs; estimates of the degree of certainty in the impact estimates (with randomized controlled trials having highest standing); and evidence of the relevance to the estimates for particular population groups and settings (Cronbach et al., 1980; Mertens & Wilson, 2018; Orr, 1999; Peck, 2017). There also has been an increased attention to synthesizing such evidence across studies using systematic review processes to arrive at “conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, policy or practice” (Cooper et al., 2019).

*Types of Questions Addressed:* “What happens as a result of subjecting individuals to a practice, program, or policy that is different from usual?” What changes in outcomes occur as a result of individuals being subjected to the practice, program, or policy rather than the usual conditions? How different are the impacts of the practice, policy of program across settings, participant groups, or time? In what ways does the data support or refute the operative theory of change associated with the practice, policy, or program?

(continued)

**Table A1.** (continued)

*Illustrative Applications:* Negative Income Tax Experiments (Burtless & Hausman, 1978), Job Corps Evaluation (Mallar, 1982), Teacher Advancement Program (Glazerman & Seifullah, 2010), Abstinence Only Education Programs (Trenholm et al., 2007), and Year Up Program Evaluation (Fein & Hamadyk, 2018).

**Design-based Implementation Research (DBIR)**

DBIR seeks to simultaneously advance basic knowledge about the relative effectiveness of practices, policies, or programs work through iterative applications of variations in them and careful observation and assessment of the results. It draws on principles of implementation science and improvement science (see below). A defining feature of DBIR is reliance on researcher–practice partnerships dedicated to simultaneously using and improving theories through iterative cycles of design, implementation, and evaluation of modifications to practice (Fishman & Penuel, 2018; Fishman et al., 2013; Penuel & Fishman, 2012; Penuel et al., 2011).

*Types of Questions Addressed:* What is the current state of knowledge regarding how a particular outcome is achieved? What are possible strategies for strengthening outcomes and/or removing barriers to achieving intended outcomes? How does practical experience inform theories of action? How generalizable is the experience in one setting or for one population group to another?

*Illustrative Applications:* Promoting teacher involvement in curriculum development (Huizinga et al., 2014), improving academic language development of students (Snow et al., 2009), and connecting out of school youth to learning opportunities (Barron et al., 2014).

**Implementation Science**

Implementation science entails disciplined examination of programs, policies, and practices to advance knowledge about the ways they influence outcomes of interest, with the goal of creating generalizable knowledge that can inform policy and practice more generally (Eccles & Mittman, 2006). It differs from DBIR and Improvement Science in that it does not explicitly involve the implementers of the program, policy, or practice in the research and the intended beneficiary of the evaluation is the implementation setting. Implementation science is described in the literature the processes and methods involved in the systematic transfer and uptake of evidence-based practices into routine, everyday practice. Its central aims are to support the understanding of relevant, contextual processes and improve the quality and effectiveness of programs, policies or practices (Bauer et al., 2015; Fisher et al., 2016; Kelly & Perkins, 2012; Odom, 2009).

*Types of Questions Addressed:* What are the key design features of a program, policy, or practice? How are those features implemented in practice? What are key performance indicators or outcomes at key points in the implementation process? What factors appear to facilitate or impede achievement of intended outcomes (Kelly & Perkins, 2012).

(continued)

**Table A1.** (continued)

*Illustrative Applications:* Approaches to achieving positive outcomes for children and youth with disabilities (Odom, 2009), effective strategies for teaching nurses quality and safety practices (Dolansky et al., 2017), Efficient Operation of the Food Stamp Employment Training Programs (Puma et al., 1988), and barriers and facilitators of an evidence-based employment and training program (Noel et al., 2017).

---

#### Improvement Science (IS)

---

Improvement science is a disciplined approach through which evaluators and practitioners collaborate to design, implement, and evaluate strategies intended to improve programs, policies, or practices. It uses and informs theory and promotes continuous program improvement through cycles of refinement and testing of the theories of change (Bryk et al., 2011; Bryk et al., 2015; Russell et al., 2017).

*Types of Questions Addressed:* What are perceived impediments to successful outcomes? What are the most promising strategies for improving outcome? What is entailed in implementing those strategies? How well were the focal improvement strategies implemented? How well did they work? Should they be adopted, tweaked, or abandoned; why and how?

*Illustrative Applications:* The Community College Pathways Networked Improvement Community (Bryk et al., 2011), networked improvement science effort to support instructional change (Hill, 2019), and development and testing of a new developmental mathematics curriculum (Norman et al., 2018).

---

#### Rapid-Cycle Evaluation (Consolidated Framework for Implementation Research)

---

Rapid-cycle evaluation draws on traditional program evaluation methods for generating highly credible estimates of the effectiveness of strategies for improving outcomes of programs, policies, or practices, as well as implementation science, DBIR, and IS. By design, it is quick turnaround and addresses very specific features of design and implementation of programs or policies. They consider a very limited set of outcome measures, a well-defined test population, and preference reliance on a randomly assigned comparison group that is not exposed to the design or implementation feature in question and their implementation (Cody & Asher, 2014; Keith et al., 2017; Shrank, 2013).

*Types of Questions Addressed:* What is the focal challenge to optimal outcomes associated with the program, policy, or practice in question? What is change in strategy that is being tested? What is the logic as to why it is expected to improve outcomes? How do outcomes change as a result of the strategy change? Do the impacts vary across setting or population group?

*Illustrative Applications:* Comprehensive Primary Care Initiative (Shrank, 2013), building stronger relationships with low-income fathers through mobile apps (Balu et al., 2018), improving human services programs (Derr et al., 2019), and improving early care and education (Paulsell et al., 2018).

---

## Authors' Note

The case example described benefited from the wisdom and support of many staff at Year Up, including Garrett Warfield and Jessica Britt, staff at the Year Up programs and leaders at the programs' college partners.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: Support for this article and the underlying research was provided by the Institute of Education Sciences (Grant # R305A150214), the Corporation for National and Community Service's Social Innovation Fund to the GreenLight Fund, and the W. T. Grant Foundation.

## ORCID iD

Rebecca A. Maynard  <https://orcid.org/0000-0003-0014-3477>

## Notes

1. Adding the additional sites to the evaluation was possible as the result of blending work on this project with ongoing, related work on an Institute of Education Sciences Development and Innovation (grant R305A150214).
2. National Student Clearinghouse data are the source of standard metrics used by colleges and universities to judge their performance and the performance of individual participants. They have been used in many past community college evaluations (see, e.g., publications at <http://www.mdrc.org/issue/higher-education>).
3. Year Up National reported that they are in the process of developing resources and protocols to support the rollout of key tools and practices demonstrated to be efficacious through this improvement study.

## References

- Bailey, T. R., Jaggars, S. S., & Jenkins, D. (2015). *Redesigning America's community colleges*. Harvard University Press.
- Balu, R., Lee, S., & Steimle, S. (2018). *Encouraging attendance and engagement in parenting programs: Developing a smartphone application with fathers, for fathers. The building bridges and bonds study* (OPRE Rep. No. 2018-68). Office of Planning, Research and Evaluation.

- Barron, B., Gomez, K., Pinkard, N., & Martin, C. K. (2014). *The digital youth network: Cultivating digital media citizenship in urban communities*. MIT Press.
- Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*, 3(1), 32.
- Belfield, C. R., & Bailey, T. (2011). The benefits of attending community college: A review of the evidence. *Community College Review*, 39(1), 46–68.
- Bloom, D., & Sommo, C. (2005). *Building learning communities early results from the opening doors demonstration at Kingsborough community college*. MDRC.
- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. Hallinan (Ed.) *Frontiers in sociology of education. Frontiers in sociology and social research* (pp. 127–162, Vol. 1). Springer.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.
- Burtless, G., & Hausman, J. A. (1978). The effect of taxation on labor supply: Evaluating the gary negative income tax experiment. *Journal of political Economy*, 86(6), 1103–1130.
- Casazza, M. E., & Silverman, S. L. (2013). Meaningful access and support: The path to college completion. Retrieved from Council of Learning Assistance and Developmental Education Associations website, [http://www.cladea.net/white\\_paper\\_meaningful\\_access.pdf](http://www.cladea.net/white_paper_meaningful_access.pdf)
- Christie, C. A., Inkelas, M., & Lemire, S. (Eds.). (2017). *Improvement science in evaluation: Methods and uses: New directions for evaluation, number 153*. John Wiley & Sons.
- Cody, S., & Asher, A. (2014). *Smarter, better, faster: The potential for predictive analytics and rapid-cycle evaluation to improve program development and outcomes*. Mathematica Policy Research.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation*. Jossey-Bass.
- Dadgar, M., & Trimble, M. J. (2015). Labor market returns to sub-baccalaureate credentials: How much does a community college degree or certificate pay? *Educational Evaluation and Policy Analysis*, 37(4), 399–418.
- Daily, S., Tout, K., Douglass, A., Miranda, B., Halle, T., Agosti, J., Partika, A., & Doyle, S. (2018). *Culture of continuous learning project: A Literature review of*

- the breakthrough series collaborative (BSC)* (OPRE Rep. No. 2018-28). Administration for Children & Families.
- Dechaysay, N., & Anzelone, C. (2016). *Cutting through complexity: Using behavioral science to improve Indiana's Child Care Subsidy Program* (OPRE Rep. No. 2016-03). Administration for Children & Families.
- Derr, M., McCay, J., & Person, A. (2019). *The innovate phase: Co-creating evidence-informed solutions to improve human services programs* (No. 1075712b5bd84e61b0eb1eebf2e7cd05). Mathematica Policy Research.
- Dolansky, M. A., Schexnayder, J., Patrician, P. A., & Sales, A. (2017). Implementation science: New approaches to integrating quality and safety education for nurses' competencies in nursing education. *Nurse Educator*, 42(5 S), S12–S17.
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to *implementation science*. *Implementation Science*, 1, 1.
- Fein, D., & Hamadyk, J. (2018). *Bridging the opportunity divide for low-income youth: Implementation and early impacts of the Year Up program* (OPRE Rep. No. 2018-65). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Fishman, B. J., & Penuel, W. R. (2018). Design-based implementation research. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 393–400). Routledge.
- Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National Society for the Study of Education*, 112(2), 136–156.
- Fisher, E. S., Shortell, S. M., & Savitz, L. A. (2016). Implementation science: A potential catalyst for delivery system reform. *JAMA*, 315(4), 339–340.
- Glazerman, S., & Saifullah, A. (2010). *An evaluation of the teacher advancement program (TAP) in Chicago: Year two impact report*. Mathematica Policy Research.
- Gueron, J., & Rolston, H. (2013). *Fighting for reliable evidence*. Russell Sage.
- Hagedorn, L. S., & Kuznetsova, I. (2016). Developmental, remedial, and basic skills: Diverse programs and approaches at community colleges. *New Directions for Institutional Research*, 2015(168), 49–64.
- Haskins, R., & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy*. Brookings Institution Press.
- Hill, M. B. (2019). *Improving instruction: An examination of a network improvement science effort to support instructional change* [Doctoral dissertation, UCLA].
- Huizinga, T., Handelzalts, A., Nieveen, N., & Voogt, J. M. (2014). Teacher involvement in curriculum design: Need for support to enhance teachers' design expertise. *Journal of Curriculum Studies*, 46(1), 33–57.

- Jacobson, L., Lalonde, R. J., & Sullivan, D. G. (2005). The impact of community college retraining on older displaced workers: Should we teach old dogs new tricks? *Industrial & Labor Relations Review*, 58(3), 398–415.
- Jacobson, L., & Mokher, C. (2009). *Pathways to boosting the earnings of low-income students by increasing their educational attainment*. Hudson Institute Center for Employment.
- Jaggars, S. S. (2011). *Online learning: Does it help low-income and underprepared students?* (CCRC Working Paper No. 26). Columbia University, Teachers College, Community College Research Center.
- Jepsen, C., Troske, K., & Coomes, P. (2009). *The labor-market returns for community college degrees, diplomas, and certificates*. University of Kentucky and University of Louisville.
- Kane, T., & Rouse, C. (1995). Labor market returns to two and four-year college. *American Economic Review*, 85, 600–614.
- Keith, R. E., Crosson, J. C., O'Malley, A. S., Cromp, D., & Taylor, E. F. (2017). Using the consolidated framework for implementation research (CFIR) to produce actionable findings: A rapid-cycle evaluation approach to improving implementation. *Implementation Science*, 12(15), 1–12.
- Kelly, B., & Perkins, D. F. (Eds.). (2012). *Handbook of implementation science for psychology in education*. Cambridge University Press.
- Kennedy, S. D. (1980). *The final report of the housing allowance demand experiment*. Abt Associates.
- Mallar, C. (1982). *Evaluation of the economic impact of the job corps program. Third follow-up report*. Mathematica Policy Research, Inc.
- Manzi, J. (2012). *Uncontrolled: The surprising payoff of trial-and-error for business, politics, and society*. Basic Books.
- Marcotte, D. E. (2010). The earnings effect of education at community colleges. *Contemporary Economic Policy*, 28(1), 36–51.
- Maynard, R., Baelen, R. N., Fein, D. J., Shivji, A., & Souvanna, P. (2018). *Improving learning and development completion: Experimenting with alternative strategies for monitoring & supporting academics in the PTC*. Post Read for Year Up staff conversations. Abt Associates.
- Maynard, R., Goldstein, N., & Nightingale, D. S. (2016). Program and policy evaluations in practice: Highlights from the federal perspective. *New Directions for Evaluation*, 2016(152), 109–135.
- Means, B., & Penuel, W. R. (2005). Research to support scaling up technology-based educational innovations. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons learned from technology-based educational improvement* (pp. 176–197). Jossey-Bass.

- Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice*. Guilford Publications.
- Munnell, A. H. (1987). Lessons from the income maintenance experiments: An overview. *New England Economic Review*, 1–21.
- Newhouse, J. P. (1993). *Free for all? Lessons from the RAND health insurance experiment*. Harvard University Press.
- Noel, V. A., Bond, G. R., Drake, R. E., Becker, D. R., McHugo, G. J., Swanson, S. J., Luciano, A. E., & Greene, M. A. (2017). Barriers and facilitators to sustainment of an evidence-based supported employment program. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(3), 331–338.
- Norman, J. R., Yamada, H., & Huang, M. (2018). *Degree attainment and transfer among Statway© students: A propensity score matched analysis of outcomes*. Carnegie Foundation for the Advancement of Teaching.
- Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education*, 29(1), 53–61.
- Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Sage.
- Paulsell, D., Harding, J. F., Burwick, A., & English, B. (2018). *Developing an equity-focused policy research agenda for low-income families with young children* (No. ae26aba26fbb40859a8abb8b98d223dd). Mathematica Policy Research.
- Peck, L. R. (Ed.). (2017). *Social experiments in practice: The what, why, when, where, and how of experimental design and analysis: New Directions for Evaluation, Number 152*. John Wiley & Sons.
- Penuel, W. R., & Fishman, B. J. (2012). Large-scale intervention research we can use. *Journal of Research in Science Teaching*, 49(3), 281–304.
- Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40, 331–337.
- Perin, D. (2011). Facilitating student learning through contextualization: A review of evidence. *Community College Review*, 39(3), 268–295.
- Proger, A. R., Bhatt, M. P., Cirks, V., & Gurke, D. (2017). *Establishing and sustaining networked improvement communities: Lessons from Michigan and Minnesota* (REL 2017–264). U.S. Department of Education, Institute of Education Sciences.
- Puma, M., Werner, A., & Hojnacki, M. (1988). *Report to congress on program implementation: Evaluation of the food stamp employment and training program*.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage Publications.

- Roder, A., & Elliot, M. (2014). *Sustained gains: Year Up's continued impacts on young adults' earnings*. Economic Mobility Corporation.
- Ross, C., Sama-Miller, E., & Roberts, L. (2018). *Using research and evaluation to support programs that promote parents' economic security and children's well-being* (OPRE Rep. No. 2018-04). Administration for Children & Families.
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach*. Sage publications.
- Russell, J. L., Bryk, A. S., Dolle, J., Gomez, L. M., LeMahieu, P., & Grunow, A. (2017). A framework for the initiation of networked improvement communities. *Teachers College Record*, 119(7), 1–36.
- Shrank, W. (2013). The center for Medicare and Medicaid innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Affairs*, 32(4), 807–812.
- Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), 325–344.
- Sommo, C., Mayer, A. K., Rudd, T., & Cullinan, D. (2012). *Commencement day: Six-year effects of a freshman learning community program at Kingsborough Community College*. MDRC.
- Tichnor-Wagner, A., Wachen, J., Cannata, M., & Cohen-Vogel, L. (2017). Continuous improvement in the public-school context: Understanding how educators respond to plan–do–study–act cycles. *Journal of Educational Change*, 18(4), 465–494.
- Trenholm, C., Devaney, B., Fortson, K., Quay, K., Wheeler, J., & Clark, M. (2007). *Impacts of four Title V, Section 510 abstinence education programs* (Final Report). Mathematica Policy Research. MPR Ref. 8549-110.
- Visher, M. G., & Stern, D. (2015). *New pathways to careers and college: Examples, evidence, and prospects*. MDRC.
- Weiss, M. J., Visher, M. G., Weissman, E., & Wathington, H. (2015). The impact of learning communities for students in developmental education: A synthesis of findings from randomized trials at six community colleges. *Educational Evaluation and Policy Analysis*, 37(4), 520–541.
- Weissman, E., Cullinan, D., Cerna, O., Safran, S., & Richman, P. (2012). *Learning communities for participants in developmental English: Impact studies at Merced College and the Community College of Baltimore County*. Teachers College, National Center for Postsecondary Research.
- Zuchowski, I., Miles, D., Woods, D., & Tsey, K. (2019). Continuous quality improvement processes in child protection: a systematic literature review. *Research on Social Work Practice*, 29(4), 389–400.