September 2016

Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program

Jim Parsons, Chris Weiss, and Qing Wei

Vera
INSTITUTE OF JUSTICE

Executive Summary

In 2012, the City of New York launched the Adolescent Behavioral Learning Experience (ABLE) program, a large-scale initiative serving 16- to 18-year-old youth detained at the New York City Department of Correction's Rikers Island jail (Rikers). The ABLE program aimed to break the cycle of reincarceration for adolescents in the jail using Moral Reconation Therapy (MRT), an evidence-based intervention that focuses on improving decision making. The project employed an innovative social impact bond (SIB) funding mechanism to support the provision of services. SIBs invite private investors to fund a public service intervention through an intermediary organization—and the government repays the funder only if the program achieves certain goals, which are specified at the outset of the initiative and assessed by an independent evaluator.

Vera's Role

The Vera Institute of Justice served as the program's independent evaluator. Vera employed rigorous research methods to determine whether the program led to reductions in recidivism for youth passing through the jail. Specifically, the evaluation was designed to assess whether the program reduced recidivism by 10 percent or more, the pre-defined threshold of success at which the city would have to pay Goldman Sachs (the SIB's private investor) back for its investment in the program.

Research Design

To determine if the ABLE program met the threshold for success, Vera assessed "recidivism bed days" (RBDs), a measure that captures the number of days that members of the study cohort were held in the jail during the 12 months following their release from Rikers. Vera used a quasi-experimental design to do this evaluation.

The quasi-experimental approach compared RBDs for 16- to 18-year-olds who were eligible to participate in the ABLE program during 2013 with a matched historical cohort who passed through the jail before the program was established (from 2006 to 2010). To ensure an "apples to

apples" comparison, these cohorts were matched on a variety of factors including charge, criminal history, gender, and age using a statistical technique called propensity score matching. To control for the impact of factors that affect recidivism, but are unrelated to the program, such as citywide changes in crime rates or changes in policing practices, researchers also tracked RBDs for 19-year-olds over the same period and then adjusted the results of the analysis accordingly. Nineteen-year-olds were chosen for this purpose because they were not eligible to receive ABLE services and historical data demonstrate that trends in recidivism for 16- to 18-year-olds and 19-year-olds were closely aligned.

Findings

The ABLE program reached most of the 16- to 18-year-olds entering the Rikers Island jail during 2013. As Figure 3 (p. 14) shows, of the 1,691 adolescents who entered the jail during this time and were held for seven or more days, the overwhelming majority (87 percent) attended at least one ABLE session. (Jail intake and housing policies meant that adolescents who were housed in the jail for fewer than seven days rarely had the opportunity to receive ABLE services.) Of the adolescents who participated in the ABLE program during the study period, 44 percent reached one of the program's key milestones.

Vera determined that the program did not lead to reductions in recidivism for participants. When external factors were taken into account, the rate of recidivism for eligible 16- to 18-year-olds was statistically equivalent to the matched historical comparison cohort. The 19-year-olds and the study cohort (16- to 18-year-olds) displayed similar trends in rates of recidivism over time, indicating that any shifts were the result of factors other than the ABLE program. The program did not reduce recidivism and, therefore, did not meet the pre-defined threshold of success, a 10 percent reduction in RBDs for the study cohort.

Contents

Introduction

In early 2010, New York City announced the creation of the Young Men's Initiative, an effort to address the challenges facing young African American and Latino men in the city. Stemming from this initiative, the New York City Department of Correction (DOC) collaborated with the Mayor's Office and a number of nonprofit and private sector partners to design and implement a large-scale program—Adolescent Behavioral Learning Experience (ABLE)—for adolescents ages 16 to 18 detained at the city's Rikers Island jail (Rikers).

In 2010, before the program was established, more than 3,700 youth between the ages of 16 and 18 entered DOC custody. Forty-seven percent of these adolescents returned to Rikers within a year of discharge. Historical data suggests that the typical adolescent who passed through Rikers would spend more than 200 days back in jail over the following six years. [1]

The ABLE program aimed to break the cycle of re-incarceration for adolescents in the jail using Moral Reconation Therapy (MRT), an evidence-based intervention that focuses on improving decision-making. ABLE was designed to reduce recidivism with the aim of positively impacting the lives of the adolescents by breaking the cycle of re-incarceration and also saving the city money. ABLE was initially piloted with a subgroup of 16- to 18-year-olds entering Rikers during 2012. The program was then expanded at the beginning of 2013 to serve the entire adolescent population.

The ABLE program and moral reconation therapy

The ABLE program utilizes Moral Reconation Therapy (MRT), a form of cognitive behavioral therapy. Developed in 1985 by Correctional Counseling, Inc., MRT has been used in prisons, jails, drug courts, residential facilities, and schools, and has been shown to be effective in a variety of settings.[a] Participants progress through a series of steps with the help of a trained facilitator, graduating from one stage to the next by delivering testimonies or presentations about themselves. There are 12 steps in total, with "Honesty" as Step 1 and "Choosing moral goals" as Step 12.

[a] MRT had been shown to be effective in reducing recidivism among adult males in jail, (S. Krueger, "Five-year recidivism study of MRT-treated offenders in a county jail," *Cognitive Behavioral Treatment Review* 9, no. 2 (1997): 1-9; and a county detention center (G. Goodwin, S. Stone, and K. Hambrock, "Recidivism study: Lake County, Florida Detention Center," *Cognitive Behavioral Treatment Review* 4, no.1 (1995): 1-20. MRT has also been shown to be effective in reducing recidivism among juveniles in residential treatment facilities, (K.D. Burnette et al., "Treating youthful offenders with Moral Reconation Therapy: A Recidivism and pre-posttest analysis," *Cognitive Behavioral Treatment Review* 13, no.1 (2004): 3-5; and among those mandated MRT through a juvenile drug court (A. Wallace, "Results of Moral Reconation Therapy (MRT) utilization in the Las Cruces, New Mexico Drug Court," *Cognitive Behavioral Treatment Review* 10, no.3/4 (2001): 1-2. A recent meta-analysis of MRT—focused on adult and juvenile offenders either in custody or the community—reported that the intervention had a small but significant effect on recidivism; see L. Ferguson and J. Wormith, "A meta-analysis of moral reconation therapy," *International Journal of Offender Therapy and Comparative Criminology* 57, no.9 (2013): 1076-106. However, a study using a randomized experimental design to assess MRT's effectiveness on reducing juvenile recidivism found no effect; see T. Armstrong, "The Effect of Moral Reconation Therapy on the Recidivism of Youthful Offenders," *Criminal Justice and Behavior* 30, no.6 (2003): 668-687. MRT had never been implemented with a juvenile population in a large metropolitan jail prior to the ABLE program.

A social impact bond (SIB) was used to fund the operation of the ABLE program on Rikers. First introduced in the United Kingdom in 2010, a SIB is an innovative form of pay-for-success contracting that leverages private funding to finance public services; private investors fund an intervention through an intermediary organization and the government repays the funder only if the program achieves certain goals, which are specified at the outset of the initiative and assessed by an independent evaluator.[2]

Figure 1: Structure of ABLE/Rikers SIB financing
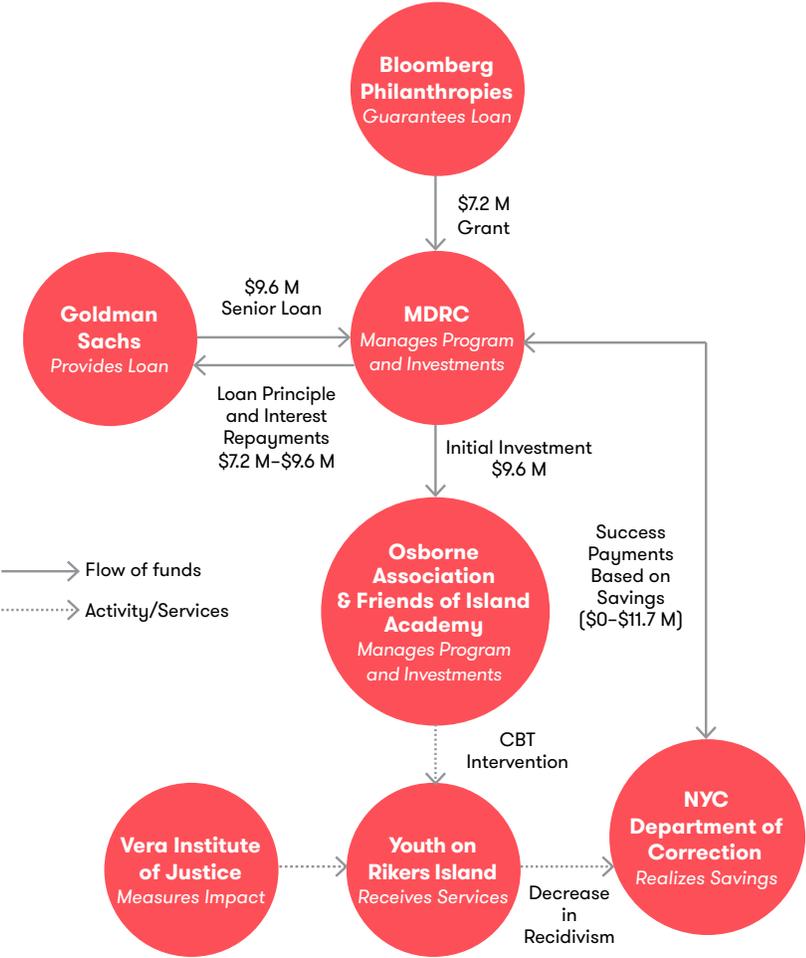
Smart Justice Advisory Council Briefing Book (2014). Smart Justice Advisory Council; Georgia. https://perma.cc/GXP3-KPTL

ABLE was the first program in the United States to be supported using a SIB funding structure. Financial support was provided by the Urban Investment Group of Goldman Sachs, which made a $9.6 million loan to

operate the ABLE program at Rikers. Bloomberg Philanthropies provided a $7.2 million grant to guarantee in part the Goldman Sachs loan.

Working with the City of New York, MDRC oversaw the implementation of the program, contracting with the Osborne Association and Friends of Island Academy, who administered the program on Rikers. The Vera Institute of Justice (Vera) served as the program's independent evaluator. The agreed-upon goal for ABLE was a 10 percent or greater decline in recidivism among participants, measured as readmission to the New York City jail for a new offense. If the program reduced reoffending by this amount or greater, the city would save on the cost of re-incarceration and use those savings to repay the investor.[3]

Vera's evaluation found that while the ABLE program reached the majority of adolescents, ages 16 to 18, who entered Rikers in 2013 and stayed for seven days or longer, it did not reduce recidivism and therefore did not meet the pre-defined threshold of success.

Evaluation

Aims and objectives

As ABLE's independent evaluator, Vera assessed whether the program achieved its target of reducing recidivism among program participants by at least 10 percent. Under the terms of the ABLE contract, a reduction in recidivism of 11 percent or more would provide an additional return to Goldman Sachs above the initial investment, with a graduated rate of return on investment up to a 20 percent reduction in recidivism.[4]

Design

To assess the program's impact on recidivism, the research team used a quasi-experimental research design to compare outcomes for program participants with a matched group of youth of the same age who entered Rikers between 2006 and 2010, before the ABLE program was established. The evaluation used a novel design, described in detail below to account for some of the temporal factors (such as changes in crime rates or policing policies) that can confound historical comparisons. Factors related to the operation of New York City jails precluded the possibility of a randomized experimental design.

There are two potential limitations of historical comparisons that could bias the evaluation of outcomes for ABLE participants. First, the profile of people entering the jail may have shifted over time, including demographics, prior arrest histories, and types of charges. These characteristics may, in turn, affect their risk of re-arrest irrespective of the interventions provided by the ABLE program. For example, compared to similarly-aged youth arrested in 2013, 16- to 18-year-olds arrested between 2006 and 2010 had fewer prior convictions, an indicator of the likelihood of future arrest.

The analysis controlled for these individual-level factors using a statistical technique called propensity score matching that is widely used in quasi-experimental studies to select groups that are matched on key baseline characteristics. In other words, 16- to 18-year-olds were selected from the 2006-2010 cohort for inclusion in the study because their attributes matched members of the ABLE program cohort. A more detailed description of propensity score matching and its application in this evaluation can be found on page 12.

Second, even if the groups are precisely matched to control for individual-level factors, wider shifts in the city's economy, crime trends, criminal justice policies, or other factors may also affect recidivism. In other words, if the same person incarcerated at Rikers was released in 2006 and in 2013, we might expect different recidivism outcomes because of changes in features of the environment. These shifts—or "system-level" factors—are unrelated to either the ABLE program or the individual characteristics of those entering DOC custody, but may influence the number of days people are held in the jail—and the likelihood that they will return to jail.

To control for the impact of these system-level factors, Vera also tracked shifts in recidivism over the same time period for 19-year-olds. This group was chosen as a control for system-level factors because 1) historical analysis of DOC records demonstrated that changes in recidivism for 19-year-olds followed a trend similar to 16- to 18-year-olds; and 2) 19-year-olds were not eligible to receive ABLE services and, therefore, any shifts in recidivism could be attributed solely to system-level factors.[5] The analysis used changes in the measure of recidivism for 19-year-olds to adjust the estimate of ABLE program effect. (See Figure 2.)

Vera used administrative records provided by MDRC and DOC to identify eligible participants entering custody during 2013 (the study cohort), select members of the comparison groups, and track program participation.

Figure 2: Measuring ABLE program effect

Factors that may influence recidivism	How these factors were accounted for
Individual factors: 16- to 18-year-olds entering DOC custody in 2006 to 2010 may differ from those admitted in 2013, such as age, gender, charge, and criminal history.	Young people from these two periods were matched based on 34 demographic and criminal justice characteristics to help ensure an "apples to apples" comparison.
System-level factors: Shifts in the city's economy, crime rates, policing practices, or other environmental factors between these two periods may affect recidivism outcomes.	Changes in recidivism were tracked over the same period for the matched sample of 19-year-olds (who were not eligible to receive ABLE services) to assess the influence of external factors.
ABLE effect: The cumulative effects of the ABLE program.	After accounting for the above factors, any changes in recidivism among 16- to 18-year-olds between 2006 to 2010 and 2013 were attributed to the ABLE program.
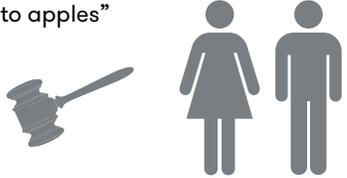
Methods

The sections below describe the research design in detail and outline key decisions that were made during the process of conducting the evaluation. For additional details of the research design, see Appendix 1.

The study cohort

The study cohort, or group that was the focus of the evaluation, was defined as 16- to 18-year-old youth who were admitted to Rikers between January 1 and December 31, 2013. The analysis of recidivism used arrest and incarceration data for the study cohort covering the period to the end of March 2015. To allow for a one-year post-release tracking period, only those youth who were released on or before March 31, 2014 were included in the cohort. In addition, only those youth who were held in the jail for seven days or longer were included in the cohort, regardless of whether they were held pretrial, serving a sentence, or in the jail for other reasons.[6] Of the 1,691 16- to 18-year-olds who were admitted to Rikers during 2013 and had at least one jail stay of seven days or longer, the majority (1,470 or 87 percent) were released prior to March 31, 2014.

A significant proportion (23 percent) of young people entering Rikers during 2013 had more than one admission during the year that met these criteria. In these cases, the research team began tracking recidivism from the date of release for the *first qualifying admission*, defined as the "index admission."[7]

Comparison groups

To control for individual-level factors: Researchers used 16- to 18-year-olds who were incarcerated at Rikers between 2006 and 2010. To ensure comparability, this group was identified using the same criteria as the study cohort (held for seven or more days and released by March 31 of the year following admission.)

To control for system-level factors: Researchers also tracked changes in recidivism for 19-year-olds incarcerated at Rikers over the same time period. To ensure comparability, this group was identified using the same criteria as the study cohort (admitted to Rikers in 2013, held for seven or more days, released by March 31, 2014) with one additional criterion: they must have been 19 years of age on January 1, 2013. Since ABLE services were provided to all 16- to 18-year-olds, this additional condition ensured that 19-year-olds included in the analysis could not have participated in the ABLE program.[8]

Vera calculated the reduction of recidivism by comparing the rate of recidivism for 16- to 18-year-olds who were eligible to participate in the ABLE program in 2013 with the recidivism rate for 16- to 18-year-olds who entered Rikers between 2006 and 2010 before the ABLE program was established. The change over time for 19-year-olds was then used to adjust for system-level changes.

Measuring effect

As the primary aim of the evaluation was to assess reductions in recidivism, the main outcome of interest was the number of days that youth are held in jail over a given time period (in this report we describe one-year recidivism outcomes).[9] This measure, referred to as "recidivism bed days" (RBD), combines the rates of return to the jail with the numbers of days that youth are held following each readmission.

Measurement of RBDs was restricted to "qualifying recidivism events," which included only those jail admissions that were connected with a new arrest. For example, following an initial spell in the jail as pretrial detainees, a number of cohort members were released and subsequently sentenced to jail on the same charge (without a new arrest). In other instances, members of the cohort were transferred from the jail to state prison and then returned to the city for a hearing or to serve as a witness in a trial. As these jail readmissions were not associated with reoffending behavior, they were not included as recidivism events when assessing program impact.

RBDs were calculated using individual-level records provided by the DOC. To calculate RBDs, all readmissions to Rikers within a 12-month period following release from the index admission were counted, including stays of any length.

Data sources

The ABLE evaluation was based on administrative records and program participation data provided by NYC government agencies and the agencies responsible for operating the ABLE program.

Department of Correction: The DOC provided complete records for individuals between ages 16 and 22 who were admitted to the jail from January 1, 2006 through March 31, 2015. DOC data included a rich array of variables, including demographic characteristics (e.g., race, sex, and age), borough of residence, characteristics of the charge (e.g., felony or misdemeanor, specific charge type), number of previous incarcerations, and assessed need

for mental health services. DOC records were used to identify members of the program cohort and comparison groups, provide information necessary to conduct propensity score matches, and measure RBD.

New York Police Department (NYPD): Vera researchers received NYPD records for all arrests involving 16- to 22-year-olds between January 2006 and March 2015. Vera merged DOC incarceration data with the NYPD arrest data to identify readmissions that were not associated with a new arrest, as described above.

MDRC: The nonprofit research and technical assistance agency, MDRC, was responsible for overseeing the implementation and operation of the ABLE program. MDRC tracked participation in and progress through the ABLE program using data collected by The Osborne Association and Friends of Island Academy, the agencies that delivered program services at Rikers. Vera researchers merged these data with records from the DOC, to assess program participation and achievement among the evaluation cohort.

Expert advisory committee

In order to guide the evaluation work and advise Vera on issues related to research design and analysis, Vera engaged an external panel of experts. The committee was comprised of J. Lawrence Aber, Wilner Family Professor in psychology and public policy at the Steinhardt School of Culture, Education and Human Development, New York University; Todd Clear, distinguished professor, School of Criminal Justice, Rutgers University – Newark; Andrew Gelman, professor of statistics and political science, Columbia University; Jennifer Hill, associate professor of social sciences at the Steinhardt School, New York University; and Michael Rempel, research director at the Center for Court Innovation. The committee met six times over the course of the study and advised on research design, statistical methodology, and model specifications.

Propensity score matching

As described above, the research team used propensity score matching to control for baseline differences between 2013 cases and the historical comparison group. Propensity score matching uses statistical techniques to define comparison groups using a single "score" or variable that accounts for multiple characteristics. This technique is often used as part of quasi-experimental evaluations to facilitate "apples to apples" comparisons, generating a more accurate estimate of program effect.[10] Researchers generated propensity scores using baseline characteristics and used these scores to identify individuals who were similar to those in the study cohort, but who were incarcerated before the ABLE program was implemented (the historical matched comparison group).

In order to create an accurate match between individuals incarcerated in 2013 and their historical counterparts, the propensity score models included 34 variables that were included in the records provided by DOC, including demographic characteristics, length of stay in the jail, severity of charge (felony or not), charge category (e.g., assault, robbery, etc.), diagnosed mental health status, and borough of residence. Details of the specific predictors included in the matching models, as well as the distribution of these measures for the evaluation and historical cohorts of 16- to 18-year-olds, are included in Appendix 2. Propensity scores were generated for 16- to 18-year-olds and 19-year-olds in separate models, each using an essentially identical set of predictors.[11] (For more detail of analytical techniques used, see Appendix 3.)

Findings

Vera's evaluation addressed five questions: 1) how many 16- to 18-year-olds admitted to Rikers in 2013 participated in ABLE; 2) what were their characteristics; 3) how many ABLE sessions did they attend and what programmatic milestones did they achieve; 4) what effect did the ABLE program have on their rate of recidivism; and 5) did the program have a greater or lesser impact on specific subgroups of program participants?

How many received ABLE?

To assess receipt of ABLE services, records from the DOC were merged with ABLE program records collected and provided to Vera by MDRC for all youth participating in at least one ABLE session between January 1, 2013, and March 31, 2015. These participation data not only included how many 16- to 18-year-olds at Rikers in 2013 participated in ABLE, but also the number of program sessions that participants attended and how many programmatic milestones they achieved.

Analysis of participation data for the 16- to 18-year-olds in the study cohort shows that the program reached the great majority of youth whose incarcerations met the criteria to be included in the evaluation. Among the 1,470 youth included in the study cohort, 85 percent (1,255) participated in at least one ABLE session between January 1, 2013 and March 31, 2015.[12]

Figure 3: ABLE participation in 2013 by quarter
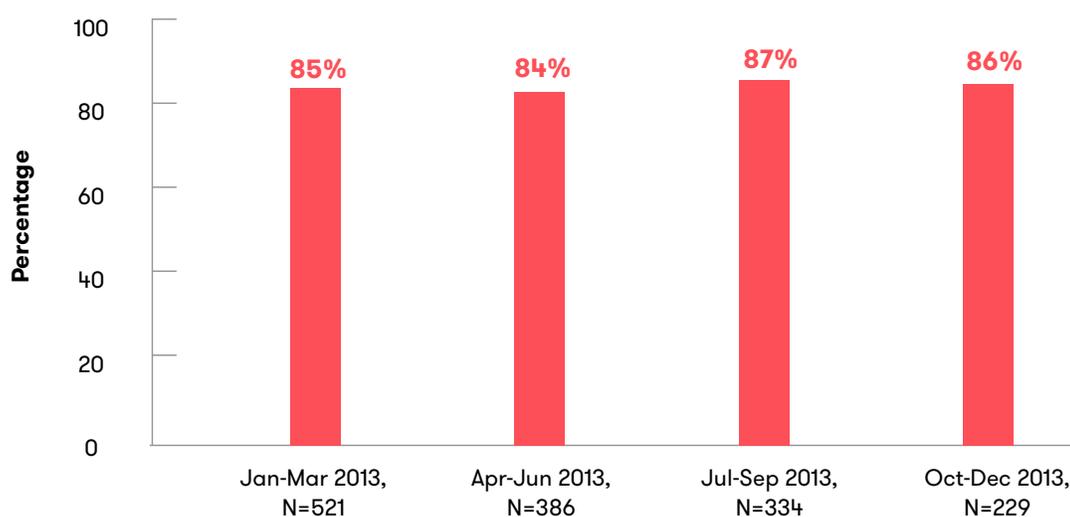
Note: The declining number of admissions over the years is partly accounted for by repeat admissions; an adolescent who entered the jail more than once during 2013 is included only in the count for the quarter when they were first admitted.

Characteristics of ABLE participants

As Figure 4 below shows, of the 1,470 adolescents in the study cohort, nearly 90 percent were male. African American and Latino people accounted for 95 percent of the overall group. Just over four out of 10 (43 percent) of the cohort were age 18 when they were first admitted during 2013, 36 percent were age 17, and 21 percent were age 16.[13]

Figure 4: Demographic characteristics of study cohort, N= 1,470
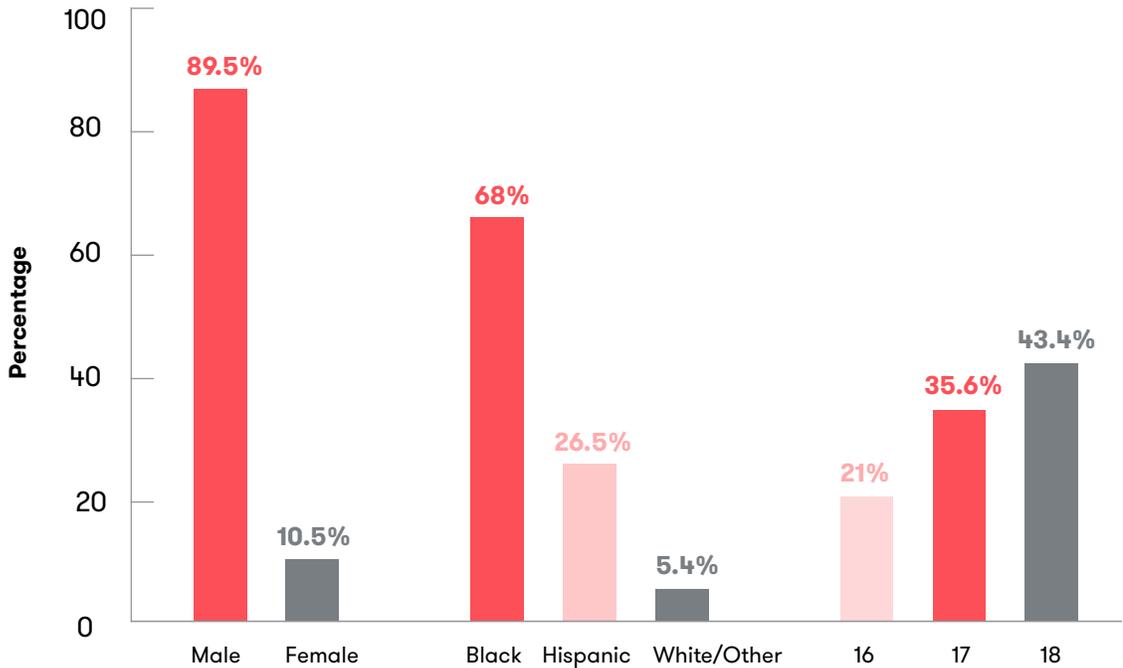
Note: Totals may not equal 100 percent due to rounding.

Two-thirds of the study cohort (66 percent) were charged with a felony as the top charge at admission as shown in Figure 5 below.[14] Twenty-three percent had a misdemeanor as their top charge at admission. With respect to the type of charge, just over half of the cohort (52 percent) faced a top charge for a violent crime. An additional 20 percent were charged with a property crime, 8 percent had a weapons charge as their top charge, 5 percent were charged with a drug crime, and the remainder faced a charge related to public order or another type of offense.

Figure 5: Top charge type and severity for study cohort, N= 1,470
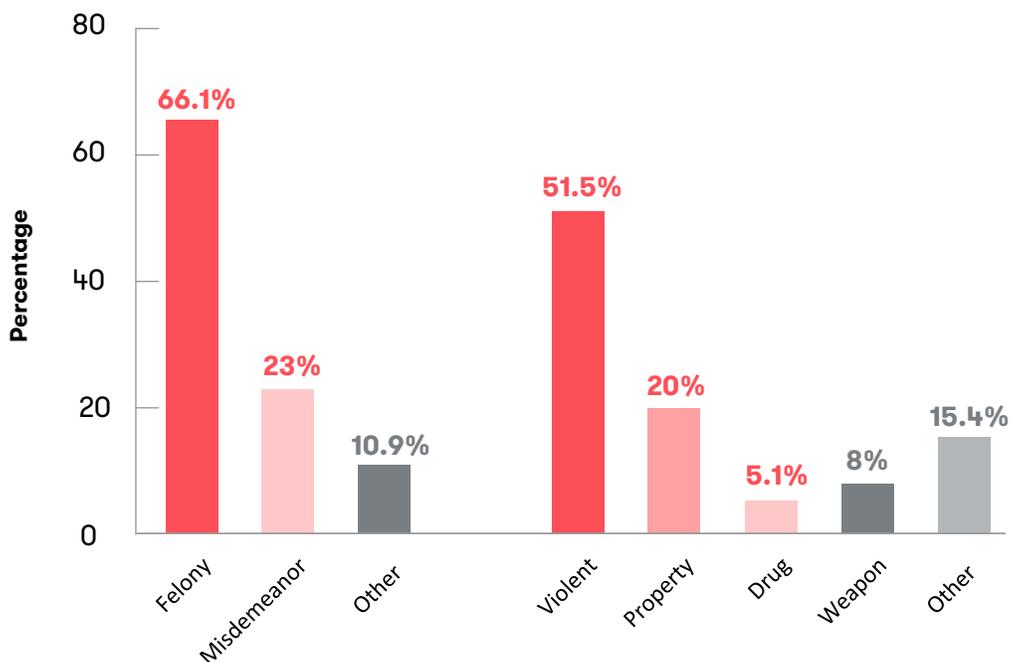
The majority of adolescents in the study cohort (53 percent) were released from Rikers as pretrial detainees.[15] Roughly equal percentages were serving their sentence at Rikers (23 percent) or awaiting transfer to state prison, or another agency or jurisdiction (22 percent). The remainder of the study cohort were released on a conditional discharge (1 percent) or had their cases dismissed (2 percent).[16]

How many ABLE sessions and milestones did participants achieve?

Adolescents in the study cohort attended a median of 12 sessions. Twenty-seven percent of the cohort attended three or fewer sessions, while 11 percent attended 50 or more.

The curriculum for the ABLE program, based on the Moral Reconation Therapy model, is stage-based.[17] These stages are operationalized as distinct "steps," with individuals progressing from one step to the next as they advance in the program. [18] The key programmatic milestones are reached at Step 3, Step 7, and Step 12.[19]

Of the 1,255 youth in the study cohort who attended at least one ABLE session, the majority (69 percent) achieved at least Step 1. Just over half (52 percent) reached one of ABLE's key milestones, achieving Step 3 or higher. In addition, more than one in 10 (11 percent) of the 16- to 18-year-olds in the study cohort reached Step 12, completing the program. Thirty-one percent of those who participated in the program did not achieve the first step. Figure 6 shows the level of program completion achieved by all ABLE participants in the study cohort.

Figure 6: Level of program completion achieved by ABLE study cohort

- No Step Achieved
- Step 1-2
- Step 3-5
- Step 6-11
- Step 12

11%
13%
28%
17%
31%

The Evaluation Cohort, N=1,225
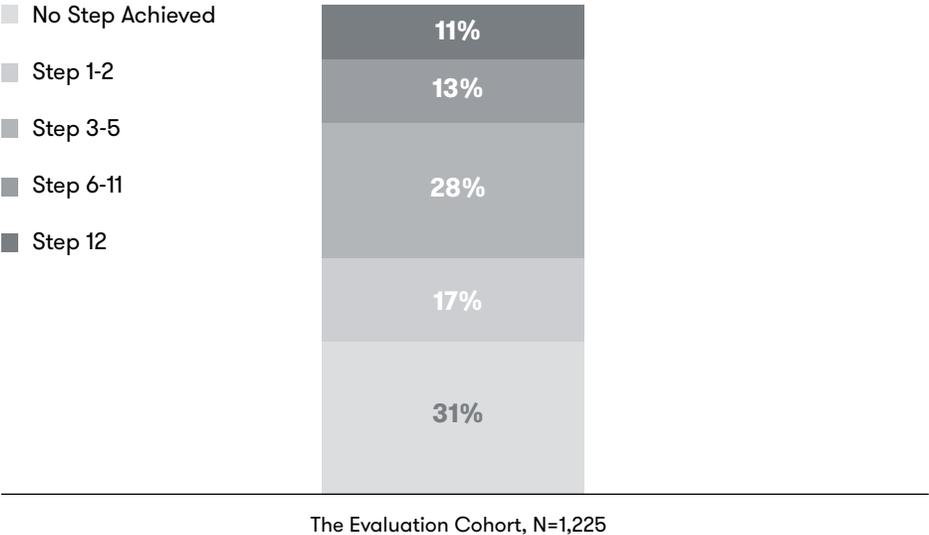
Note: These percentages are based on only those individuals in the study cohort who received any ABLE services. Not included are the 15 percent of the study cohort who did not attend at least one ABLE session during their incarceration.

Program effect analysis

The primary purpose of the evaluation was to determine whether the implementation of the ABLE program reduced recidivism among eligible 16- to 18-year-olds admitted to Rikers in 2013 (the study cohort).[20] More specifically, the evaluation assessed any change in the average number of days these adolescents spent in jail (measured as recidivism bed days or RBD) following arrest on a new charge in the 12-month period following their initial release.

As a first step, the Vera researchers compared the change in RBD for the ABLE study cohort with the matched historical comparison group. Among the study cohort, the average number of RBDs was 37.2.[21] The same outcome among the matched sample of youth incarcerated at Rikers between 2006 and 2010 was 32.8. The difference in outcomes between these two groups represents a 13.4 percent increase in recidivism for the ABLE cohort, without controlling for system-level factors that may impact recidivism.

To assess the impact of system-level factors on recidivism, researchers compared RBDs for matched samples of 19-year-olds. The average number of RBDs for 19-year-olds incarcerated between 2006 and 2010 was 26.8 days, while among those incarcerated in 2013, the average number of RBDs was 28.9 days; equivalent to an 8.0 percent increase in RBDs for 19-year-olds who did not receive ABLE services. The analysis assumes that, absent of any program effect, recidivism among 16- to 18-year-olds would have increased by the same amount as 19-year-olds. In other words, 16- to 18-year-olds would have experienced an 8 percent increase in RBDs as the result of factors that are unconnected to the ABLE program. Figure 8 below shows the changes in RBDs in study and comparison cohorts.

Figure 7: Changes in recidivism bed days (RBDs) in study and comparison cohorts
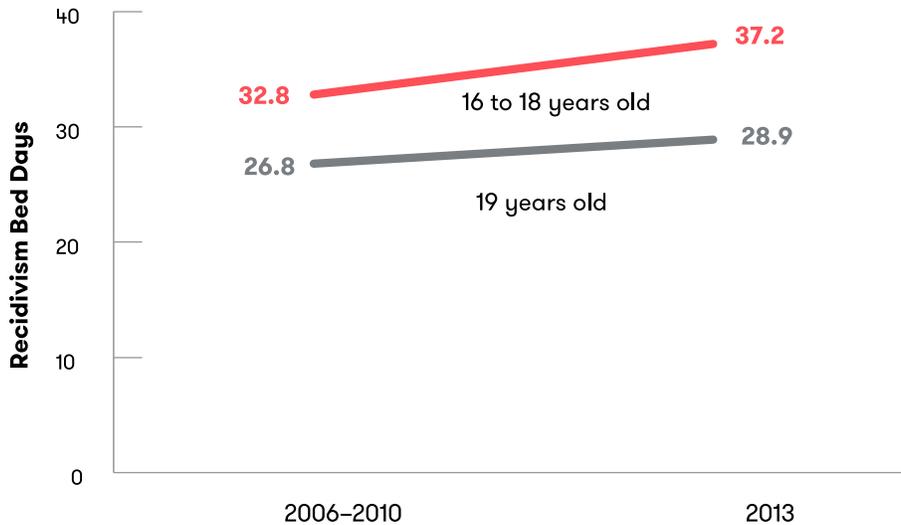
In the final step of the analysis, Vera researchers used the calculated change in RBDs for 19-year-olds to adjust for that of the 16- to 18-year-olds in order to estimate the program effect.

Results of RBD Analysis

Change in RBDs for 16- to 18-year-olds (ABLE cohort)	13.4% increase between 2006 to 2010 and 2013
Change in RBDs for 19-year-olds (system-level factors)	8.0% increase between 2006 to 2010 and 2013
Adjusted ABLE program effect (ABLE cohort — system-level factors)	**Non-significant increase (5.1%) between 2006 to 2010 and 2013**

To adjust for system-level factors, the RBDs for the historical comparison group were increased by 8 percent. In other words, if members of the historical comparison group were arrested in 2013, they would be expected to accrue 8 percent more RBDs because of shifts in system-level factors in the city. If recidivism among the study cohort had increased by the same amount as it did among 19-year-olds (8.0 percent), the average number of RBDs for the comparison cohort would have been 35.4 days (the adjusted

comparison cohort's RBDs). See Appendix 1 for further details of the formulae used to measure program effect.

The actual value for RBDs among the study cohort is 37.2 days, which is 5.1 percent *higher* than the adjusted comparison cohort RBD. This represents the adjusted program effect. While this finding may appear to suggest that the program was associated with higher levels of recidivism, this is not the case. The difference is not statistically significant and therefore is equivalent to a zero program effect.[22]

In summary, Vera's analysis shows a general upward trend in RBDs for both the study and comparison cohorts, indicating that changes in recidivism were the result of factors other than the ABLE program.

Subgroup analysis

Vera researchers also assessed whether ABLE was beneficial for any subgroup of the study cohort. They separately analyzed RBDs for subgroups by gender, age, race, assessed need for mental health services, and whether the top charge at discharge was a felony, classified as violent, or a property offense. Outcome data were also analyzed separately for cohort members admitted to Rikers during the first six months of 2013. This last factor, based on month of admission, was included to explore whether the ABLE program might have been more effective in the second half of the year, after adjustments were made to address any startup difficulties.

Vera researchers compared the average number of RBDs in these study cohort subgroups with those for comparison cohort subgroup members using the propensity score matching technique described on page 12. Overall, these analyses did not show any significant differences in recidivism for any of the subgroups examined. See Appendix 1 for further details of this analysis.

Finally, Vera researchers explored the impact of different rates of program participation or "dosage." Specific measures included the number of sessions that a participant attended and the programmatic level or step they achieved. However, the analysis of dosage was confounded by length of stay in the jail; although adolescents who were held for longer periods received a higher dose of services, their length of stay was also highly correlated with recidivism. Researchers controlled for length of stay by comparing the RBDs within dosage subgroups for those whose stay in Rikers was 60 days or longer with RBDs of those whose initial stay was shorter. This analysis did not detect significant differences in recidivism outcomes. See Appendix 1 for further details of the subgroup analysis.

Conclusion

———————

Overall, Vera researchers determined that the program did not lead to reductions in recidivism for youth incarcerated at Rikers. Researchers found that recidivism for the study cohort of 16- to 18-year-olds, adjusted for individual and system-level factors, did not achieve the pre-determined threshold of success, a 10 percent reduction in recidivism bed days. Moreover, the study and comparison cohorts displayed similar upward trends in rates of recidivism over time, indicating that any shifts were the result of factors other than the ABLE program. As a result of these findings, the ABLE program was discontinued on August 31, 2015.

Vera's evaluation was not designed to collect and analyze information on how the intervention model was provided at Rikers. MRT had never before been implemented at this scale nor had it been tested with an adolescent population in a large urban jail. More research is needed to understand whether the model could be adapted for an environment as challenging as that of Rikers or, if not, what types of intervention are better equipped to support adolescents as they transition out of the jail.

While the evaluation found that the ABLE program as implemented was not effective at addressing the problem of recidivism, the social impact bond funding mechanism operated as intended. The program included a rigorous outcome evaluation that demonstrated that the intervention was not effective and should be discontinued.

Vera's evaluation of the ABLE program contributes to the growing body of research on the effectiveness of specific interventions aimed at reducing recidivism. The fact that the program was quickly taken to scale demonstrates the feasibility of large-scale programming within correctional facilities.

Research Design and Methods

This technical appendix expands upon the descriptions of methods included in the main body of the report. The following section provides additional details on the definition of the evaluation cohort, selection of comparison groups, recidivism measures, evaluation design, and the specifications of analytic models.

Study cohort and comparison groups

As noted in the main body of this report, the research compared outcomes for ABLE participants and a historical comparison group of 16- to 18-year-olds who were admitted to the New York City jail between 2006 and 2010.[23] The research team used propensity score matching to create two groups drawn from these time periods with similar demographic characteristics, charging patterns and histories of criminal justice system involvement.

To create the historical comparison cohort, data on 16- to 18-year-olds admitted to the jail between 2006 and 2010 were first grouped by year, and then cleaned and processed (e.g., categories of charge type were created from the specific charge). These five sets of annual data were then combined to create a pool of potential comparison cases.

Some individuals had multiple incarcerations within the same calendar year; for example, the same person could have been admitted to the jail in both February and October of 2007. To ensure equivalence between treatment and comparison cases, the first event that met the study's inclusion criteria (length of stay of seven or more days, released by March 31 of the following year, etc.) was selected as the index event.

In addition, there were a number of individuals who were incarcerated in multiple years when they were between the ages of 16 and 18. For example, a person might have been held at Rikers during 2008, when they were age 17, and again in 2009 at age 18. Including multiple entries for the same individual in the pool of potential comparison cases could lead to biased estimates, since those individuals incarcerated in multiple years were, by definition, more likely to recidivate. Duplicate records were eliminated through a process of random selection, leaving one record per individual in the resulting pooled data file.

Propensity score matching

Propensity score matching techniques are frequently used as part of quasi-experimental evaluations to create valid comparison groups in order to estimate the effect of a program on some outcome or outcomes. Propensity scores are attractive because they summarize multiple co-variates into a single metric that can then be used for matching purposes. In consultation with the External Expert Advisory Committee, Vera elected to use a technique, called caliper matching, to select comparison group cases. This technique involves defining a numerical range around each treatment case's statistically generated propensity score. All comparison group cases with propensity scores within this range are defined as matches. This study employed a 0.0009 caliper and the propensity score matching was conducted "without replacement," meaning that each comparison group case was used only once in the matching procedure.

Propensity score models – measures included as predictors

As described in the body of this report, propensity score matching was used to conduct historical matches for cases involving 16- to 18-year-olds (to account for individual-level changes) and 19-year-olds (to account for system-level changes). The same predictors were included in both models—specifically gender, race, borough of residence, whether the individual was "discharged to the street" or not (whether he or she was free to recidivate upon release), whether he or she was transferred to state prison, whether the individual was released pretrial, whether the admission charge was a felony, the number of prior admissions, age at admission, the length of stay for the qualifying ("index") event, the individual's assessed need for mental health services, and the charge that resulted in the admission to Rikers.

Charges at admission were grouped into 24 distinct, non-overlapping measures, all of which were included in the match, save those for arson, bribery, vulnerable persons, and white collar crimes. As might be expected among people under age 20, there were an insufficient number of cases with these charges to include them in the analysis. Finally, a small number of individuals had very high values for the number of prior incarcerations at Rikers and the number of prior events was capped at six to remove outliers.

Assessing balance

As an initial step, researchers assessed the balance of the measured variables, or the level of similarity in the distribution of observed covariates between treatment and comparison groups. Balancing measured covariates reduces bias in the estimate of the treatment effect.[24] Balance was assessed using two different statistical procedures, one (using the pstest command in Stata) is a standard measure of testing covariate balance. The second method is based upon standardized differences.

The results of the balance in covariates, both before and after matching, can be found in Appendix 3. Two rows of data are presented for each of the characteristics shown in the table: "unmatched" and "matched" values for both the 2006-2010 and 2013 cohorts. The "unmatched" values correspond to the unadjusted values for the two groups. For example, 89.5 percent of the 2013 cohort was male, compared to 92.1 percent of the 2006-2010 group. The "matched" values display the characteristics after statistical adjustment through propensity score matching. For the values for males, the percentage of the 2013 cohort remains 89.5 percent, while the historical cohort group is 89.8 percent.

To summarize the findings from this table: the propensity score matching created two similar groups based on a range of characteristics that are theoretically associated with recidivism (e.g. demographics, current charge, and criminal history). The unmatched values highlight some of the differences between the historical comparison and evaluation cohorts. Among those to note are:

> compared to the 2006-2010 cohort, a somewhat greater proportion of youth incarcerated in 2013 were African American and fewer were Latino. The percentage of white youth entering DOC custody is essentially the same for both periods;
> the proportion of youth entering the jail from Manhattan increased compared to a minor decrease in the proportion of youth from Brooklyn;
> the average number of prior incarcerations increased slightly between 2006 to 2010 and 2013;
> the percentage of 16- to 18-year-olds with felony charges at admission was slightly lower among the evaluation cohort than the historical comparison group; and
> a greater proportion of the 2013 cohort was released pretrial (52.5 percent) compared to the historical comparison group (40.8 percent).

The largest difference between the historical and evaluation cohorts is in the percentage of individuals identified as requiring mental health services, which grew from 23 percent among the 2006-2010 cohort to 41 percent in the 2013 evaluation cohort. This increase may be partly due to the procedures used to identify individuals in need of mental health services—and the increased likelihood that youth with a history of receiving mental health services at Rikers will be identified in subsequent incarcerations.

Calculating adjusted program effect

The following formulae were used to calculate the overall (adjusted) program effect.

Figure 8: Formulae for calculating overall program effect

$$\text{Adjusted comparison RBDs (accounting for individual- and system-level changes)} = \text{\% change in RBDs for 19-year-olds, 2006 to 2010 and in 2013} \times \text{RBDs of matched 16- to 18-year-olds, 2006 to 2010}$$

$$\text{Adjusted program effect} = \frac{(\text{Actual RBDs for ABLE cohort}) - (\text{adjusted comparison RBDs})}{\text{adjusted comparison RBDs}}$$

Assessing statistical significance

Based on these formulae, the study cohort of 16- to 18-year-olds as a whole, the *unadjusted increase* in recidivism was 13.4 percent (from 32.8 among the 2006-2010 cohort to 37.2 among the 2013 study cohort). Once adjusted for system-level changes, the overall adjusted program effect is a non-significant positive 5.1 percent. This is statistically equivalent to a zero program effect.

In order to test the significance of this difference, we pooled data from the historical and 2013 cohorts for both 16- to 18-year-olds and 19-year-olds into one large data set. We then followed a procedure for assessing statistical significance similar to that used in a difference-in-difference model. Specifically, dummy variables were created for year (2013 vs. 2006 to 2010) and age (16 to 18 vs. 19) and an interaction term (equal to 1 if in the 16- to 18-year-old cohort in 2013). A regression model predicting RBD was run with the full set of predictors used in the propensity score matching models and these dummies. The interaction variable was not statistically significant in the model, indicating that the increase of 5.1 percent was not statistically significant.

Subgroup analysis

Further analysis demonstrated that recidivism among most subgroups increased by about the same amount. For example, as Figure 9 on page 27 shows, the average *unadjusted* number of RBDs among African American youth increased from 33.3 days among the 2006-2010 cohort to 37.6 days among the 2013 study cohort, an increase of 12.9 percent (compared to an *unadjusted* increase for the entire sample of 13.4 percent). Among youth of other races, RBDs increased by 14.1 percent, from 31.9 days to 36.4 days. Similarly, the average number of RBDs for adolescents facing felony charges increased by 13.9 percent (from 32.3 days between 2006 to 2010 to 36.8 days in 2013), which tracks closely the 12.5 percent increase between these two time periods for adolescents facing other, non-felony charges (33.8 to 38.0 RBDs).

Figure 9: Examples of subgroup differences (16- to 18-year-olds)

	2006-2010 comparison cohort RBDs	2013 program cohort RBDs	Difference
OVERALL	32.8	37.2	13.4%
African American	33.3	37.6	12.9%
Other Race	31.9	36.4	14.1%
Felony Charge	32.3	36.8	13.9%
Non-felony Charge	33.8	38.0	12.5%

Vera researchers also tested several approaches to examine differences based upon measures of participation in the ABLE program. This investigation was designed to assess whether higher "dosage" levels of the ABLE program might be related to lower rates of recidivism. Researchers aimed to assess the relationship between the number of ABLE sessions attended or the level of Step achievement and recidivism outcomes. However, these results were deemed to be unreliable, since all measures of ABLE participation or dosage were so highly correlated with the length of stay in the jail—which is itself highly correlated with recidivism. The degree of interrelatedness between these measures meant that it was not possible to disentangle dosage measures, length of stay, and future recidivism. These measures were not used in the final subgroup analysis.

Characteristics of the Evaluation Cohort

	Evaluation Cohort, N=1,470	
	Number	**Percent**
Gender		
Female	155	10.5
Male	1,315	89.5
Race/Ethnicity		
Black	1,000	68.0
White non-Hispanic	34	2.3
Hispanic	390	26.5
Other	46	3.1
Age		
16 yr olds	309	21.0
17 yr olds	523	35.6
18 yr olds	638	43.4
Identified as requiring mental health services	603	41.0
Top charge type at admission		
Violent	757	51.5
Property	294	20.0
Drug	75	5.1
Weapon	117	8.0
Public order	58	3.9
Other	110	7.5
Missing	59	4.0
Top charge severity at admission		
Felony	971	66.1
Misdemeanor	338	23.0
Other	102	6.9
Missing	59	4.0
Discharge status		
City sentenced	337	22.9
Pretrial release	772	52.5
Conditional discharge	15	1.0
State sentenced/cont. incarceration	314	21.5
Case dismissed	32	2.2
Average length of stay	67.8 days	

Characteristics of 16- to 18-year-olds (treated), 2013 Study Cohort and Historical Comparison Cohort (control)

	Unmatched Treated	Unmatched Control	Matched Treated	Matched Control
Demographics				
Male	0.895	0.921	0.895	0.898
African American	0.680	0.616	0.675	0.672
Latino	0.265	0.318	0.269	0.269
Brad H*	0.410	0.233	0.400	0.402
Borough				
Queens	0.196	0.204	0.197	0.205
Manhattan	0.199	0.146	0.193	0.181
Bronx	0.227	0.244	0.231	0.233
Brooklyn	0.254	0.316	0.256	0.261
Discharge status				
Discharged to street	0.786	0.771	0.786	0.790
Transferred to prison	0.134	0.132	0.134	0.133
Released pretrial	0.525	0.408	0.521	0.515
Felony charge at admission	0.661	0.730	0.675	0.664
Top charge				
Solicitation & conspiracy	0.012	0.003	0.009	0.008
Assault	0.148	0.138	0.151	0.154
Homicide	0.020	0.023	0.021	0.020
Sex crime	0.012	0.017	0.013	0.013
Kidnapping	0.001	0.002	0.001	0.001
Burglary	0.070	0.063	0.072	0.071
Vandalism	0.014	0.015	0.014	0.013
Theft	0.106	0.083	0.109	0.109
Fraud & forgery	0.010	0.005	0.010	0.011
Robbery	0.333	0.345	0.343	0.335
Drugs	0.051	0.140	0.053	0.052
Custody & procedural	0.027	0.027	0.027	0.029
Public order	0.010	0.012	0.010	0.012
Prostitution	0.001	0.002	0.001	0.002
Weapons	0.080	0.075	0.082	0.083
Other violations	0.003	0.004	0.003	0.003
Warrant	0.029	0.005	0.012	0.018
VTL	0.001	0.002	0.001	0.001
Parole/Probation violation	0.001	0.008	0.001	0.002
Number of prior admits	0.869	0.751	0.847	0.867
Average age at admission	17.224	17.274	17.222	17.216
Length of stay	67.822	69.827	67.453	67.684

*Identified as requiring mental health services.

Endnotes

1 Timothy Rudd, Elisa Nicoletti, Kristin Misner, and Janae Bonsu, *Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond* (New York: MDRC, 2013).

2 There are a number of social programs funded by social impact bonds, primarily in the U.S., UK, and Australia, focusing on issues related to criminal justice, education, homelessness, and unemployment.

3 For additional information about the organizations involved in the SIB and their roles, see Timothy Rudd, et al., MDRC, 2013.

4 For specific levels of return on investment by level of reduction in recidivism, see "Fact Sheet: The NYC ABLE Project for Incarcerated Youth" (New York: Office of the Mayor, 2012), https://perma.cc/6X25-7ZS5.

5 The 19-year-old cohorts were matched on key characteristics to control for individual level differences using the same procedures as 16- to 18-year-olds.

6 Initially, the evaluation included all youth who were jailed for five or more days; however, the threshold was extended after analysis conducted by Vera in early 2014 found that youth who were in custody for six days or fewer rarely had the opportunity to receive ABLE services, for a range of reasons that were beyond the control of the agencies implementing the program; the majority of those incarcerated in the New York City jail are either held pretrial or serving sentences. Pretrial detainees have not been convicted of a crime but are being held on bail or have been remanded to custody. Sentenced youth have been convicted and are serving a custodial sentence of a year or less in the jail. A smaller number of people held are awaiting transfer to another facility or jurisdiction, including immigration detainees, and sentenced inmates en-route to state prison facilities.

7 If an individual had only one admission during the year, that event was defined as his or her index admission.

8 The ABLE program was piloted among 16- to 18-year-olds in Rikers in early 2012—and some of the individuals in the 19-year-old comparison group participated in the pilot when they were age 18. Specifically, five percent (n=23) of the comparison group of 19-year-olds participated in the ABLE pilot. These cases were included in the analysis.

9 The original design of the evaluation called for assessing recidivism within 12 and 24 months following release. However, after the findings of Vera's evaluation were released in early July 2015, the city decided to discontinue the ABLE program. See "Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program at Rikers Island" (New York: Vera Institute of Justice, 2015), https://perma.cc/6986-AFHD.

10 In evaluations that use experimental designs, such as randomized control trials (RCTs) in which individuals are randomly assigned to receive a treatment, the process of random assignment usually creates groups that are similar. Through randomization, assignment to either the treatment or non-treatment group does not depend on any background characteristic, thus minimizing selection bias.

11 The model for 19-year-olds contained 33 predictors, since the measure of age did not vary among those included in the model.

12 Adolescents admitted to DOC custody more than once during the one-year follow up period may participate in ABLE during multiple stays.

13 Appendix 1 contains a detailed description of the study cohort.

14 An individual can be charged with multiple crimes in one criminal incident. We report and analyze the offense that attracts the most severe penalty—commonly referred to as the "top charge." Top charge type and severity were measured using the top charge information at admission to the jail.

15 This proportion is lower than that of the overall population because those held for fewer than seven days are excluded from the study cohort.

16 Totals do not equal 100 percent due to rounding.

17 Timothy Rudd, et al., MDRC, 2013, 30-32.

18 Ibid.

19 These steps were designated as the key milestones in consultation between MDRC and Correctional Counseling, Inc., the developer of MRT (D. Butler, personal communication, October 14, 2015).

20 See Methods section on page 10 for a discussion of eligibility criteria.

21 Values of RBD throughout this section are adjusted through propensity score matching. For more information, see page 12.

22 The methods used to assess the statistical significance of this difference are described in the Statistical Significance section of Appendix 1.

23 All 16- to 18-year-old youth who met the qualifying criteria were included in the "treatment" group, regardless of whether they received ABLE services, since the evaluation assessed the reduction of recidivism overall, not only for those who participated in the ABLE program.

24 P.R. Rosenbaum, *Design of Observational Studies* (New York: Springer, 2009). K. Imai and D.A. Van Dyk, "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association* 99 (2004): 854-866.

Acknowledgments

The authors would like to thank Kristin Misner-Gutierrez from the New York City Mayor's Office, Katie Appel Duda at Bloomberg Philanthropies, and Andrea Phillips for their support throughout this project. We are grateful to our colleagues at the Osborne Association, Friends of Island Academy and MDRC for sharing their insights and experiences administering the ABLE program. In particular, we would like to acknowledge Elizabeth Gaynes and Susan Gottesfeld at the Osborne Association, and David Butler, Cindy Redcross, and Timothy Rudd at MDRC. We would also like to thank the New York City Mayor's Office and Department of Correction and notably DOC Commissioner Joseph Ponte for supporting this project and facilitating access to the jail, and Eric Sorensen for providing essential data for this study. We are immensely grateful to our expert advisors who were extremely generous with their time and expertise: J. Lawrence Aber, Todd Clear, Andrew Gelman, Jennifer Hill and Michael Rempel. A special thank you to our current and former Vera colleagues who helped develop the project and provided ongoing advice and support: Shelley Azumbrado, Reagan Daly, Ayesha Delany-Brumsey, Chelsea Davis, Adair Iacono, Michael Jacobson, Julia Kennard, and Vera's Director, Nick Turner. We are particularly grateful for the support of Vera's communication team who assisted with the editing, design and production of this report: Mary Crowley, Ram Subramanian, Erika Turner, and Patricia Connelly. The evaluation of the ABLE program is supported by the Mayor's Fund to Advance New York City, a not-for-profit corporation incorporated under the laws of the State of New York.

About Citations

As researchers and readers alike rely more and more on public knowledge made available through the Internet, "link rot" has become a widely acknowledged problem with creating useful and sustainable citations. To address this issue, the Vera Institute of Justice is experimenting with the use of Perma.cc (https://perma.cc), a service that helps scholars, journals, and courts create permanent links to the online sources cited in their work.