

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Evaluating a Cognitive Behavioral Approach for Improving Life Outcomes of Underserved Young Women: A Randomized Experiment in Chicago

Author(s): Harold Pollack, Jonathan Guryan, Monica Bhatt

Document Number: 310163

Date Received: February 2025

Award Number: 2016-JU-FX-0002

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

FINAL REPORT

U.S. Department of Justice Office of Justice Programs
Office of Juvenile Justice and Delinquency Prevention

Project/Award Number: 2016-JU-FX-0002

Project Title: Evaluating a Cognitive Behavioral Approach for Improving Life Outcomes of
Underserved Young Women: A Randomized Experiment in Chicago

Principal Investigators:

Harold Pollack

Faculty Co-Director, University of
Chicago Crime Lab and Health Lab
Professor, University of Chicago
School of Social Service
Administration
haroldp@uchicago.edu
773-702-4414

Jonathan Guryan

Faculty Co-Director, University of
Chicago Education Lab
Professor, School of Education and
Social Policy at Northwestern
University
j-guryan@northwestern.edu
773-848-9408

Monica Bhatt

Senior Research Director, University of
Chicago Education Lab
mbhatt@uchicago.edu
773-834-9960

Submission Date: December 29, 2022

Project/Grant Period: October 1, 2016 to September 30, 2022

Reporting Period End Date: September 30, 2022

Report Term or Frequency: Semi-annual

Current IRB Approval/Renewal and Expiration Date: Approved 2/28/2022; Expires 2/27/2023

Goals and Accomplishments:

- *What are the major goals and objectives of this project?*

Goals: We sought to work closely with Youth Guidance to learn more about the effectiveness of cognitive behavioral mentoring (CBM) programs as a means to improve mental health outcomes for adolescent girls growing up in Chicago neighborhoods that are experiencing dramatically elevated levels of crime. To do so, we worked with Youth Guidance to execute a rigorous randomized controlled trial (RCT) testing the Working on Womanhood (WOW) intervention in academic years (AY) 2017-18 and 2018-19. In the face of increasingly constrained budgets, it becomes even more important to identify and test programs that are both effective and cost-effective to help policymakers focus scarce resources on efforts that accomplish the most social good per dollar spent. At the beginning of the grant period we aimed to:

1. Carry out the first experimental test of a school-based CBM program for young women that seeks to improve mental health outcomes first and foremost, as well as understand how the program affects schooling outcomes and violence involvement, including victimization;
2. Estimate the cost-effectiveness of this intervention and compare to other strategies that aim to improve life outcomes among youth, especially young women.

Objectives: During the RCT, we planned to:

1. Partner with Youth Guidance, the Chicago Public Schools (CPS), and partner schools to use administrative data to identify a sample of ~2,500 eligible 9th, 10th, and 11th grade girls at schools in high violence Chicago community areas, all of whom are at elevated risk for violence exposure and/or violent victimization;
2. Randomly assign eligible students to either be offered the chance to participate in WOW or to be in a control group receiving status-quo school and community services; oversee and support participant recruitment to ensure program enrollment targets are met (so that the study will have sufficient statistical power); provide replacement participants if program slots open up to preserve the integrity of the study; and work with Youth Guidance on participant retention;
3. Monitor and measure program participation and fidelity of program implementation;
4. Access administrative records from the Chicago Police Department to measure victimization and juvenile and adult-system arrests for violent and other crimes;
5. Access student-level CPS records to measure other forms of delinquency such as truancy, dropout, and disciplinary actions and to measure school engagement and persistence;
6. Carry out in-person survey data collection to gather data on symptoms of anxiety, depression, and PTSD, as well as a range of other mental and social health outcome indicators;

7. Should we gain access to such data sets, obtain data from the Illinois Department of Public Health and/or the Chicago Department of Public Health to measure reproductive and sexual health outcomes captured by administrative data;
 8. Conduct statistical analyses comparing violent behavior and delinquency, victimization, schooling outcomes, and mental/reproductive health outcomes across treatment and control groups, and conduct comprehensive benefit-cost analysis; and
 9. Disseminate findings broadly to reach academic, policymaker, and practitioner audiences.
- *What was accomplished under these goals during the reporting period?*

In the most recent reporting period (July 1, 2022 through September 30, 2022), we accomplished all goals we established for this timeframe, specifically focusing on objective 9, as outlined below. (Objectives 1-6 and Objective 8 are complete. Objective 7 was unable to be accomplished due to constrained capacity in city and state health organizations during the COVID-19 pandemic.)

Objective 9: Disseminate findings broadly to reach academic, policymaker, and practitioner audiences.

- We have continued to hold meetings with WOW leadership and the Chicago Public Schools to discuss project updates, analysis updates and publication updates as needed. This open line of communication allows us to solve problems collaboratively and identify areas where our respective organizations can offer support to each other over the course of the study, such as obtaining pertinent implementation data from Youth Guidance for our ongoing data analysis.
- Our research team has drafted a paper presenting analyses from Cohort 1 girls' first year in the WOW program that indicate significant impact on three key aspects of girls' mental health: depression, anxiety, and post-traumatic stress disorder (PTSD). We submitted this paper to two peer-reviewed journals: *Science* and *Science Advances*. While *Science* did not opt to publish the paper, *Science Advances* has requested that we revise and resubmit our manuscript in December 2022. Our objective is to highlight the implications for policies and funding to support school-based programs available to young women who need additional mental health supports and services. The results to be published in a peer-review journal will include not only results on the program impact on our three primary mental health outcomes but also on the program's cost-effectiveness and implementation data analyses. Youth Guidance and Chicago Public Schools both approved the publication prior to submission, with specific focus on the data discussed in the paper and language on the program model, respectively.
- Our research team has been working closely with Evergreen Strategy Group, a communications firm, to develop a communications plan for the WOW paper to disseminate findings to the larger policy and academic communities to inform them

on the promising impacts of programs like WOW. We have shared this plan with the Youth Guidance team to collaboratively plan for the dissemination of these findings, specifically those described in the publication, to community members and other stakeholders in an effective and meaningful way. For example, we have developed a list of objectives that include (1) reinforcing the Education Lab's expertise and leadership in nationally relevant research among policymakers, practitioner, thought leaders, and journalists in order to maximize the impact of our findings; (2) demonstrating the efficacy of WOW; (3) capitalizing on the expansion of WOW to reach other agencies in the mist of directing federal ARP funding; (4) bringing trauma-informed programming/supports centered around young women, particularly Black and Latina women, to the forefront of the conversation around school-based support services and destigmatizing mental health supports by and for Black and Latina women; and (5) bringing the experiences of program participants, educators, and counselors working on these issues to contribute to national conversations on pandemic schooling. In doing so, we have developed topline messaging to address the problem, the need, and the results as they relate to the goals and findings of WOW. We have also developed a list of tactics to disseminate our findings and meet the objectives outlines above.

- As a result of these conversations, Youth Guidance and the Education Lab devised a coordinated communications plan, with Youth Guidance focusing largely on depicting the story of WOW and first-hand accounts of its impact on participants, and the Education Lab working to disseminate the research findings and policy implications. During the reporting period, both teams finalized op-ed articles to achieve these goals. These op-eds will be submitted once the paper is published.
- *What was accomplished under these goals during the full grant period?*

Over the course of the grant period, we accomplished all objectives we set out to achieve with this project, with the exception of Objective 7 (due to the COVID-19 pandemic; full explanation below). We detail these accomplishments by objective below:

Objective 1

In collaboration with Youth Guidance, the Chicago Public Schools (CPS), and partner schools, we successfully used administrative data to identify a sample of 3,751 eligible 9th, 10th, and 11th grade girls at schools in Chicago community areas experiencing high rates of violence. Using data from CPS, our analyst generated a list of all eligible girls from the 10 study schools based on a few limiting criteria:

1. Enrolled in 9th, 10th, or 11th grade during the 2017-2018 academic year
2. 75% attendance or better during the 2016-2017 academic year
3. No severe or profound disability (as indicated by district codes)

Objective 2

By January 2018, we randomized all 3,751 eligible young women into a treatment or control group, blocking by school and grade. Through our power analysis, we determined that 750 individuals receiving WOW services was sufficient to detect effects on our outcomes of interest; accounting for attrition and average program take-up rates, we randomized a total of 1,233 individuals to the treatment group to safely reach the previous figure of 750. All individuals not assigned to the treatment group were assigned to the control group, for a total of 2,518 control youth. In early October 2018, we randomized a second cohort of 1,355 eligible incoming 9th graders to fill program spots that had opened due to a small amount of attrition (e.g., students who have transferred to non-study schools, or moved outside of Chicago). Of these individuals, 362 were assigned to treatment, and the remaining 993 were assigned to the control group, for a total of 5,106 study participants including a treatment group of 1,595 and a control group of 3,511.

Overall, our sample was balanced across a set of observable student characteristics, indicating that randomization was carried out successfully. See Table 1 for the associated p-values on the variables we tested and the results from a Joint-F test.

Objective 3

Throughout the grant period, the research team monitored and measured program participation and fidelity of program implementation. The research team and Youth Guidance held weekly meetings, during which Youth Guidance provided detailed updates on program recruitment and enrollment and provided updates on the progress of WOW groups within the curriculum. This open line of communication allowed us to solve problems collaboratively and identify areas where our respective organizations could offer support to each other over the course of the study.

In addition to these meetings, the research team had planned to conduct systematic observations of WOW sessions to assess program fidelity. However, due to recent and more stringent HIPAA and human subject requirements, our research team was unable to complete this task and instead substituted these observations with a series of 60-90 minute semi-structured interviews of WOW counselors, school leadership, and school staff conducted by a trained qualitative researcher. Through a qualitative analysis, these interviews revealed a set of barriers and facilitators faced by the counselors to implement WOW with fidelity. We found that the barriers to program fidelity seemed to generally fall into two categories:

- **School-level structural barriers:** Lack of buy-in, space, and budget
- **Language barriers:** Many of the participants' families primarily spoke Spanish

Facilitators to program fidelity fell into three categories:

- **Youth Guidance-based facilitators:** Receiving the training, budget, and supervision necessary to succeed in program implementation

- **Counselor-based facilitators:** Relevant work experience, a shared identity, and problem-solving independently or together to tackle issues such as low attendance or language barriers
- **School-based facilitators:** School administrators who took the time to understand and support WOW

Objectives 4 and 5

The research team fully executed three statements of work under our existing master data sharing agreements:

1. A statement of work with CPS allowing us to measure program impacts using student information such as test scores, grades, attendance and enrollment records, and disciplinary infractions.
2. A statement of work with the Chicago Police Department (CPD) allowing us to measure the program's impact on participants' criminal justice involvement by using arrest records.
3. A statement of work with Youth Guidance allowing us to monitor implementation data, such as the take-up rate and retention, and match the WOW roster to the CPS and CPD administrative datasets to measure program impacts.

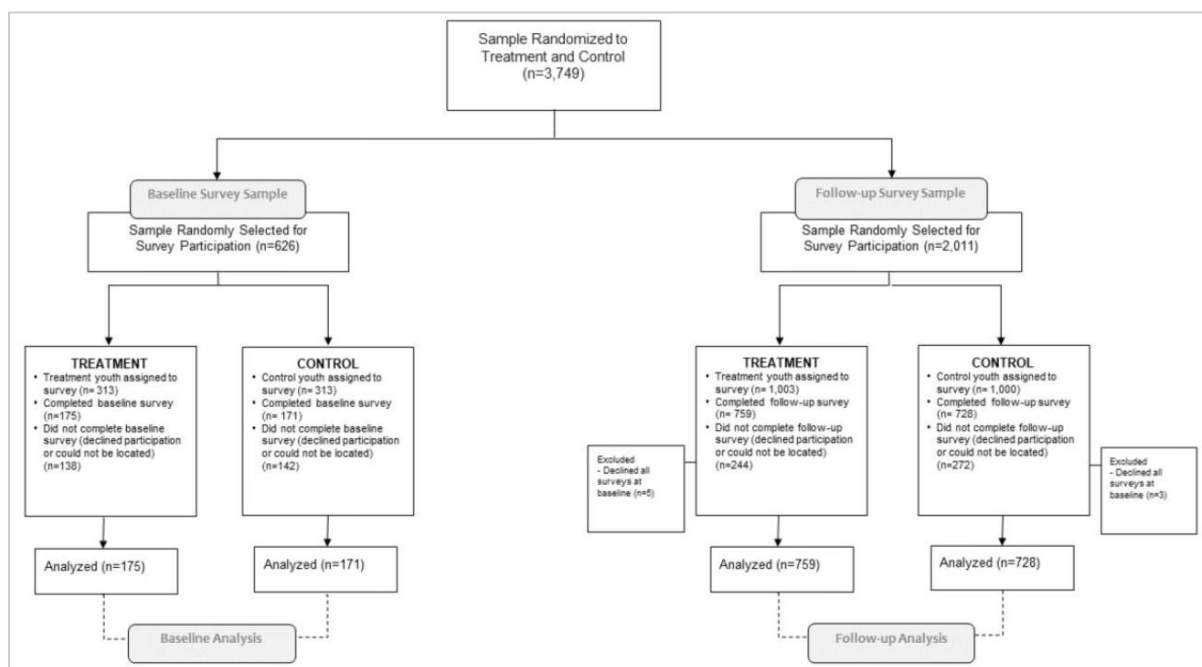
Objective 6

In collaboration with the University of Michigan Institute for Social Research (ISR), the research team successfully collected primary data through surveys designed to assess the prevalence of symptoms of anxiety, depression, and PTSD. Using additional resources raised by the research team, we conducted a baseline survey prior to WOW program delivery in order to (a) establish baseline equivalence among non-administrative data outcomes, (b) obtain baseline rates of key mental health outcomes, and (c) enable our research team to carry out an exploratory pre-/post- analysis on survey data outcomes. From a randomly selected survey sample of 626, we were able to consent and survey 346 young women, for a final response rate of 55.3 percent (Figure 1). The survey subsample that completed surveys was balanced across a set of demographics, school-related outcomes, mental health outcomes, and potential mediators, as demonstrated by Table 2 in the attachment.

Table 2 shows the baseline prevalence of depression, anxiety, and PTSD within this understudied population. Approximately 17% of girls surveyed at baseline reported as at-risk for depression, and 17-18% for anxiety. About 10% of girls surveyed were within the clinically significant range for depression and about 7% were within the clinically significant range for anxiety. Almost 30% of surveyed girls scored within the probable PTSD range on the Child and Adolescent Trauma Screen (CATS) and on average girls reported 2.5 lifetime exposures to traumatic events. These data give us an important sense of what girls in our study population were experiencing prior to program delivery and were instrumental in contextualizing program impacts and potential future program scaling efforts.

We followed our baseline survey with a large-scale survey to assess the impact of one year of WOW programming on key mental health outcomes, potential mediators, and other social-emotional and behavioral outcomes not captured in the administrative data. To deliver this larger survey, we partnered again with ISR for survey administration. The ISR team was able to collect 1,487 responses out of the roughly 2,000 students they intended to reach, making an effective response rate of 84% (Figure 1). Balance on observable student characteristics for the full combined survey sample is demonstrated in Table 3. This increased sample size strengthened our ability to detect program effects on key academic and justice-related outcomes.

Figure 1: Survey population consort diagram



Objective 7

The COVID-19 pandemic unfortunately did not allow for us to work with the Illinois Department of Public Health or the Chicago Department of Public Health CDPH to acquire health data in order to study the reproductive or sexual health impacts of participation in WOW, as the public health agencies paused non-essential research requests. Due to the backlog in these requests, we did not obtain these datasets in the course of this grant period.

Objective 8

Over the span of the grant period, our research team both conducted the statistical analyses required to compare violent behavior and arrest, victimization, schooling outcomes, and

mental/reproductive health outcomes across treatment and control groups and conducted a comprehensive benefit-cost analysis.

Our results show statistically significant reductions in anxiety, depression and PTSD aggregated scores when comparing results for the treatment group and the control group (Table 4). Because these constructs are categorized as *clinical scale measures*, an observed reduction in score means that the treatment group is, on average, better off than the control group. It is notable, however, that we see statistically significant differences around the clinical thresholds for PTSD, but not anxiety or depression.

In the case of the non-primary BASC-3 outcomes, we observed a significant reduction in the aggregated score for Social Stress (Table 5), which can be interpreted as the treatment group being better off than the control group per the reason outlined above, and a significant increase in the Self-Reliance aggregated score (Table 6). Because Self-Reliance is categorized as an *adaptive scale measure*, an observed increase in score means that the treatment group is, on average, better off than the control group.

In our analysis of academic outcomes in the administrative data for the full sample of 3,749 girls, we observed no statistically significant differences between treatment and control participants in overall attendance, GPA, or freshman on-track status. Treatment and control participants displayed virtually identical outcomes on these measures (Table 7).

We also use Chicago Police Department data to understand the effect of WOW on arrests (Table 8) and victimization (Table 9). We see relatively low rates of students in the study sample experiencing contact with the criminal justice system, though these low rates bely high numbers of young girls experiencing arrests. For example, just 2% of the girls in our sample had been arrested, but this represents 100 young high school girls experiencing arrests. We do not see any differences between girls in the treatment and control group either during the program or up to 12 months after the program who have been arrested. We observe similar victimization rates (3% in the control mean) to arrests; however, we see statistically significant increases in the effect of being offered the WOW program (an increase of 2 percentage points) or participating in the WOW program (an increase of 3 percentage points) on victimization. These results are puzzling, but inconclusive as the CPD administrative data is a proxy for a crime occurring. If the WOW program increased a students' perception of self-worth or self-esteem, or taught them about harmful relationships, this could increase a students' likelihood of *reporting* being a victim of a crime without altering their likelihood of being involved as the victim of a crime. Since we do not have measures that allow us to distinguish between students' likelihood of reporting from involvement, we cite these as findings that are worth replicating, but inconclusive in their current form.

Cost-benefit analysis

In addition to the analyses of primary mental health outcomes, student health, and academic/arrest data, the research team also performed a cost-benefit analysis. This analysis allows us to estimate the program's strengths and weaknesses relative to other similar programs or interventions—or no intervention at all. As outlined above, our main results suggest that WOW improved mental health outcomes for program participants relative to controls. We then used cost-utility analysis to examine the economic value of such mental health improvements relative to WOW's associated program costs. We focused on observed treatment effects in reducing depression and PTSD symptoms, converting these mental health benefits to changes in Quality Adjusted Life Year (QALY) units gained through the intervention.¹

Our findings suggest that WOW is highly cost-effective when evaluated solely in terms of its impact on PTSD, or on the combined PTSD/depression outcome. It is borderline cost-effective if evaluated solely as a depression prevention intervention. These results will be included in the WOW paper.

This pattern is especially important for policy and practice because WOW's main program benefits arise from its ability to directly improve student mental health. Within education and policy research, such mental health benefits are frequently overlooked or disregarded as important endpoints in and of themselves. Rather, they are frequently modeled as mediating outcomes or impact measures within interventions whose ultimate metrics for success are to improve grades, reduce arrests for violent offenses, or to move other outcomes.

Effects duration analysis

Our research team also sought to understand the degree to which the evaluation's primary mental health results of anxiety, depression, and PTSD, varied as a function of time. In order to model this, we measured the number of days in between: (i) randomization day; and (ii) the day in which a participant responded to the follow-up survey; the former being constant for everyone enrolled in the same cohort (Sep 18, 2017 for cohort 1), the latter unique for each participant (some day in between May and November 2018).

The variation induced by different follow-up survey dates provided us with a powerful tool to study the degree to which our primary mental health outcomes depend on the duration of/exposure to the WOW program, given that some participants (~500/1,487) were surveyed up to a month before the end of the 2017-2018 CPS calendar year (June 18, 2018), and others were surveyed several months after.

¹ Briefly summarized, QALYs are constructed by multiplying the length of time a program effect is presumed to occur by a utility weight that corresponds to the quality of life of individual. One year in perfect health is equivalent to one QALY as is two years in a very poor health state with the assigned utility weight of 0.5 (1 year * 1 = 1 QALY = 2 years * 0.5).

We studied the effects of program duration with a classic heterogeneous treatment effects approach. After creating the time-to survey indicator, we included an interaction term in between this indicator with the treatment indicator in the main regression analysis for the mental health primary outcomes. The coefficient obtained for the interaction term can be thought of as the marginal contribution of an extra day of programming to the average treatment and control group difference. Results are suggestive of sustained and possibly increasing reductions in terms of anxiety, depression and PTSD, as a function of time since randomization; however, these results are suggestive and exploratory only and should not be considered causal or confirmatory (Table 10).

Objective 9

Throughout the grant period, the research team has worked to ensure that this study's findings reach a wide range of policymakers, practitioners, and interested community members. We met regularly with Youth Guidance and CPS to share preliminary updates and findings, and as mentioned in the previous section, we have recently partnered with the communications firm Evergreen Strategy Group to create a detailed dissemination plan and communications assets to circulate once the study results are public. To reach academic audiences, we submitted our manuscript to various peer-reviewed journals and have been invited to revise and resubmit our draft to *Science Advances*.

- *What opportunities for training and professional development has the project provided?*

This project has provided opportunities for all research team members to be engaged in the collaborative process of a researcher-practitioner partnership, which includes holding regular meetings and check-ins with Youth Guidance to make mutual decisions. Members of our research team have been able to learn in-depth about the WOW program and to learn how to consider questions from the practitioner perspective.

Research team members were also able to continue practicing presentation and communication skills through a series of meetings with Youth Guidance staff and leadership, stakeholders at the Chicago Public Schools, and other key community stakeholders, to share preliminary results and discuss next steps for the research study.

Finally, we invited current and previous members or contributors of our research team to collaborate with us on the writing, editing, and development of the WOW publication. We acknowledge that each current and previous member of our research team has provided meaningful contributions to the study and want to recognize their work by inviting them to be co-authors on the publication.

Several of our research staff have gone on to pursue graduate or Ph.D. programs in the field of social science.

- *How have the results been disseminated to communities of interest?*

We have continued to discuss preliminary baseline mental health data, secondary outcomes data analysis, cost-benefit analysis, and study progress in meetings with leadership from the

Chicago Public Schools—including individuals from the district’s Office of Social and Emotional Learning and school leaders of participating schools—Youth Guidance staff and leadership, funders, and other key community stakeholders such as the WOW counselors, Youth Guidance board and Youth Guidance supporters at their annual gala. We plan to continue these meetings with study school leadership and administrators in the near future. Most recently, in September 2021 and December 2022, our research team organized a meeting with the larger Youth Guidance team to share updates on the WOW paper and communications plan. We discussed the data analyses and language around the WOW program model, specifically its curriculum.

Products/Deliverables:

- *What has the project produced?*

Our research team has continued to submit grant reports with our findings to our funders, and we have also produced a draft of the WOW which we submitted to *Science* and *Science Advances* in February 2022. We have received a request for revisions and resubmission of the paper to *Science Advances* which we will submit in December 2022.

Participants and Other Collaborating Organization (if applicable):

- *What individuals have worked on the project?*

Harold Pollack, Principal Investigator
UChicago Crime Lab and Health Lab
UChicago School of Social Service Administration

Jonathan Guryan, Co-Principal Investigator
UChicago Education Lab
Northwestern University

Monica Bhatt, Senior Research Director
UChicago Crime Lab and Education Lab

Molly Clark, Research Manager
UChicago Crime Lab and Education Lab

Makenzi Sumners, Research Manager
UChicago Crime Lab and Education Lab

Phoebe Lin, Research Manager
UChicago Crime Lab and Education Lab

Juan Carlos Castrejon, Project Associate
UChicago Crime Lab and Education Lab

Cristobal Pinto, Analytics Manager
UChicago Crime Lab and Education Lab

Lucia Delgado Sanchez, Data and Governance Fellow
UChicago Crime Lab and Education Lab

Ben Shaver, Data and Governance Fellow
UChicago Crime Lab and Education Lab

Ashley Cureton Turner, Implementation Manager
UChicago Crime Lab and Education Lab
UChicago School of Social Service Administration

Ashley Allcorn, Research Assistant
UChicago Crime Lab and Education Lab

Dinah Hudson, Research Assistant
UChicago Crime Lab and Education Lab

Olga Nehme, Research Assistant
UChicago Crime Lab and Education Lab

Charlotte Acra, Research Assistant
UChicago Crime Lab and Education Lab

Michelle Ochoa, Portfolio Director
UChicago Education Lab

- *What other organizations have been involved as partners?*
 - *Organization Name:* Youth Guidance
 - *Location of Organization:* 1 N LaSalle Dr #900, Chicago, IL 60602
 - *Partner's Contribution:* Youth Guidance is a co-grantee on this researcher-practitioner joint award and has been working closely with our team to provide a deep understanding of the WOW curriculum. They have collaborated with us on our RCT design and implementation plans, as well as the design and implementation of the mental health survey. Perhaps most significantly, as program provider, Youth Guidance staff have executed programming agreements with all participating schools to ensure they agree to the terms of the research project, hired and trained all WOW counselors involved in the study, coordinated and overseen the recruitment of all treatment group study participants, and are providing all treatment services to youth in this study. Youth Guidance has also collaborated with the Education Lab on the study's communications and dissemination strategy.
 - *Organization Name:* Chicago Public Schools (CPS)
 - *Location of Organization:* 42 W Madison ST, Chicago, IL 60602

Partner's Contribution: CPS has been working closely with our team at the district and schools to provide program implementation, research, and data support. We have active master data sharing agreement with CPS, which we will be able to utilize to assess short- and long-term effects of the program. Specifically, longitudinal CPS records are central to our study. Our team has successfully conducted several RCTs in CPS over the past decade.

- *Organization Name:* Chicago Police Department (CPD)
- *Location of Organization:* 3510 S Michigan Ave, Chicago, IL 60653

Partner's Contribution: We have active master data sharing agreements to access CPD data, which we will be able to utilize to assess short- and long-term effects of the program. Specifically, we will have CPD data on arrests and victimization for all youth (all adult arrests and all juvenile-system felonies and serious misdemeanors), which include arrest date, offense, and circumstances of victimization. These data will let us measure impacts during both the program period and in future years.

- *Organization Name:* University of Michigan institute for Social Research (ISR)
- *Location of Organization:* P.O. Box 1248, 426 Thompson St. Ann Arbor, MI 48106-1248

Partner's Contribution: We worked with ISR on the large-scale survey we conducted to measure key mental health outcomes and other behavioral outcomes at the end of the first year of WOW study programming in 2018. While the Education Lab regularly met with ISR to oversee project management efforts and implement the survey in schools prior to the start of the COVID-19 pandemic, we ended our partnership with ISR in Summer 2020 as we cancelled the second follow-up survey that was set to commence in SY2021-22 as a result of the pandemic and no longer have an intention to conduct a large-scale survey.

- *Have other collaborators or contacts been involved?*

Kathryn Grant, Consultant, DePaul University Department of Psychology:

To assist in the design of our mental health survey and to consult on surveying processes and data analysis, we have brought an experienced mental health researcher to our team, Dr. Kathryn Grant. Dr. Grant is a professor of clinical psychology at DePaul University and has conducted extensive research on the effects of stress among children and adolescents in urban settings, and interventions to protect youth, young women in particular, from the negative effects of stress and trauma.

Impacts or Contributions to the field:

Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project.

Through initial analysis of the baseline surveys, we have been able to produce a descriptive snapshot of a population that is generally difficult to reach, on outcomes that are rarely

studied on a large scale. Additionally, through our extensive large-scale follow-up survey, we have generated preliminary results on a trauma-informed program with a positive effect on mental health outcomes for young women. We have evidence that just four months of programming led to statistically significant improvements in participants' mental health, specifically, depression, anxiety, and PTSD symptoms. After only a few months of WOW programming, we observed statistically significant and meaningful reductions in PTSD symptoms (-21.7%), anxiety (-9.8%), and depression (-14.1%) among young women who participated in the program. In addition to the decrease in PTSD symptom severity scores, we observed a 38% decrease in the number of youth whose scores indicate "moderate trauma-related distress." We also see promising but statistically insignificant impacts on "probable PTSD," with a decrease of about 20% in the number of youth who scored at or above this clinical threshold. These preliminary findings suggest the promise of this relatively light-touch approach. Encouraging early signs also suggest the importance of understanding the full two-year program impact and persistence of these beneficial changes.

As outlined above, our analysis of post-one year programming impact suggests statistically significant decreases in PTSD, depression, and anxiety. In addition to these primary outcomes, we want to understand the effect of WOW on other studies as well. We observe a significant reduction social stress and an increase in self-reliance. Our results also show statistically significant effects for whether program participants want to finish high school, want to graduate from college, and want to attend higher level of schooling after graduating from college. We do not, however, see statistically significant effects of WOW on GPA, % course failure, attendance, PSAT/SAT scores, school persistence index, freshman on-track indicator, school suspensions and school incidents.

Our research team has also produced a cost-utility analysis of the WOW program, based solely on the program's demonstrated impacts to improve mental health. As we learned from our post-one year programming impact analysis, WOW improved mental health outcomes for program participants relative to controls. Building on these findings, our research team used a cost-benefit analysis to examine the economic value of such mental health improvements relative to WOW's associated program costs. We specifically focused on the treatment effects in reducing depression and PTSD symptoms and converted these mental health benefits to changes in Quality Adjust Life Year (QALY) units gained through the intervention. Comparing the cost per QALY gained from WOW to these thresholds, WOW appears to be highly cost-effective when evaluated solely in terms of its impact on PTSD, or on the combined PTSD/depression outcome. It is borderline cost-effective if evaluated solely as a depression prevention intervention.

We believe this intervention has the potential to make outsize impact in the lives of Chicago's young women – and all those they touch, and we are excited to share our findings at a larger scale through the publication of our manuscript detailing our data analyses. We are deeply grateful to the Office of Juvenile Justice and Delinquency Prevention for its generous support of this ambitious project to-date and look forward seeing this partnership into the future. We believe our work on this project will allow us to generate rigorous evidence on

what works to address the impacts of pervasive community violence and help build brighter futures for all of Chicago's young women.

Changes/Problems and Solutions:

- *Changes in approach and reasons for change.*

Previously we had reported a potential difficulty in obtaining Year 2, Cohort 2 data from Chicago Public Schools. We have received that data and it is incorporated into the academic primary outcomes displayed in Table 12 and described earlier in this report under Objective 8.

- *Actual or anticipated problems or delays and actions or plans to resolve them.*

We encountered two challenges at the beginning of the award period which delayed the start of various study activities. First, there were logistical challenges in accepting the award. Due to the way the award was initially set up, we were unable to accept the award in a timely fashion. Working with our program officers, we were able to work out a partial solution and finally were able to draw funds in January 2017. Second, we learned that data from Youth Guidance was considered personal health information, which required additional processes that delayed the exploratory analysis several months.

During the pandemic, our research team and Youth Guidance decided to completely suspend the second follow-up survey as a result of the pandemic, school closures, and the implications of such conditions on the research and survey participants. This decision was communicated with ISR and other relevant survey stakeholders.

Finally, due to the COVID-19 pandemic, we were unable to obtain data from city and state health agencies, as planned, as they had restricted all non-essential research questions not related to the COVID-19 pandemic.

- *Changes that have a significant impact on expenditures.*

In February 2020, our research team submitted a budget modification request to OJP to redirect funds to support the second follow-up survey. The proposed modification redirected \$300,000 from personnel, fringe benefits, and indirect towards contractual/subcontracts. The budget modification GAN was approved in April 2020. Once we decided to completely suspend the second follow-up survey, we submitted a scope of change to our award to OJP to address the impact of this decision on expenditures. The scope of work addressed our project funding and proposed a modification to redirect \$300,000 from contractual/subcontracts to personnel, fringe benefits, and indirect costs. We submitted this in July 2020, and it was approved during this reporting period in September 2020.

We submitted a request for a 12-month no-cost extension (NCE) in June 2021 and have been approved. Given the anticipated delays in administrative data, this extension

provided us with enough time to complete major components of our analysis, and to continue conducting study activities as planned.

- *Significant changes in use or care of animals, human subjects, and/or biohazards.*

N/A

- *Changes to Key Staff:*

Michelle Ochoa, Portfolio Director at the University of Chicago Education Lab, will be responsible for grant closeout.

Next Steps:

- *What do you plan to do following the conclusion of the grant?*

In the future, the research team hopes to publish our paper on our Year 1 findings of WOW's program impact on young women and the WOW program's cost-effectiveness to *Science Advances*. We have been asked to revise our manuscript and resubmit by December 2022.

We will continue working on a communications plan to disseminate our results to a broad policymaker, academic, and practitioner audience to inform them of the promising impact. We will also continue to meet regularly with the Youth Guidance team to discuss program implementation and collaboratively plan for the dissemination of findings, specifically those described in the publication, to community members and other stakeholders in an effective and meaningful way. Once the results are made public, the research team plans to publish a press release, various articles directed to policymaker and practitioner audiences, and amplify the technical research results among academic audiences and social media. Now more than ever, we believe there is a great deal that we can learn from an additional round of data collection, and incredible value in understanding the full impact of the WOW program on the lives of participants.

Table 1: WOW Cohorts 1 & 2 Randomization Baseline Check, All Schools

Variable	Mean T	Mean C	Dif.	p-value	Obs
Old for grade	.147	.12	.017	.105	5106
age by Sep 9th 2017/2018	14.917	14.834	.015	.22	5106
learning disability	.088	.077	.003	.77	5106
free/reduced lunch recipient	.951	.946	.002	.765	5106
baseline Count of out-of-school suspension days	.245	.198	-.008	.792	5102
baseline Count of unexcused absences	5.592	5.452	-.027	.886	5102
baseline Count of excused absences	2.894	3.127	-.059	.647	5102
black	.389	.327	-.003	.764	5106
hispanic	.585	.639	.002	.865	5106
baseline GPA	3.006	3.015	.011	.561	5102
Count of disciplinary incidents	.288	.25	-.01	.764	5106
LEP indicator	.116	.104	.019*	.053	4965
Primary Language Spanish	.5	.536	.012	.291	5106
Joint F Test	Number of obs. =	4965			
	F(13 , 4922) =	.724			
	Prob > F =	.741			

Significance between the mean differences for the treatment vs. control groups was estimated using a linear regression with school & grade fixed effects. Standard errors are robust. A joint F test was performed on all baseline covariates grouped together to test their joint significance. Baseline misconduct, attendance & GPA covariates are SY 1516 for first year randomization, SY 1617 for second year randomization.

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Baseline Survey Balance Results (Updated)

Variable	Treatment mean	Control mean	Adjusted difference	P-value	Obs
CPS DATA					
African American	0.38	0.46	-0.06 ⁺	0.09	346
Hispanic/Latina	0.59	0.52	0.06	0.13	346
Age as of Sep 9th 2017	15.19	15.15	0.02	0.67	346
Learning Disability	0.08	0.06	0.01	0.64	346
Old for grade	0.14	0.14	0.02	0.64	346
# of unexcused absences (AY16-17)	7.84	9.37	-1.3	0.29	346
# of excused absences (AY16-17)	4.24	4.01	0.33	0.54	346
# of out-of-school-suspensions (AY16-17)	0.06	0.05	0.01	0.69	346
# of disciplinary incidents (AY16-17)	0.31	0.24	0.07	0.45	346
GPA (AY1617)	2.98	2.91	0.06	0.46	346
SURVEY DATA					
Depression raw score (BASC-3 SRP-A)	8.87	8.53	0.5	0.53	339
Anxiety raw score (BASC-3 SRP-A)	15.90	15.98	-0.06	0.94	341
PTSD severity raw score (CATS)	12.94	13.77	-0.42	0.78	346
Depression (at-risk)	0.16	0.17	-0.01	0.76	339
Depression (clinically significant)	0.10	0.10	0.00	0.98	339
Anxiety (at-risk)	0.16	0.19	-0.03	0.48	341
Anxiety (clinically significant)	0.05	0.08	-0.03	0.20	341
PTSD (probable)	0.27	0.29	-0.01	0.86	346
# of trauma exposures	2.49	2.47	0.07	0.74	346

Note: Depression and anxiety are coded as dummy variables which equal 1 if a student's score falls within the at-risk or clinically significant range, 0 otherwise. The PTSD variable equals 1 if a student's score is 21 or above (0 otherwise), which indicates that a student is probable for PTSD. These data do not represent diagnoses of depression, anxiety, or PTSD. All differences adjusted for school and grade effects. Unadjusted treatment and control means will differ due to differential proportions of treatment and control groups across schools that display compositional differences in race/ethnicity and other factors.

** p<0.01, * p<0.05, + p<0.1

Table 3: Balance Table – Follow-up Survey Respondents

Variable	Treat	Control	Est	pvalue	Obs
MENTAL HEALTH INDICATORS					
Anxiety Raw Score	15.893	15.969	-.066	.938	329
Anxiety Per Score	59.25	58.565	.603	.839	329
Anxiety At Risk or Clinically Significant	.208	.267	-.064	.165	329
Depression Raw Score	8.754	8.553	.337	.677	326
Depression Per Score	61.395	60.535	1.363	.678	326
Depression At Risk or Clinically Significant	.257	.277	-.016	.745	326
PTSD Raw Score	12.641	13.38	-.372	.805	333
PTSD At Risk or Clinically Significant	.388	.362	.049	.364	333
DEMOGRAPHIC INDICATORS					
Old for grade	.153	.132	.017	.341	1487
Age as of randomization	15.163	15.124	.01	.61	1487
Learning disability	.091	.073	.015	.277	1487
Free/reduced lunch recipient	.947	.949	-.004	.735	1487
Black	.428	.435	-.005	.743	1487
Hispanic	.556	.545	.009	.605	1487
Out-of-school suspensions	.072	.073	0	1	1487
In-of-school suspensions	.116	.113	-.002	.948	1487
Unexcused absences	8.746	9.013	-.353	.495	1487
Excused absences	3.944	4.121	-.335	.196	1487
Days present at school	163.257	163.697	-.16	.855	1487
GPA	2.879	2.875	.015	.683	1487
Disciplinary incidents	.318	.279	.036	.438	1487
Violent disciplinary incidents	.025	.027	-.003	.726	1487
Any disciplinary incidents?	.173	.151	.02	.278	1487
LEP indicator	.108	.087	.017	.238	1487
Primary Language Spanish	.476	.451	.022	.246	1487

Joint F Test I	Number of obs. =	326			
(Mental health & Demographic)	F(25 , 271) =	.745			
	Prob > F =	.809			

Joint F Test II	Number of obs. =	1487			
(Demographic only)	F(17 , 1440) =	.722			
	Prob > F =	.782			

Significance between the mean differences for the treatment vs. control groups was estimated using a linear regression with block level fixed effects. All missing cases were imputed using block level averages. A joint F test was performed on demographic, academic & mental health covariates grouped together to test their joint significance.

P-values estimated using heteroscedasticity-robust standard errors.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 4: Mental Health Primary Outcomes: Anxiety, Depression, and PTSD (including percentile scores when available)

Sample	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
	Anxiety Raw Score					Depression Raw Score					PTSD Raw Score				
Baseline	341	15.98	-.064 (.821)	-.096 (1.178)	16.01	339	8.53	.5 (.789)	.752 (1.132)	8.5	346	13.77	-.416 (1.485)	-.62 (2.116)	14
Year One FU	1537.4	14.83	-.976* (.455)	-1.513* (.692)	15.48	1530.9	8.02	-.794+ (.41)	-1.227* (.625)	8.7	1620.4	12.93	-1.956** (.752)	-3.016** (1.14)	13.91
	Anxiety Per Score					Depression Per Score									
Baseline	341	58.68	.459 (2.902)	.691 (4.164)	58.04	339	60.27	2.299 (3.192)	3.461 (4.583)	59.47					
Year One FU	1537.4	54.89	-3.713* (1.656)	-5.755* (2.52)	57.37	1530.9	58.11	-4.041* (1.697)	-6.249* (2.584)	60.75					
	Anxiety At Risk					Depression At Risk					PTSD At Risk				
Baseline	341	.18	-.029 (.041)	-.044 (.059)	.21	339	.17	-.013 (.042)	-.019 (.06)	.16	346	.09	.053 (.033)	.079+ (.047)	.06
Year One FU	1537.4	.15	.006 (.021)	.01 (.032)	.16	1530.9	.13	-.009 (.019)	-.013 (.028)	.13	1620.4	.13	-.042* (.017)	-.065* (.026)	.17
	Anxiety Clinically Significant					Depression Clinically Significant					PTSD Clinically Significant				
Baseline	341	.08	-.033 (.026)	-.049 (.037)	.11	339	.1	.001 (.034)	.001 (.049)	.13	346	.29	-.009 (.049)	-.013 (.07)	.3
Year One FU	1537.4	.04	-.003 (.011)	-.005 (.017)	.05	1530.9	.09	-.019 (.017)	-.029 (.026)	.12	1620.4	.27	-.035 (.024)	-.055 (.036)	.28
	Anxiety At Risk or Clinically Significant					Depression At Risk or Clinically Significant					PTSD At Risk or Clinically Significant				
Baseline	341	.27	-.062 (.045)	-.094 (.065)	.32	339	.28	-.012 (.049)	-.018 (.071)	.29	346	.37	.044 (.053)	.066 (.076)	.36
Year One FU	1537.4	.2	.003 (.023)	.004 (.035)	.21	1530.9	.22	-.027 (.023)	-.042 (.035)	.25	1620.4	.39	-.078** (.026)	-.12** (.04)	.44

Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property. N(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.

Heteroscedasticity-robust standard errors shown in parentheses.
 *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 5: Clinical Scales

	N	Mean C	ITT	TOT	CCM
Social Stress Raw Score	1419	7.738	-0.674 * (0.297)	-1.031 * (0.455)	8.339
Somatization Raw Score	1442	3.246	-0.112 (0.186)	-0.171 (0.282)	3.415
Sensation Seeking Raw Score	1412	10.11	-0.387 (0.262)	-0.594 (0.402)	10.583
Attention Problems Raw Score	1434	7.621	-0.106 (0.273)	-0.161 (0.418)	8.066
Atypicality Raw Score	1408	4.786	-0.176 (0.253)	-0.271 (0.389)	5.06
Hyperactivity Raw Score	1425	5.481	-0.029 (0.227)	-0.044 (0.350)	5.707
Locus of Control Raw Score	1411	5.135	-0.2 (0.233)	-0.308 (0.360)	5.562
Attitude to School Raw Score	1431	7.765	0.162 (0.237)	0.248 (0.363)	7.853
Attitude to Teachers Raw Score	1404	7.06	0.104 (0.244)	0.159 (0.374)	7.2
Sense of Inadequacy Raw Score	1418	9.91	-0.283 (0.327)	-0.433 (0.499)	10.256

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 6: Adaptive Scales

	N	Mean C	ITT	TOT	CCM
Self Esteem Raw Score	1433	12.709	0.43 + (0.247)	0.658 + (0.379)	12.359
Self Reliance Raw Score	1429	17.62	0.104 (0.224)	0.158 (0.340)	17.602
Interpersonal Skills Raw Score	1419	18.422	0.198 (0.222)	0.303 (0.329)	18.232
Relationship with Parents Raw Score	1405	22.426	0.164 (0.372)	0.252 (0.570)	22.212

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 7: Academic primary outcomes

Sample	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
GPA						Failed any course				Credits Earned			
Cohort I (year 1)	3749	2.61	.003 (.026)	.005 (.042)	2.67	.29	.01 (.015)	.017 (.024)	.23	6.65	.007 (.031)	.011 (.05)	6.78
Cohort II (year 1)	1355	2.59	-.072 (.049)	-.169 (.116)	2.9	.28	.031 (.025)	.073 (.059)	.17	6.44	.025 (.063)	.059 (.147)	6.49
Pooled (year 1)	5104	2.6	-.016 (.023)	-.029 (.041)	2.71	.29	.013 (.013)	.023 (.023)	.23	6.59	.018 (.028)	.033 (.05)	6.73
Cohort I (year 2)	3749	2.62	-.009 (.025)	-.015 (.041)	2.69	.28	.003 (.014)	.006 (.023)	.24	6.64	.025 (.033)	.042 (.054)	6.72
Cohort II (year 2)	1355	2.77	-.016 (.043)	-.037 (.098)	2.82	.25	.001 (.025)	.003 (.057)	.27	6.65	-.005 (.058)	-.012 (.132)	6.67
Pooled (year 2)	5104	2.66	-.014 (.022)	-.025 (.039)	2.71	.28	.003 (.012)	.006 (.022)	.25	6.64	.015 (.029)	.027 (.051)	6.71
Days present						School persistence index				Any school incident ¹			
Cohort I (year 1)	3749	154.68	.892 (.809)	1.469 (1.32)	158.35	0	.015 (.02)	.025 (.032)	.13	.15	.002 (.012)	.004 (.02)	.18
Cohort II (year 1)	1355	155.56	-.556 (1.57)	-1.308 (3.653)	159.63	0	-.02 (.037)	-.046 (.086)	.17	.21	.036 (.025)	.086 (.058)	.17
Pooled (year 1)	5104	154.94	.564 (.73)	1.006 (1.293)	158.79	0	.005 (.017)	.008 (.031)	.13	.17	.008 (.011)	.015 (.02)	.18
Cohort I (year 2)	3749	148.59	-.504 (.949)	-.828 (1.55)	152.08	0	-.018 (.018)	-.029 (.029)	.06	.12	.018 (.012)	.029 (.019)	.14
Cohort II (year 2)	1355	138.55	.6 (1.451)	1.384 (3.311)	137.42	0	.017 (.033)	.039 (.074)	-.01				
Pooled (year 2)	5104	145.75	-.003 (.803)	-.006 (1.417)	149.4	0	-.005 (.016)	-.009 (.028)	.04				
Any L4-6 school incident ¹						In-school suspensions ¹				Out-of-school suspensions ¹			
Cohort I (year 1)	3749	.08	.004 (.01)	.006 (.016)	.09	.08	.008 (.015)	.012 (.025)	.1	.11	.004 (.015)	.007 (.024)	.11
Cohort II (year 1)	1355	.1	.054** (.02)	.127** (.046)	.03	.11	-.003 (.024)	-.008 (.056)	.09	.13	.035 (.027)	.083 (.063)	.07
Pooled (year 1)	5104	.09	.012 (.009)	.021 (.016)	.09	.09	.001 (.013)	.002 (.023)	.11	.11	.007 (.013)	.012 (.023)	.12
Cohort I (year 2)	3749	.06	.004 (.008)	.007 (.013)	.06	.06	.033* (.015)	.054* (.025)	.05	.06	.005 (.01)	.008 (.017)	.06
PSAT/SAT math score (std.) ²						PSAT/SAT read score (std.) ²				Freshman on-track ³			
Cohort I (year 1)	3749	0	-.019 (.029)	-.031 (.047)	0	0	-.044 (.029)	-.072 (.048)	.02	.86	.025 (.018)	.043 (.031)	.88
Cohort II (year 1)	1355	0	-.023 (.05)	-.053 (.117)	.05	0	-.049 (.048)	-.116 (.113)	.07	.830	.016 (.02)	.037 (.046)	.820
Pooled (year 1)	5104	0	-.015 (.025)	-.026 (.044)	.01	0	-.044+ (.025)	-.078+ (.045)	.03	.84	.02 (.013)	.04 (.026)	.86
Cohort I (year 2)	2595	-.03	-.053 (.035)	-.09 (.06)	.04	-.02	-.031 (.034)	-.053 (.058)	.03				

Baseline covariates and randomization block fixed effects included in all models. Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the following arrest categories: violent, drug, and property. Heteroscedasticity-robust standard errors are shown in parentheses.

Year 1(2) outcomes for Cohort I correspond to CPS school year 1718(1819); Year 1(2) outcomes for Cohort II correspond to CPS school year 1819(1920); Year 1(2) outcomes for the pooled case includes both cohort I & II year 1(2) outcomes.

¹Misconduct data from CPS was only available up to SY 1819 at the moment this table was put together.

²PSAT/SAT test are not administered to senior year students - i.e. the sample used for these estimates does not include participants that were enrolled in 12th grade during the testing year. Moreover, we don't have PSAT test data for Cohort II, year 2; the mandatory administration of PSAT/SAT tests was cancelled by CPS in spring 2020 due to the covid-19 pandemic.

³Freshman on-track indicator only applicable for students enrolled in 9th grade. The number of Cohort I participants for which we have a freshman on-track indicator is equivalent to the number of participants enrolled in 9th grade at baseline (N=1,199).

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 8: Arrests after randomization

Sample	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
		Any arrest during program			Number of arrests during program			Number of violent arrests during program					
Cohort I	3749	.02	.004 (.005)	.006 (.008)	.01	.02	.007 (.007)	.011 (.012)	.02	.01	.005 (.005)	.009 (.008)	.01
Cohort II	1355	.02	-.003 (.007)			.02	-.001 (.009)			.01	.001 (.005)		
Pooled	5104	.02	.002 (.004)			.02	.004 (.006)			.01	.005 (.004)		
		Number of drug arrests during program			Number of property arrests during program			Number of other arrests during program					
Cohort I	3749	0	-.001 (.001)	-.001 (.001)	0	0	.005+ (.003)	.008+ (.005)	0	.01	-.003 (.003)	-.005 (.004)	.01
Cohort II	1355	0	-.001 (.001)			0	.001 (.004)			.01	-.002 (.004)		
Pooled	5104	0	-.001 (0)			0	.003 (.002)			.01	-.003 (.002)		
		Any arrest 6 months after			Number of arrests 6 months after			Number of violent arrests 6 months after					
Cohort I	3749	.03	0 (.005)	.001 (.009)	.02	.03	.005 (.009)	.008 (.014)	.03	.02	.002 (.006)	.003 (.009)	.02
Cohort II	1355	.02	.015 (.01)			.02	.016 (.011)			.01	.006 (.007)		
Pooled	5104	.02	.004 (.005)			.03	.008 (.007)			.02	.004 (.005)		
		Number of drug arrests 6 months after			Number of property arrests 6 months after			Number of other arrests 6 months after					
Cohort I	3749	0	-.001 (.001)	-.002+ (.001)	0	.01	.006 (.004)	.009+ (.006)	0	.01	-.001 (.003)	-.002 (.005)	.01
Cohort II	1355	0	-.001 (.001)			0	.007 (.005)			.01	.003 (.006)		
Pooled	5104	0	-.001+ (.001)			.01	.006+ (.003)			.01	-.001 (.003)		
		Any arrest 12 months after			Number of arrests 12 months after			Number of violent arrests 12 months after					
Cohort I	3749	.03	.002 (.006)	.003 (.01)	.03	.05	-.001 (.011)	-.001 (.017)	.05	.02	-.001 (.006)	-.002 (.011)	.03
Cohort II	1355												
Pooled	5104												
		Number of drug arrests 12 months after			Number of property arrests 12 months after			Number of other arrests 12 months after					
Cohort I	3749	0	0 (.001)	-.001 (.002)	0	.01	.003 (.005)	.006 (.007)	.01	.01	-.002 (.004)	-.004 (.007)	.01
Cohort II	1355												
Pooled	5104												

Note: arrests after randomization include events that happened after or during randomization day (Sep 18, 2017, for cohort I; Oct 23, 2018, for cohort II), and up to: (i) the end of the WOW program period (year 1); (ii) 6 months after the end of WOW program; (iii) 12 months after the end of WOW program (cohort I only). Baseline covariates and randomization block fixed effects included in all models. Baseline covariates are: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property.

Heteroscedasticity-robust standard errors shown in parentheses.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 9:Victimizations after randomization

Sample	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
		Any victim during program				Number of victims during program				Number of violent victims during program			
Cohort I	3749	.03	.012+ (.007)	.019+ (.011)	.02	.03	.015+ (.009)	.024+ (.014)	.02	.02	.016* (.008)	.026* (.013)	.02
Cohort II	1355	.01	.004 (.009)			.02	.003 (.009)			.01	.002 (.008)		
Pooled	5104	.03	.01+ (.006)			.03	.013+ (.007)			.02	.013* (.006)		
		Number of drug victims during program				Number of property victims during program				Number of other victims during program			
Cohort I	3749	0	.	.	.	0	-.002 (.001)	-.003 (.002)	0	.01	.001 (.003)	.001 (.005)	.01
Cohort II	1355	0	0 (0)			0	.003 (.003)			0	-.003 (.002)		
Pooled	5104	0	0 (0)			0	-.001 (.001)			0	0 (.002)		
		Any victim 6 months after				Number of victims 6 months after				Number of violent victims 6 months after			
Cohort I	3749	.04	.022** (.008)	.036** (.014)	.03	.05	.032** (.012)	.052** (.02)	.03	.04	.03** (.011)	.05** (.018)	.01
Cohort II	1355	.01	.004 (.009)			.02	.003 (.009)			.01	.002 (.008)		
Pooled	5104	.04	.018** (.007)			.04	.025** (.009)			.03	.024** (.009)		
		Number of drug victims 6 months after				Number of property victims 6 months after				Number of other victims 6 months after			
Cohort I	3749	0	0 (0)	-.001 (.001)	0	0	0 (.002)	-.001 (.004)	0	.01	.002 (.004)	.004 (.006)	.01
Cohort II	1355	0	0 (0)			0	.003 (.003)			0	-.003 (.002)		
Pooled	5104	0	0 (0)			0	0 (.002)			.01	.001 (.003)		
		Any victim 12 months after				Number of victims 12 months after				Number of violent victims 12 months after			
Cohort I	3749	.05	.02* (.009)	.033* (.014)	.04	.06	.031* (.013)	.051* (.02)	.03	.04	.03** (.011)	.049** (.019)	.02
Cohort II	1355												
Pooled	5104												
		Number of drug victims 12 months after				Number of property victims 12 months after				Number of other victims 12 months after			
Cohort I	3749	0	0 (0)	-.001 (.001)	0	0	-.001 (.002)	-.001 (.004)	0	.01	.002 (.004)	.003 (.007)	.01
Cohort II	1355												
Pooled	5104												

Note: victimizations after randomization include events that happened after or during randomization day (Sep 18, 2017, for cohort I; Oct 23, 2018, for cohort II), and up to: (i) the end of the WOW program period (year 1); (ii) 6 months after the end of WOW program; (iii) 12 months after the end of WOW program (cohort I only). Baseline covariates and randomization block fixed effects included in all models. Baseline covariates are: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property.

Heteroscedasticity-robust standard errors shown in parentheses.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 10: Effect duration estimates: Treatment interacted with time-to-survey indicator

Variable	Mean T	Mean C	ITT	CCM	TOT	Obs	Obs (w)
Anxiety Raw Score							
Treat	13.69	14.691	-.859+ (.444)	15.518	-1.361* (.678)	1403	1537.4
Time to Survey x Treat			-.011 (.009)		-.018 (.014)		
Anxiety Per Score							
Treat	50.671	54.355	-3.32* (1.612)	57.723	-5.28* (2.455)	1403	1537.4
Time to Survey x Treat			-.048 (.032)		-.078 (.048)		
Anxiety At Risk or Clinically Significant							
Treat	.194	.194	.009 (.022)	.2	.013 (.034)	1403	1537.4
Time to Survey x Treat			0 (0)		-.001 (.001)		
Depression Raw Score							
Treat	7.217	7.95	-.72+ (.403)	8.682	-1.134+ (.615)	1401	1530.9
Time to Survey x Treat			-.008 (.008)		-.014 (.013)		
Depression Per Score							
Treat	54.044	57.648	-3.842* (1.675)	61.297	-6.074* (2.558)	1401	1530.9
Time to Survey x Treat			-.052 (.034)		-.084+ (.051)		
Depression At Risk or Clinically Significant							
Treat	.201	.224	-.023 (.023)	.246	-.036 (.034)	1401	1530.9
Time to Survey x Treat			-.001 (0)		-.001 (.001)		
PTSD Raw Score							
Treat	10.931	12.867	-2.067** (.736)	14.095	-3.2** (1.122)	1486	1620.4
Time to Survey x Treat			-.008 (.014)		-.015 (.022)		
PTSD At Risk or Clinically Significant							
Treat	.318	.389	-.081** (.025)	.449	-.126** (.039)	1486	1620.4
Time to Survey x Treat			-.001 (.001)		-.001 (.001)		

Heteroscedasticity-robust standard errors shown in parentheses.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

TECHNICAL REPORT

U.S. Department of Justice Office of Justice Programs
Office of Juvenile Justice and Delinquency Prevention

Project/Award Number: 2016-JU-FX-0002

Project Title: Evaluating a Cognitive Behavioral Approach for Improving Life Outcomes of
Underserved Young Women: A Randomized Experiment in Chicago

Principal Investigators:

Harold Pollack

Faculty Co-Director, University of
Chicago Crime Lab and Health Lab
Professor, University of Chicago
School of Social Service
Administration
haroldp@uchicago.edu
773-702-4414

Jonathan Guryan

Faculty Co-Director, University of
Chicago Education Lab
Professor, School of Education and
Social Policy at Northwestern
University
j-guryan@northwestern.edu
773-848-9408

Monica Bhatt

Senior Research Director, University of
Chicago Education Lab
mbhatt@uchicago.edu
773-834-9960

Submission Date: December 29, 2022

Project/Grant Period: October 1, 2016 to September 30, 2022

Current IRB Approval/Renewal and Expiration Date: Approved 2/28/2022; Expires 2/27/2022

Randomized Evaluation of a School-Based, Trauma-Informed Group Intervention for Young Women in Chicago

Monica P. Bhatt, Ph.D.^{1*}, Jonathan Guryan, Ph.D.², Harold A. Pollack, Ph.D.^{3,4}, Juan C. Castrejon^{1†}, Molly Clark^{1†}, Lucia Delgado-Sanchez^{1‡}, Phoebe Lin^{1†}, Max Lubell^{1†}, Cristobal Pinto^{1‡}, Ben Shaver^{1‡}, Makenzi Sumners^{1†}

Affiliations:

¹Crime Lab and Education Lab, University of Chicago; Chicago, Illinois.

*Corresponding Author. Email: mbhatt@uchicago.edu.

²School of Education and Social Policy, Northwestern University; Evanston, Illinois.

³Crown Family School of Social Work, Policy, and Practice, University of Chicago; Chicago, Illinois.

⁴University of Chicago Health Lab, University of Chicago; Chicago, Illinois.

†These authors supported with operationalizing randomization, survey design and administration.

‡These authors supported statistical analysis.

Abstract: This study explores whether a school-based group counseling program for adolescent girls, implemented at scale, can mitigate trauma-related mental health harms. In a randomized trial involving 3,749 Chicago public high school girls, we find participating in the program for four months induces a 22% reduction in PTSD symptoms, and we find significant decreases in anxiety and depression. Results surpass widely-accepted cost-effectiveness thresholds, with estimated cost-utility well below \$150,000 per Quality Adjusted Life Year. We find suggestive evidence that effects persist and may even increase over time. Our results provide the first efficacy trial of such a program specifically designed for girls, conducted in America's third-largest city. These findings suggest the promise of school-based programs to mitigate trauma-related harms.

One-Sentence Summary: A RCT of a school-based group therapy program in Chicago has shown positive effects on three mental health outcomes and was found to be cost-effective.

Main Text: Decades of psychology research establish the developmental mental health harms of trauma exposure, particularly during adolescence. The COVID-19 global pandemic has sharply increased these mental health challenges: the proportion of mental health–related emergency department visits for children aged 5–11 and 12–17 years increased approximately 24% and 31%, respectively, in 2020 compared to prior years (1). While these experiences are not unique to young women, their burdens are often disproportionately borne by girls, particularly Black and Latinx girls. These burdens include effects on mental health, development, and schooling attainment (2). Young women who have been exposed to a traumatic event such as violence are more likely to report symptoms of post-traumatic stress disorder (PTSD) than their male peers (3); depressive disorders are more common among women (4); and women are more likely than men to develop an anxiety disorder over their lifespans (5-7). Exposure to trauma, through its effect on mental health, also influences academic performance and increases the likelihood of high school dropout (8). In Chicago, for example, students lose roughly one week of schooling out a month because of proximity to homicides (9).

Despite the dramatic increase in trauma exposure for adolescent girls and accompanying mental health challenges such as anxiety and depression, there is a dearth of evidence about what works to address these challenges. In a review of over 1,400 studies, the influential Blueprints for Healthy Youth Development identified fewer than ten “model programs” for reducing the harmful effects of trauma exposure. Furthermore, because studies often focus on the general population, we know even less about what works for girls specifically (10). Existing data also suggest that investing in programming for young women may provide benefits beyond the individuals served. For example, research suggests that programs that help males tend to primarily benefit participating individuals, whereas programs that help females also yield benefits for families and communities (11).

In this paper, we provide evidence that it is possible to reduce PTSD, anxiety, and depression among adolescent girls through a school-based, group counseling program informed by cognitive-behavioral therapy (CBT), acceptance and commitment therapy, and narrative therapy. The program we study includes a curriculum that intentionally helps students participating in the program develop self-awareness, build self-esteem and self-efficacy, and enhance their own individual competencies to make positive and healthy decisions.

Students randomized to treatment received programming through Youth Guidance’s Working on Womanhood (WOW) program. WOW is a school-based, trauma-informed, relationship-centered, group counseling and mentoring program designed specifically by and for Black and Latinx women to disrupt the cycle of intergenerational trauma.

WOW’s 39-lesson curriculum is designed around five core themes: *self-awareness*, *emotional intelligence*, *healthy relationships*, *visionary goal setting*, and *leadership*. The program is delivered by masters-level educated social workers and counselors. Each of these staff members works with 50-55 girls across four to five groups in a school. Groups meet once per week, during the school day. For needs that cannot be addressed in a group setting, WOW counselors provide individual counseling and referrals to other services. The

positive network created by WOW is posited to improve peer relationships and social supports, while strengthening girls' ties with important adults, such as parents, teachers and counselors. The program model and curriculum were refined and revised following a formal formative evaluation in 2018 (12).

WOW is rooted in Cognitive Behavioral Therapy (CBT), Acceptance and Commitment Therapy (ACT), and narrative therapy, and key learnings from other school-based mentoring programs. CBT helps individuals recognize and understand connections between thoughts, feelings and behavior, as well as the context within which those thoughts, feelings and behaviors are developed. Prior research has found that CBT is effective in providing symptomatic relief for specific psychiatric disorders, including depression (13-17), PTSD (18), and anxiety disorders (19-21). CBT has also been found to be effective when adapted to a school-based setting for youth exposed to trauma in urban settings (21). There is evidence that school-based CBT can improve school grades (23) and reduce negative behavioral outcomes such as violent crime arrests (24).

Acceptance and Commitment Therapy (ACT) is used primarily to strengthen psychological flexibility, consisting of six primary components: defusion, acceptance, self as context, contact with the present moment, values, and committed action. Psychological flexibility promotes behavior change in the service of chosen values by connecting the present moment with what the situation affords. Narrative Therapy (NT) is a psychotherapeutic approach based on the notion that people construct narratives to define themselves and give meaning to their daily experiences and life events; by supporting narrating their life experiences in richer and more gratifying ways, the individual becomes empowered to consider life alternatives not previously considered. The narrative therapy process has seven core components: collaboration, externalizing the problem, identifying empowerment, identifying values, linking past story to new narrative, inviting witness to new narrative, and documenting gains (25-27). These therapies and treatments have been used to treat depression and anxiety, with no discernible differences documented to date on individual outcomes. The WOW program draws from each therapeutic tradition to form the foundation of the program curriculum (28).

Methods and Analyses

Our research team and Youth Guidance together selected ten study-site Chicago Public Schools (CPS) high schools to be invited to participate in a randomized control trial (RCT) during academic year 2017-18 and 2018-19. Each school had sufficient enrollment to provide the necessary number of treatment and control students, could accommodate WOW, and is located in an under-served community on Chicago's south or west side that has experienced high rates of violent crime (see Fig. 1 below).

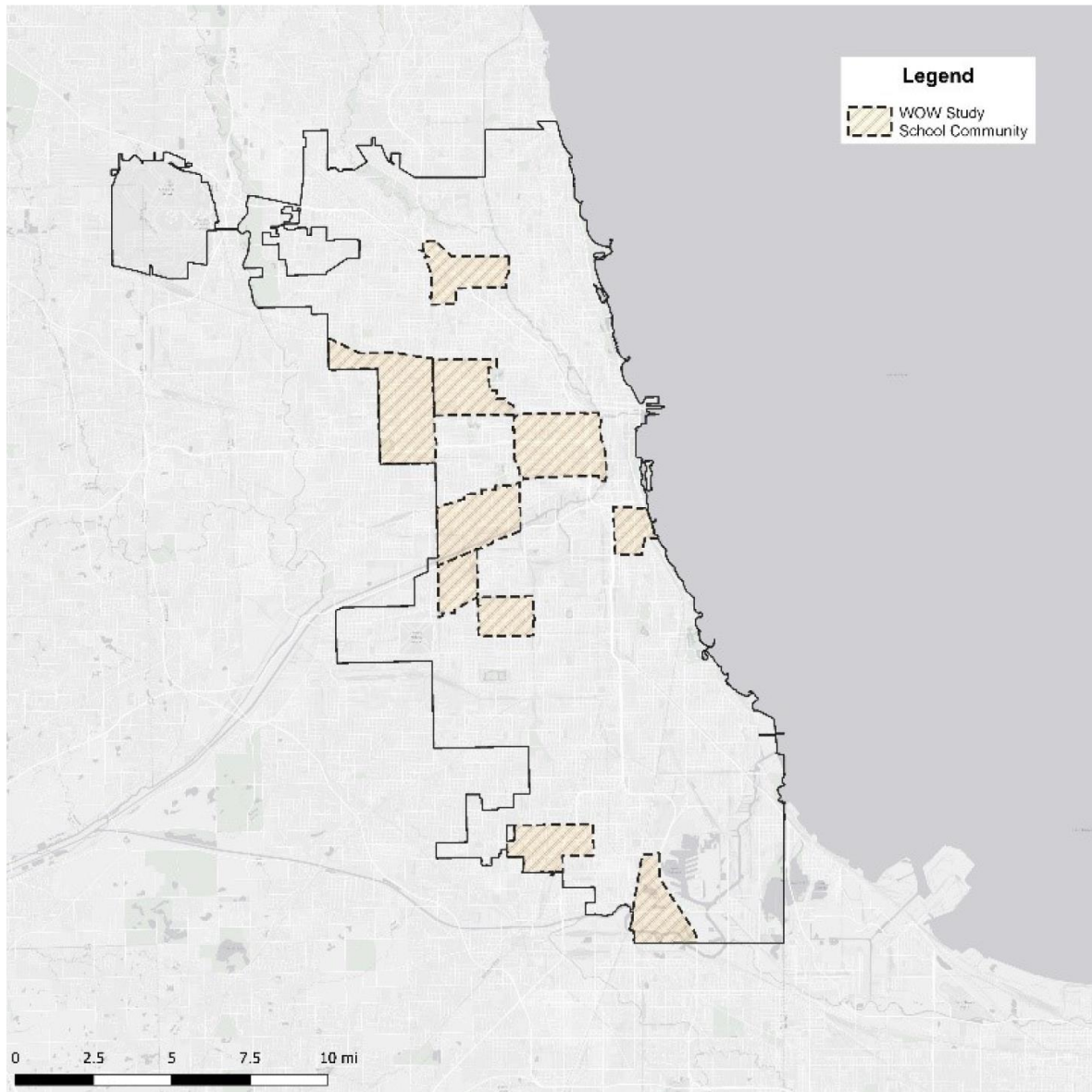


Fig. 1: WOW Study School Communities

During the first week of AY2017-18, we received lists of all 9th – 11th grade girls from schools participating in the study (29). We used CPS administrative data to randomize 3,749 young women who met Youth Guidance’s guidelines for participating in WOW, which included: (1) at least 75% overall school attendance during 2016-17 academic year; (2) no specific diagnosis of significant intellectual disability, per CPS records by the end of the 2016-17 academic year (30); and (3) no signs of proactive aggression, self-harm, or active suicidal ideation. These guidelines were intended to ensure that the eligible students had a reasonable chance of engaging with and benefiting from WOW by attending a sufficient number of sessions, and by being able to productively participate in and understand group activities. Based on these requirements we randomized >95% of all 9th – 11th grade girls in schools participating in the study.

The combined roster lists contained 3,875 potential candidates ready to enter the randomization stage; out of those, 126 individuals (about 3.25%) were excluded from the randomization because of not meeting WOW's requirements for participation.

Data Collection

Survey Measures

In consultation with Youth Guidance and our team's expert mental health researcher (Dr. Kathryn Grant, DePaul University), we compiled a comprehensive survey that contained several rigorous diagnostic instruments. We selected the nationally normed and clinically validated Behavioral Assessment System for Children (BASC-3) (31) to measure anxiety and depression (among several other mediators and secondary outcomes), enabling us to compare our study sample to the national distribution of young women aged 15-18. We also elected to use the well-established Child and Adolescent Trauma Screen (CATS) to measure trauma exposure and PTSD symptomology. We relied on the BASC-3 to gather self-reported data on our hypothesized mechanisms of self-esteem and self-efficacy. The self-efficacy measurement was supplemented using the Self-Efficacy Questionnaire for Children (SEQ-C) drawn from Muris and Bandura et al. (32).

We drew upon a variety of validated instruments and other survey questions to complement these instruments with a range of additional survey items that explored physical and sexual health, risk-taking behavior, self-image, future orientation, relationships, and more. We employed the Difficulties in Emotional Regulation Scale – Short Form (DERS-SF) (33) to measure emotional regulation, and the Future Outlook Inventory (FOI) (34-37) to capture future orientation. Our surveys also included items from the National Longitudinal Study of Adolescent to Adult Health (Add Health) (38), which provided information on reproductive health outcomes and substance use.

Baseline Survey

We administered a pre-program survey in the fall of School Year 2017-18 to obtain mental and physical health outcome measures not well-captured in administrative data. We used this baseline survey data to test for and establish balance between treatment and control groups, improve the precision of our estimates, and provide information on the overall mental health characteristics of our population of interest.

Due to funding constraints, we could not conduct a baseline survey of our entire sample to complement the administrative data to which we had access for the full study sample. We were able to draw a random subset of 626 students from our full sample, blocking on school and grade, for baseline survey. Of the 626, we successfully consented, assented, and surveyed 346 to achieve a response rate of 55.3%. As shown in Tables 1 and 2 below, we achieved good balance in our surveyed sample on demographic and other variables, including our primary outcomes and proposed mechanisms.

Follow-up Survey

Our research team conducted a follow-up survey from May through October 2018 to capture first-year post-program impacts on key outcomes for our first study cohort, which

received services from Fall 2017 through Spring 2019. We worked with the Survey Research Center at the University of Michigan's Institute for Social Research (ISR) to administer the follow-up survey to a random sample of approximately 2,000 students using the survey instruments that we employed for the baseline survey. We employed a two-phase sampling to obtain responses from a representative sub-sample of hard-to-locate respondents, with a goal of an effective response rate of 85%, which ISR had achieved in our prior studies of CPS students (39). We use the follow up survey items in two ways: 1), to assess program impact and 2) to identify potential mediators of treatment effects.

Administrative Data

Master data sharing agreements with CPS and the Chicago Police Department (CPD) provided the administrative data foundation for our statistical analysis of education and arrest outcomes. Because the study sample was drawn from CPS administrative data, we had access to longitudinal student-level records for every student in the treatment and the control groups, including student-level school records on attendance, course grades, test scores, graduation and college enrollment, and school disciplinary actions (40). These data were available prior to the intervention, throughout the intervention period, as well as post-intervention, even for students who changed schools as long as they remained connected to CPS, and were particularly valuable to track youth for in-person surveys, as students could have changed schools, dropped out, or changed addresses during the study period.

Our team also has a master data-sharing agreement with the Chicago Police Department (CPD) to access arrest data (41). These include data on the identity of the offender, date and location of the crime event, and criminal charges (for juvenile as well as adult offenders). We also had access to CPD victimization data covering juvenile and adult arrests citywide. We matched our study sample to these arrest records using probabilistic matching on name and date-of-birth.

Research Design

Intent-to-treat random assignment

We randomly assigned 3,749 9th, 10th, and 11th grade girls to either a treatment group or control group. The consort diagram (see Fig. 2 below) shows the overall structure at each point, depicting both the baseline and follow-up survey's progressions.

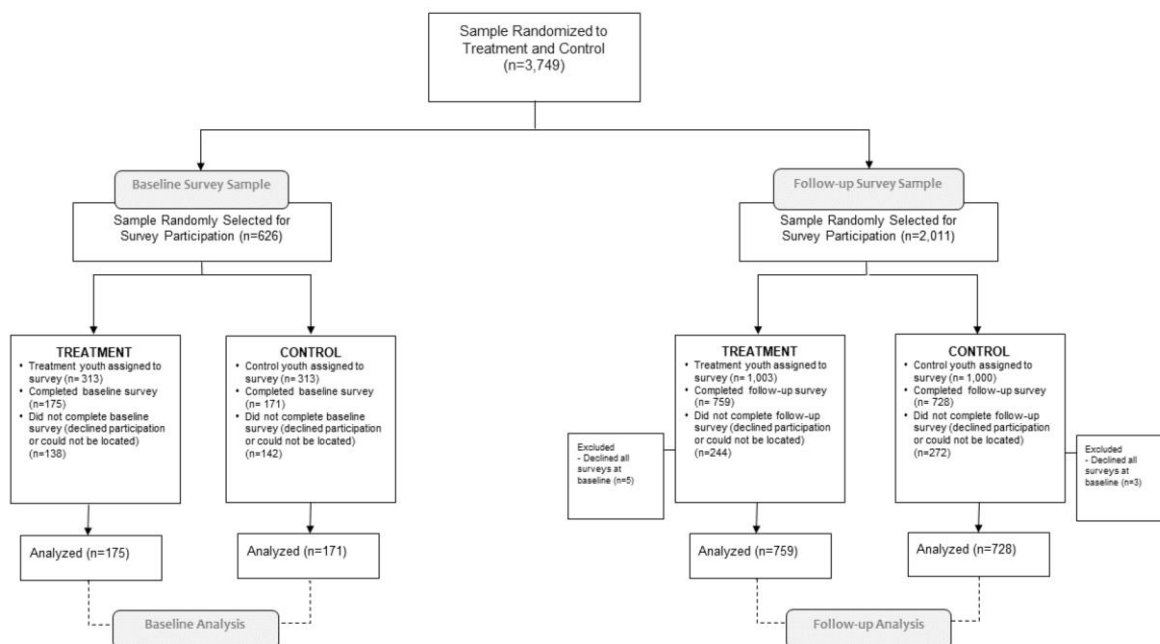


Fig. 2: WOW Study Consort Diagram - Cohort I

Randomization was implemented conditional on school-by-grade randomization blocks, or strata. Within each randomization block, students were randomly selected for either the treatment group, which was offered the chance to participate in WOW, or for the control group, which was not.

The young women in our study are inclusive of almost all 9th-11th grade girls in ten neighborhood high schools across Chicago (95% of enrolled girls are part of the study sample). 1,232 young women were randomized to treatment and 2,517 were assigned to the control group in Fall 2017. 85% of young women in our study are Black and Hispanic/Latinx. About one-half of young women in the study reported speaking Spanish as their primary language at home.

Randomization

After applying the eligibility screens, 3,749 young women remained in the primary randomization pool. Because WOW programming capacity varied by school and sometimes by grade, we carried out a random assignment conditional on school-by-grade “randomization blocks” and varied the probability of assignment to the treatment condition across randomization blocks. 1,232 young women in the study were assigned to the treatment group; 2,517 were assigned to the control group.

Two additional randomization phases followed the one described above. First, we randomly selected a sample of 626 students (out of the 3,749 students in the study) for recruitment for a baseline survey, conducted before the beginning of WOW programming; this sample was created in a way such that the number of students assigned to the treatment group was as

similar as possible to the number assigned to the control group (42). Second, we randomly selected a sample of 2,011 students (out of the 3,749 students in the study) to be recruited for a follow-up survey, conducted towards the end of the first year of WOW programming; this sample was also created in a way such that the number of students in the study assigned to the treatment group was as similar as possible to the number assigned to the control group (43).

For the baseline case, out of the 313 students in the study assigned to the treatment group, 175 (~56%) were effectively surveyed; and out of the 313 assigned to the control group, 171 (~55%) were effectively surveyed. All other students in the study were either unable to be located or declined to participate in the baseline survey. In the case of the follow-up survey, out of the 1,008 students in the study assigned to the treatment group, 759 (~76%) were effectively surveyed (44); and out of the 1,003 assigned to the control group, 728 (~73%) were effectively surveyed (45). Eight students assigned to the follow-up survey sample declined to participate in the study before the follow-up survey began and were excluded from the targeted sample altogether.

Sample

Cohort 1 Sample

In Fall 2017, we randomized a sample of 3,749 eligible youth across the ten schools participating in the study to form Cohort 1, of which 1,232 individuals were assigned to treatment group and 2,517 were assigned to a control group. As seen in Table 1 above, we find that a majority of students in Cohort 1 are free and reduced price lunch (FRL) recipients (a common indicator of low income), and are primarily Black and Hispanic, with about one half of participants recorded as speaking Spanish as their primary language at home. We see a slightly higher proportion of students with limited English proficiency (LEP) in the treatment group, likely driven by significant imbalances within Cohort 1, and a somewhat higher proportion of youth in the control group who are flagged as having a learning disability in Cohort 1.

Follow-up Survey Sample

The first follow-up survey sample included 2,011 study youth randomly sampled for survey participation. Girls selected to be part of the follow-up survey sample were representative of girls in the cohort 1 sample (table 1). Out of the 2,011 youth, we found 8 girls had declined survey participation at baseline and were excluded from the study; therefore, 1,003 treatment youth and 1,000 control youth were assigned to the survey. A total of 1,487 youth took the survey, with 759 treatment youth and 728 control youth completing the survey. Within the 759 follow-up survey respondents in the treatment group, 494 (65.1%) participated in WOW while 265 were non-participants. Across all survey respondents, a majority of youth are FRL recipients, and are primarily Black and Hispanic, with nearly one half of respondents recorded as speaking Spanish as their primary language at home (see Table 2 above). We observed no statistically significant differences in response rates across treatment and control participants. 1,487 participants responded to at least one question of either the SHS or BASC-3 questionnaires. We received 1,484 completed surveys.

Cohort I				Follow-up Survey Sample		
Variable	Treat Mean	Control Mean	p-value	Treat Mean	Control Mean	p-value
Old for grade	.149	.122	.135	.151	.127	.118
Age as of randomization	15.15	15.12	.256	15.16	15.13	.190
Learning disability	.096	.073	.099+	.091	.081	.393
Free/reduced lunch recipient	.952	.951	.929	.946	.954	.402
Black	.393	.328	.705	.402	.407	.871
Hispanic	.582	.64	.859	.574	.570	.944
Out-of-school suspensions	.067	.07	.319	.072	.072	.985
In-school suspensions	.101	.081	.729	.109	.107	.910
Unexcused absences	8.92	9.43	.412	8.96	9.37	.417
Excused absences	4.10	4.47	.260	4.03	4.36	.161
Days present at school	162.7	162.8	.620	162.6	162.9	.631
GPA	2.86	2.81	.235	2.86	2.84	.532
Disciplinary incidents	.287	.25	.898	.298	.278	.567
Violent disciplinary incidents	.025	.027	.312	.027	.025	.805
Any disciplinary incidents?	.157	.14	.861	.162	.156	.666
LEP indicator	.11	.099	.063+	.102	.097	.703
Primary Language Spanish	.497	.538	.397	.491	.474	.403
Obs.	3,749			2,003		
Joint F Tests	F(17 , 3702) = 1.044 Prob > F = .406			F(17 , 1956) = .732 Prob > F = .772		

Significance between the mean differences for the treatment vs. control groups was estimated using a linear regression with block level fixed effects. All missing cases were imputed using block level averages. A joint F test was performed on demographic/academic covariates grouped together to test their joint significance.

P-values estimated using heteroscedasticity-robust standard errors.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 1: Balance Table - Cohort I & Follow-up Survey Sample

Balance and randomization

2,003 students in the study were offered the opportunity to complete a follow-up survey between June and November 2018. We received 1,484 completed surveys, an effective

response rate of 84%. Non-responders in the follow-up were more likely to be Hispanic, averaged an additional day of unexcused absence, and had slightly lower GPAs than respondents did. We found strong balance, with no statistically significant differences in response rates across students assigned to treatment and control (table 2).

Variable	Treat	Control	Est	pvalue	Obs	Obs (w)
MENTAL HEALTH INDICATORS						
Anxiety Raw Score	15.893	15.969	-.066	.938	329	329
Anxiety Per Score	59.25	58.565	.603	.839	329	329
Anxiety At Risk or Clinically Significant	.208	.267	-.064	.165	329	329
Depression Raw Score	8.754	8.553	.337	.677	326	326
Depression Per Score	61.395	60.535	1.363	.678	326	326
Depression At Risk or Clinically Significant	.257	.277	-.016	.745	326	326
PTSD Raw Score	12.641	13.38	-.372	.805	333	333
PTSD At Risk or Clinically Significant	.388	.362	.049	.364	333	333
DEMOGRAPHIC INDICATORS						
Old for grade	.15	.133	.011	.541	1487	1621.4
Age as of randomization	15.179	15.112	.008	.687	1487	1621.4
Learning disability	.093	.066	.024+	.085	1487	1621.4
Free/reduced lunch recipient	.945	.945	-.002	.87	1487	1621.4
Black	.435	.422	.001	.944	1487	1621.4
Hispanic	.55	.561	.001	.941	1487	1621.4
Out-of-school suspensions	.07	.071	-.004	.84	1487	1621.4
In-of-school suspensions	.119	.116	-.007	.824	1487	1621.4
Unexcused absences	8.589	8.925	-.422	.425	1487	1621.4
Excused absences	3.935	4.108	-.324	.217	1487	1621.4
Days present at school	163.327	163.874	-.146	.867	1487	1621.4
GPA	2.883	2.883	.02	.587	1487	1621.4
Disciplinary incidents	.309	.272	.025	.614	1487	1621.4
violent disciplinary incidents	.023	.028	-.006	.481	1487	1621.4
Any disciplinary incidents?	.171	.14	.027	.143	1487	1621.4
LEP indicator	.106	.093	.012	.463	1487	1621.4
Primary Language Spanish	.473	.469	.012	.564	1487	1621.4
Joint F Test I (Mental health & Demographic)	Number of obs. = F(25 , 271) = Prob > F =	326 1.03 .428	(Obs (w) = 326)			
Joint F Test II (Demographic only)	Number of obs. = F(17 , 1440) = Prob > F =	1487 .918 .553	(Obs (w) = 1621.4)			

Significance between the mean differences for the treatment vs. control groups was estimated using a linear regression with block level fixed effects. All missing cases were imputed using block level averages. A joint F test was performed on demographic, academic & mental health covariates grouped together to test their joint significance. Obs(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.

P-values estimated using heteroscedasticity-robust standard errors.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 2: Balance Table - Follow-up Survey Respondents

Data Analysis

Because of our randomized experimental design, our analysis plan is straightforward. We estimate both the intent-to-treat (ITT) effect and the effect of treatment-on-the-treated (TOT). The ITT estimate comes from estimating equation (1):

$$(1) Y_i = \pi_0 + \pi_1 Z_i + X_i \pi_2 + B_i + \varepsilon_i$$

Where Y_i is an outcome for participant i measured after random assignment, Z_i is an indicator for having been offered WOW programming, B_i is a full set of randomization block fixed effects, ε_i is a random error term, and X_i is a set of baseline controls to improve precision (46).

To estimate the treatment-on-the-treated (TOT) effect we use random assignment (Z_i) as an instrumental variable (IV) for participation (D_i), as in equations (2) and (3) (47, 48). The first-stage equation is:

$$(2) D_i = \gamma_0 + \gamma_1 Z_i + X_i \gamma_2 + B_i + \mu_i$$

Where D_i is an indicator for having participated in WOW programming (defined as having participated in at least one WOW group session), the γ 's are parameters to be estimated, μ_i is a random error term, and all other variables are defined as above. The relationship of interest is:

$$(3) Y_i = \beta_0 + \beta_1 D_i + X_i \beta_2 + B_i + \vartheta_i$$

The identifying assumption here is that treatment assignment has no effect on the outcomes of those assigned to treatment who do not participate. The IV estimate for the parameter β_1 equation (3) is essentially a ratio of two ITT estimates—the ITT effect on the outcome of interest in the numerator and the ITT effect on participation in the denominator.

Results

Baseline descriptive statistics

We used the baseline survey to test for and establish balance between treatment and control groups, improve the precision of our estimates, and describe the overall mental health characteristics of our population of interest. As shown in Table 1 above, we achieved strong balance between treatment and control subjects across schools participating in the study. We observed no statistically significant differences in our thirteen baseline measures.

At baseline, young women in the study had suffered on average at least two serious traumatic experiences in their lifetimes; nearly 30% had personally witnessed someone being attacked, stabbed, shot at, hurt badly, or killed. More than 45% had someone close die suddenly or violently. Twenty-four percent scored within the clinical range or the at-risk range for anxiety; twenty-seven percent scored within the clinical range or the at-risk range for depression (49).

The baseline survey also helps us understand the baseline prevalence of depression, anxiety, and PTSD within this population (N=346). Approximately 17% of girls surveyed at baseline appeared at-risk for depression and anxiety. About 10% appeared within the clinically significant range for depression and about 7% for anxiety (50). Our baseline surveys indicate that 38 percent of 10th and 11th grade young women in our study sample exhibit signs of PTSD; the observed prevalence of probable PTSD among these young women is twice that of service members returning from Iraq and Afghanistan (51). Almost

30% scored within the probable PTSD range on the CATS, and on average girls reported 2.5 exposures to traumatic events.

Follow-up survey outcomes

We observed large and statistically significant treatment-control differences in PTSD, depression, and anxiety scores (see Fig. 3-5 below).

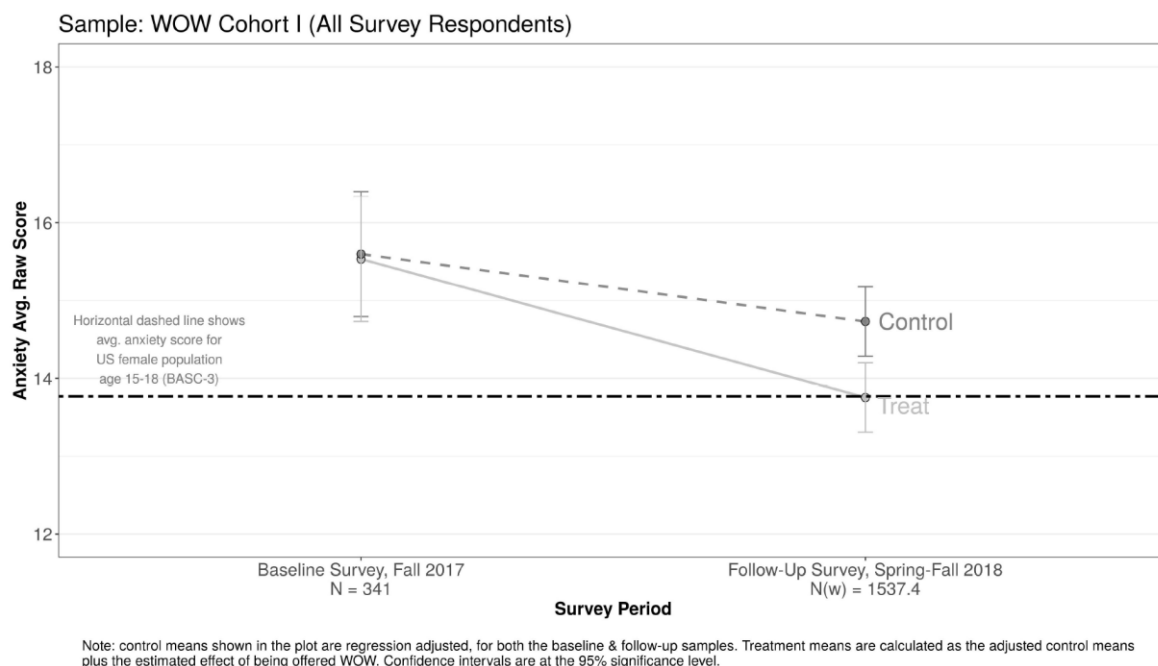


Fig. 3: Mental Health Trajectory for Anxiety (Effect of Being Offered WOW)

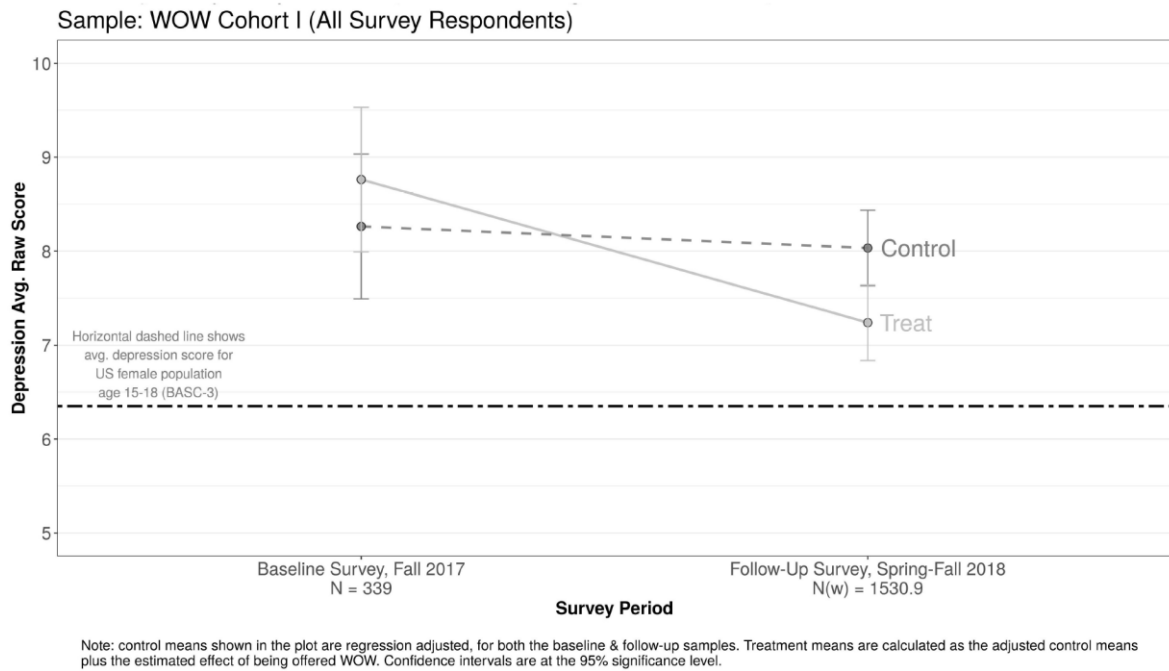


Fig. 4: Mental Health Trajectory for Depression (Effect of Being Offered WOW)

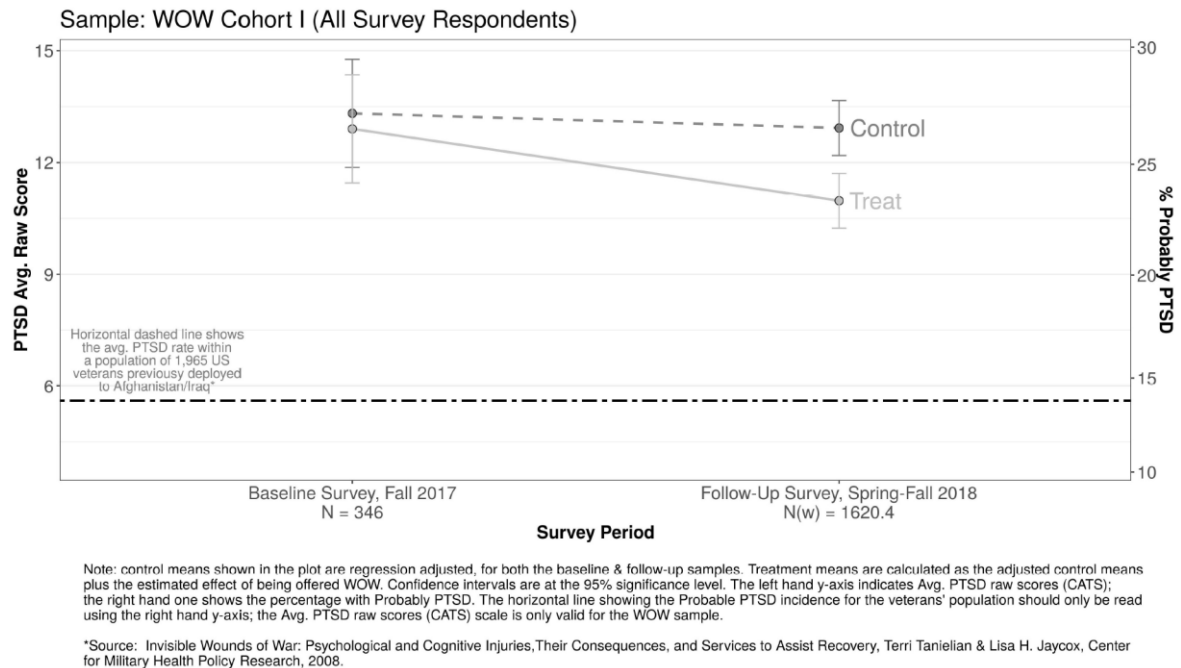


Fig. 5: Mental Health Trajectory for PTSD (Effect of Being Offered WOW)

We find that participation in WOW causes a 22% decrease in PTSD symptom severity scores, which measures the frequency and intensity of PTSD symptoms, and a 38%

decrease in scores that indicate “moderate trauma-related distress.” We also find that WOW participation causes significant decreases in measures of anxiety (9.77%) and depression (14.1%) (see Table 3 below).

Sample	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
Anxiety Raw Score						Depression Raw Score					PTSD Raw Score				
Baseline	341	15.98	-.064 (.821)	-.096 (1.178)	16.01	339	8.53	.5 (.789)	.752 (1.132)	8.5	346	13.77	-.416 (1.485)	-.62 (2.116)	14
Year One FU	1537.4	14.83	-.976* (.455)	-1.513* (.692)	15.48	1530.9	8.02	-.794+ (.41)	-1.227* (.625)	8.7	1620.4	12.93	-1.956** (.752)	-3.016** (1.14)	13.91
Anxiety At Risk						Depression At Risk					PTSD At Risk				
Baseline	341	.18	-.029 (.041)	-.044 (.059)	.21	339	.17	-.013 (.042)	-.019 (.06)	.16	346	.09	.053 (.033)	.079+ (.047)	.06
Year One FU	1537.4	.15	.006 (.021)	.01 (.032)	.16	1530.9	.13	-.009 (.019)	-.013 (.028)	.13	1620.4	.13	-.042* (.017)	-.065* (.026)	.17
Anxiety Clinically Significant						Depression Clinically Significant					PTSD Clinically Significant				
Baseline	341	.08	-.033 (.026)	-.049 (.037)	.11	339	.1	.001 (.034)	.001 (.049)	.13	346	.29	-.009 (.049)	-.013 (.07)	.3
Year One FU	1537.4	.04	-.003 (.011)	-.005 (.017)	.05	1530.9	.09	-.019 (.017)	-.029 (.026)	.12	1620.4	.27	-.035 (.024)	-.055 (.036)	.28
Anxiety At Risk or Clinically Significant						Depression At Risk or Clinically Significant					PTSD At Risk or Clinically Significant				
Baseline	341	.27	-.062 (.045)	-.094 (.065)	.32	339	.28	-.012 (.049)	-.018 (.071)	.29	346	.37	.044 (.053)	.066 (.076)	.36
Year One FU	1537.4	.2	.003 (.023)	.004 (.035)	.21	1530.9	.22	-.027 (.023)	-.042 (.035)	.25	1620.4	.39	-.078** (.026)	-.12** (.04)	.44

Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property. N(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.

Heteroscedasticity-robust standard errors shown in parentheses.
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 3: Mental Health Primary Outcomes: Anxiety, Depression, and PTSD

We analyze academic outcomes in the administrative data for the full sample of 3,749 girls randomized to treatment and control conditions. We observed no clinically or statistically significant differences between treatment and control subjects in overall attendance, GPA, or freshman on-track status. Treatment and control subjects displayed virtually identical outcomes on these measures (see Table 4 below).

Sample	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
		GPA				Failed any course				Credits Earned			
Cohort I (year 1)	3,749	2.61	.003 (.026)	.005 (.042)	2.67	.29	.01 (.015)	.017 (.024)	.23	6.65	.007 (.031)	.011 (.05)	6.78
Cohort I (year 2)	3,749	2.62	-.009 (.025)	-.015 (.041)	2.69	.28	.003 (.014)	.006 (.023)	.24	6.64	.025 (.033)	.042 (.054)	6.72
		Days present				School persistence index				Any school incident¹			
Cohort I (year 1)	3,749	154.68	.892 (.809)	1.469 (1.32)	158.35	0	.015 (.02)	.025 (.032)	.13	.15	.002 (.012)	.004 (.02)	.18
Cohort I (year 2)	3,749	148.59	-.504 (.949)	-.828 (1.55)	152.08	0	-.018 (.018)	-.029 (.029)	.06	.12	.018 (.012)	.029 (.019)	.14
		Any 1-4-6 school incident¹				In-school suspensions¹				Out-of-school suspensions¹			
Cohort I (year 1)	3,749	.08	.004 (.01)	.006 (.016)	.09	.08	.008 (.015)	.012 (.025)	.1	.11	.004 (.015)	.007 (.024)	.11
Cohort I (year 2)	3,749	.06	.004 (.008)	.007 (.013)	.06	.06	.033* (.015)	.054* (.025)	.05	.06	.005 (.01)	.008 (.017)	.06
		PSAT/SAT math score (std.)²				PSAT/SAT read score (std.)²				Freshman on-track³			
Cohort I (year 1)	3,749	0	-.019 (.029)	-.031 (.047)	0	0	-.044 (.029)	-.072 (.048)	.02	.86	.025 (.018)	.043 (.031)	.88
Cohort I (year 2)	2,595	-.03	-.053 (.035)	-.09 (.06)	.04	-.02	-.031 (.034)	-.053 (.058)	.03				

Baseline covariates and randomization block fixed effects included in all models. Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the following arrest categories: violent, drug, and property.

Year 1(2) outcomes for Cohort I correspond to CPS school year 1718(1819).

¹Misconduct data from CPS was only available up to SY 1819 at the moment this table was put together.

²PSAT/SAT test are not administered to senior year students - i.e., the sample used for these estimates does not include participants that were enrolled in 12th grade during the testing year.

³Freshman on-track indicator only applicable for students enrolled in 9th grade. The number of Cohort I participants for which we have a freshman on-track indicator is equivalent to the number of participants enrolled in 9th grade at baseline (N=1,199).

Heteroscedasticity-robust standard errors shown in parentheses.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 4: Academic Primary Outcomes - Cohort I

Economic policy analysis (cost-utility results)

Our main results suggest that WOW improved mental health outcomes for young women participating in the program relative to those who did not. Building on these findings, we used cost-utility analysis to examine the economic value of such mental health improvements relative to WOW's associated program costs. We focused on observed treatment effects in reducing depression and PTSD symptoms, converting these mental health benefits to changes in Quality Adjusted Life Year (QALY) units gained through the intervention (52).

To generate an estimate of the cost per QALY gained through WOW, we applied utility weights for depression and PTSD, drawn from existing research, to the responses in the follow up survey administered to the young women in the study. For depression, we assumed that an individual who is depression-free has a utility weight of 1 (perfect health) and that an individual who satisfies screening criteria for clinically significant depression has a utility weight of 0.59 (53). We also assumed that an individual at risk for clinically significant depression has a utility weight between 0.59 and 1 proportionate to their BASC-3 depression t score. The assignment of utility weights for PTSD followed a similar pattern. We assumed an individual who is PTSD-free has a utility weight of 1 and that an individual with clinically significant PTSD has a utility weight of 0.61 (54). Those at-risk for PTSD receive a utility weight between 0.61 and 1 proportionate to their survey CATS PTSD score.

The program cost was estimated to be \$2,300 per actual participant in WOW. This number was obtained from Youth Guidance, the WOW program provider. That cost figure is based on a caseload of 55 youths served per Counselor. It includes the direct salary and benefit costs of Counselors and the Curriculum Specialists who provide training and coaching. It

also includes further expensing of program management, other direct programming expenses, and overhead costs for information technology, human resources, and Youth Guidance financial and program evaluation/quality assurance required to operate the program.

We performed an intent-to-treat (ITT) analysis to compare mental health outcomes among all youth invited to participate in WOW with those observed in the control group. Based on the program take-up rate of 65%, we converted our program cost data to be comparable to the ITT measure. The cost per program participant in the ITT treatment-assigned group was thus $\$2300 \times 0.6525$, or \$1,501.

Following the medical cost-effectiveness literature, we then computed the cost per QALY gained by computing an incremental cost effectiveness ratio (ICER) for each outcome:

$$\left(\frac{Cost_T - Cost_C}{QALY_T - QALY_C} \right) = \left(\frac{\$1,501 - \$0}{QALY_T - QALY_C} \right)$$

Where $Cost_T$ refers to the average cost of implementing WOW calculated within the full treatment-assigned group (55) (e.g. $\$2300 \times 0.6525 = \sim \1500 for the PTSD case). We take an estimate of β_1 from the following regression model as an estimate of $QALY_T - QALY_C$:

$$QALY_i = \beta_0 + \beta_1 WOW_1 + BX + BXmiss + \mu_i + \varepsilon_i$$

where:

- $QALY_i$: is an individual's utility weight multiplied by an assumed effect duration of 365 days. (We perform sensitivity analyses of assumed effect duration below.)
- WOW_i : is a WOW treatment binary indicator.
- X : is a vector of baseline covariates.
- $Xmiss$: a dummy indicator identifying those cases where the baseline covariates included in $\$X\$$ were imputed, using randomization block-level means,
- μ_i : is block level fixed effects,
- ε_i : normally distributed random error term.

We compute an incremental cost-effectiveness ratio (ICER) for depression and PTSD separately, where the cost per QALY refers to the cost per QALY gained in terms of just that outcome.

We also compute the cost per QALY gained for the combination of PTSD and depression. To avoid complex assumptions about the cumulative impact of co-occurring conditions, we assigned each student in the study a utility weight that corresponds to the minimum utility weight they scored across depression and PTSD (56). Table 5 reports the cost per QALY gained for each outcome.

Outcome	Cost per QALY Gained presuming 365-day benefit	Days of posited benefit required to meet \$150,000/QALY threshold	Days of posited benefit required to meet \$100,000/QALY threshold
Depression as the only valued outcome	\$180,522/QALY	439	659
PTSD as the only valued outcome	\$64,274/QALY	156	235
Combined depression and PTSD	\$58,025/QALY	141	212

Showing cost per QALY gained assuming a 365-day benefit. We also show the number of posited benefit days required to meet relevant cost-benefit thresholds.

Table 5: Cost per QALY Gained

The Institute for Clinical and Economic Research reports two standard thresholds to use for economic evaluations of pharmaceutical pricing, as well as medical and public health interventions in the U.S.: \$150,000/QALY and \$100,000/QALY (57). Comparing the cost per QALY gained from WOW to these thresholds, WOW appears to be highly cost-effective when evaluated solely in terms of its impact on PTSD, or on the combined PTSD/depression outcome. It is borderline cost-effective when evaluated solely as a depression prevention intervention.

To assess the sensitivity of these results to variance in the estimates of the program effects, we performed a bootstrap analysis to examine the variation in our computed ICER for the combined outcome using 100,000 repetitions. We find that WOW meets the \$150,000 threshold in 92.4% of samples and the \$100,000 threshold in 81.2 % of samples. Figure 8 is the cumulative distribution function of the bootstrapping results.

As a further robustness check, we conducted a bounding exercise where we generate utility weights for a third mental health outcome, anxiety, using the utility weights for depression and students' responses to anxiety questions of the BASC-3. This does not change our fundamental result. When we combine utility weights using the minimum quality-of-life weights for an individual across their PTSD, depression, and imputed anxiety scores, and construct an ICER, we find the cost per QALY gained to be \$67,505. Assuming the quality of life for anxiety is no better than the quality of life for depression, this ICER provides an estimate of the cost per QALY gained across our three primary outcomes.

A second uncertainty (see Table 5 above) relates to the posited duration of program effects. If the true program effect were shorter than 365 days, we may overstate program cost-utility. Examining the combined depression/PTSD outcome, we find that the intervention meets the \$100,000 benchmark with a posited intervention duration of 235 days of benefit. The intervention meets the \$150,000 benchmark with a posited benefit duration of 156 days.

Study limitations

Our analysis has several study limitations that must be considered in evaluating our results.

Within-school randomization creates some potential for spillovers, which could have led us to understate program effects. Spillovers occur if control students directly enroll in WOW, if WOW staff provide informal mentoring or other supports to control-group students, or if treatment-group services indirectly benefit control-group students in other ways.

Our mental health outcomes were based on the BASC-3 instrument rather than a diagnostic clinical interview. Our results thus should not be taken as providing clinical diagnosis for any specific individual. Because we could not conduct diagnostic clinical interviews, we adopted a conservative approach to the quality-of-life impact of co-occurring depression and PTSD symptoms, in which we assumed quality of life was equivalent to the minimum value obtained for each condition independent of co-morbidity. To the extent that the quality-of-life impact of co-occurring symptoms is worse than the impact of either condition alone, we may understate the harms associated with co-occurring conditions.

Given study exigencies arising from the COVID-19 pandemic, we could not gather data on long-term mental health outcomes among WOW participants or controls. Our findings are thus based on respondents' mental health status as captured in the follow-up survey. We were therefore not able to measure whether WOW induced benefits that lasted beyond the study period. We also could not observe intervention benefits that may have faded out for some participants before the follow-up survey was administered. WOW was also designed to be a two-year program. We may understate the program's full benefits given the timing of our mental health surveys.

Given these limitations, we conducted sensitivity analyses to examine the duration of program benefit required for WOW to pass the standard cost-effectiveness threshold. As noted above, WOW's observed mental health benefits associated with reduced PTSD and depression symptoms would need to last approximately 157 days (roughly five months) for WOW to be deemed cost-effective, based on the \$150,000-per-QALY threshold.

Conclusion

Young women, particularly those attending school in low-income, predominantly-minority communities experience high rates of depression, post-traumatic stress disorder, and other mental health challenges. Designing and fielding feasible and cost-effective interventions to address these challenges remains a key challenge.

The Working on Womanhood (WOW) intervention induces marked and statistically significant improvement in depression, anxiety, and PTSD symptoms among the young women who participate in this intervention. WOW was not designed to move academic outcomes such as standardized test scores and grades. And over the period observed, WOW did not improve these outcomes.

Evaluated on the basis of its ability to ameliorate mental health symptoms, WOW appears highly cost-effective when judged on the basis of standard cost-utility metrics used to evaluate medical and public health interventions. These benefits were achieved within the challenging real-world environment of ten Chicago public high schools. At a per-participant cost of \$2,300, WOW provides one promising model that can be replicated at scale within resource-challenged public schools across the country.

Discussion

The burden of mental health is often unseen and overlooked, particularly for young women of color, in part because it is not always associated with externalizing behaviors. This study marks one of the first studies of adolescent mental health for young women of color that documents prevalence of mental health challenges such as anxiety, depression, and PTSD in Chicago. It also provides rigorous evidence about how to systematically reduce the prevalence of these challenges at scale through an innovative, group-based, in-school model of therapy.

These findings are novel in part because of the limited evidence base for existing interventions and programs intended to address the high rates of trauma exposure and consequent anxiety, depression and PTSD found among young women in our study. Our results suggest that group-based, in-school therapy programs like WOW can be effective in reducing these rates, and highlight the dearth of alternate services available to the young women in our study. More attention should be directed to WOW and other feasible, cost-effective interventions that support the mental health of young women.

This study was conducted in ten schools within the real-world environment in America's third-largest city and in its third-largest school district. The study included the vast majority of young women in these schools; approximately 95 percent of female students were randomized to either the treatment or control conditions.

Given that approximately 70% of adolescents' mental health services are received in a school-based setting, this study marks an important advancement in quantifying the efficacy of the WOW approach of combining CBT, ACT and narrative therapy into group-based counseling. Despite this advancement, there were young women who had more profound mental health challenges, and more significant needs, that were not appropriate for group-based intervention; future work should examine interventions to support the needs of these girls.

In our analysis of costs and benefits, we applied the best cost-utility metrics available. More analyses are needed to specifically study young women, and to provide cost- and outcome measures tailored to this population. Young women experience high mental health disease burdens but are understudied in the literature.

It is also important to evaluate school-based, mental health interventions by the same yardsticks we use to evaluate other interventions to extend life and improve health. If as a society we are willing to cover the cost of medication to support mental health through

Medicaid and other payment sources, we should also finance psycho-social interventions that prove cost-effective when viewed through the lens of these same cost-utility metrics.

References and Notes:

1. R. T. Leeb, R. H. Bitsko, L. Radhakrishnan, P. Martinez, R. Njai, K. M Holland. "Mental Health–Related Emergency Department Visits Among Children Aged <18 Years During the COVID-19 Pandemic — United States, January 1–October 17, 2020." *MMWR Morb. Mortal Wkly. Rep.* **69** (45), 1675-1680 (2020).
2. J. A. Fairbank, F. W. Putnam, W. W. Harris, "The prevalence and impact of child traumatic stress" in *Handbook of PTSD: Science and practice*, M. J. Friedman, T. M. Keane, P. A. Resick, Eds. (Guilford Press, 2007), ed 1, ch. 13.
3. K. M. Fitzpatrick, J. P. Boldizar, The prevalence and consequences of exposure to violence among African-American youth. *Journal of the American Academy of Child & Adolescent Psychiatry.* **32** (2), 424-430. (1993).
4. C. J. L Murray, A. D. Lopez, "The global burden of disease" in *Global burden of disease and injury*. (Harvard University Press, Cambridge, ed. 1, 1996), 117–200 (1996).
5. J. Angst, J., A. Dobler-Mikola, The Zurich Study--a prospective epidemiological study of depressive, neurotic and psychosomatic syndromes. IV. Recurrent and nonrecurrent brief depression. *European Archives of Psychiatry and Neurological Sciences.* **234** (6), 408-416 (1985).
6. S. E. Bruce, K. A. Yonkers, M. W. Otto, J. L. Eisen, R. B. Weisberg, M. Pagano, M. T. Shea, M. B. Keller, Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *American Journal of Psychiatry.* **162**, 1179-1187 (2005).
7. D. A. Regier, M. E. Farmer, D. S. Rae, B. Z. Locke, S. J. Keith, L. L. Judd, F. K. Goodwin, Comorbidity of mental disorders with alcohol and other drug abuse: Results from the epidemiologic catchment area (ECA) study. *Journal of the American Medical Association.* 264(19), 2511-2518 (1990).
8. M. V. Porche, L. R. Fortuna, J. Lin, M. Alegria, Childhood Trauma and Psychiatric Disorders as Correlates of School Dropout in a National Sample of Young Adults. *Child Dev. SRCD.* **82** (3), 982-98 (2011).
9. P. Sharkey, The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences.* **107** (26), 11733-11738 (2010).
10. D. F. Tolin, E. B. Foa, Sex differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin.* **132** (6), 959-992 (2006).
11. J. G. Stotsky, Gender and its relevance to macroeconomic policy: a survey. *IMF Working Paper.* Working Paper No. 06/233 (2006).

12. R. E. Ford-Paz, L. Crown, K. Lawton, H. Goldenthal, G. Day, C. A. Coyne, T. Gill, N. Harris, S. Blakemore, C. Cicchetti, Working on Womanhood (WOW): A participatory formative evaluation of a community-developed intervention. *Evaluation and Program Planning*. **72**, 237-249 (2019).
13. D. A. Brent, D. Holder, D. Kolko, A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive treatments. *Archives of General Psychiatry*. **54**, 877-885 (1997).
14. G. N. Clarke, P. Rohde, P. M. Lewinsohn, H. Hops, J. R. Seeley, Cognitivebehavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. *Journal of the American Academy of Child and Adolescent Psychiatry*. **38**, 272-279 (1999).
15. P. M. Lewinsohn, G. N. Clarke, H. Hops, J. Andrews, Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy*. **21** (4), 385-401 (1990).
16. P. Rohde, G. N. Clarke, D. E. Mace, J. S. Jorgensen, J. R. Seeley, An efficacy/effectiveness study of cognitive-behavioral treatment for adolescents with comorbid major depression and conduct disorder. *Journal of American Academy of Child Adolescent Psychiatry*. **43**, 660-668 (2004).
17. A. Wood, R. Harrington, A. Moore, A controlled trial of a brief cognitive behavioral intervention in adolescent patients with depressive disorders. *Journal of Child Psychology and Psychiatry*. **37**, 737-746 (1996).
18. P. Smith, W. Yule, S. Perrin, T. Tranah, T. Dalgleish, D. Clark, Cognitive behavioral therapy for PTSD in children and adolescents: A preliminary randomized controlled trial. *Child & Adolescent Psychiatry*. **46** (8), 1051-1061 (2007).
19. P. M. Barrett, A. L. Duffy, M. R. Dadds, R. M. Rapee, Cognitive-behavioral treatment of anxiety disorders in children: Long-term (6-year) follow-up. *Journal of Consulting and Clinical Psychology*. **69**, 135-141 (2001).
20. T. In-Albon, S. Schneider, Psychotherapy of childhood anxiety disorders: A metaanalysis. *Psychotherapy and Psychosomatics*. **76**, 15-24 (2007).
21. P. C. Kendall, E. Flannery-Schroeder, S. Panichelli-Mindel, M. Southam-Gerow, A. Henin, M. Warman, Therapy for youth with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology*. **65**, 366-380 (1997).
22. B. D. Stein, L. H. Jaycox, S. H. Kataoka, M. Wong, W. Tu, M.N. Elliott, et al., A mental health intervention for schoolchildren exposed to violence: A randomized control trial. *Journal of the American Medical Association*. **290**, 603-611 (2003).

23. S. Kataoka, L. Jaycox, M. Wong, E. Nadeem, A. Langley, L. Tang, B. Stein, Effects on school outcomes in low-income minority youth: Preliminary findings from a community-partnered study of a school trauma intervention. *Ethnicity & Disease*. **21**, 71-79, (2011).
24. S. B. Heller, A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, H. A. Pollack, Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics*. 132 (1), 1-54 (2017).
25. We thank Andrea Tentner for assistance summarizing these modalities.
26. J. White, Wellbeing and Education: Issues of Culture and Authority. *Journal of Phil. of Edu.* **41** (1), 17-28 (2007).
27. M. White, D. Epston, *Narrative Means to Therapeutic Ends* (Norton & Co, New York, London, 1990).
28. R. T. Lopes, M. M. Gonçalves, P. P. P. Machado, D. Sinai, T. Bento, J. Salgado, Narrative Therapy vs. Cognitive-Behavioral Therapy for moderate depression: Empirical evidence from a controlled clinical trial, *Psychotherapy Research*. **24** (6), 662-674 (2014).
29. In practice, many of these lists came with extra information—e.g. including additional grades and/or male students—and had to be cleaned before the randomization stage.
30. A student is considered to have a significant intellectual disability if they were listed in any of the following six disability categories: autistic; emotional and behavior disorder; educable mental handicap; intellectual disability – profound; severe/profound handicap; trainable mental handicap.
31. C. R. Reynolds, R. W. Kamphaus, K. J. Vannest, BASC3: Behavior Assessment System for Children. *PsychCorp.* (2015).
32. P. Muris, A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioral Assessment*, **23** (3), 145-149.ph (2001).
33. E. A. Kaufman, M. Xia, G. Fosco, M. Yaptangco, C. R. Skidmore, S. E. Crowell, The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and Replication in Adolescent and Adult Samples. *Psychopathol Behav Assess*. **38**, 443-455 (2016).
34. E. Cauffman, J. Woolard, Future Outlook Inventory. (1999, unpublished).
35. M. F. Scheier, C. S. Carver, Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*. **4** (3), 219–247 (1985).

36. A. Strathman, F. Gleicher, D. S. Boninger, C.S. Edwards, The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*. **66** (4), 742-752 (1994).
37. P. G. Zimbardo, Strategies for Coping with Social Traps: Times Perspectives Influences. Paper presented at the 98th Annual Convention of American Psychological Association, Boston, August. (1990).
38. K. M Harris, J. R. Udry, National Longitudinal Study of Adolescent to Adult Health (Add Health). *Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor]*. (1994-2018).
39. S. B. Heller, A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, H. A. Pollack, Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics*. **132** (1), 1-54 (2017).
40. Students who switch schools or drop out are tracked and analyzed by original group assignment. Our data suggest that 3% of CPS 9th and 10th graders switch schools each year; 9% dropout or leave Chicago.
41. P. J. Cook, S. T. Parker, H. A. Pollack, Sources of guns to dangerous people: what we learn by asking them. *Preventive Medicine*. **79**, 28-36 (2015).
42. Note that the final number of participants within each group has a random component to it. We can only define ex-ante the desired sampling probabilities within each randomization block. When creating a random sample for the baseline survey, we used sampling probabilities such that within each block, we *expected* an equal number of participants landing in the treatment vs control group. It just happened to be the case that we ended up with an equal number of 313 participants on each group for this case.
43. For the follow-up survey sample, the methodology used was the same as the one used for the baseline survey sample, only that now targeting a larger sample size (~2,000). The final number of participants assigned to the treatment group was that of 1,008; with 1,003 being assigned to the control group.
44. Note that 5 out of 1,008 participants assigned to the treatment group in the follow-up survey belong to the group of eight that were excluded from the follow-up survey ex-ante.
45. Within the control group, 3 out of 1,003 cases belong to the group of eight that were excluded from the follow-up survey ex-ante.
46. These include age, school and grade fixed effects; an indicator for having a learning disability, an indicator for having limited English proficiency, an indicator for being old

for grade, indicators for black and Latinx; and the following academic measures measured in the 2016-17 baseline school year: number of each type of grade received (A, B, C, D, F), days present, days of out-of-school suspensions, days of in-school suspensions. We also include indicators for ever being arrested for a violent, drug, or property related charge. Missingness of baseline covariates is balanced across treatment and controls groups. Missing baseline covariates were imputed using randomization-block means, by treatment status. For each imputed baseline covariate, we included an indicator identifying those observations that were imputed.

47. J. D. Angrist, G. W. Imbens, D. B. Rubin, Identification of Causal Effects Using Instrumental Variables, *Journal of the American Statistical Association*. **91** (434), 444-455 (1996).
48. H. S. Brown, Accounting for No-Shows in Experimental Evaluation Designs, *Evaluation Review*. **8** (2), 225-246 (1984).
49. T. Schell, G. Marshall, “Survey of Individuals Previously Deployed for OEF/OIF” in Invisible Wounds of War: Summary and Recommendations for Addressing Psychological and Cognitive Injuries (RAND Corporation, ed. 1, 2008), pp. 87-116.
50. Findings from a small-scale baseline survey (n=346) of study youth conducted by our research team in Fall 2017. Findings on trauma exposure do not account for experiences of direct physical abuse, sexual abuse, or sexual assault due to mandatory reporting requirements, and may be a conservative estimate.
51. Our depression and anxiety measures are drawn from the BASC-3 SRP-A. The clinically significant range is defined as two standard deviations above the mean when scores are normalized to the appropriate national norm. In our case, we use General Female, aged 15-18.
52. Briefly summarized, QALYs are constructed by multiplying the length of time a program effect is presumed to occur by a utility weight that corresponds to the quality of life of individual. One year in perfect health is equivalent to one QALY as is two years in a very poor health state with the assigned utility weight of 0.5 ($1 \text{ year} * 1 = 1 \text{ QALY}$ $2 \text{ years} * 0.5$).
53. J. R. Lave, R. G. Frank, H. C. Shulberg, M. S. Kamlet, Cost-effectiveness of Treatments for Major Depression in Primary Care Practice. *Arch Gen Psychiatry*. **55** (7), 645-651 (1998).
54. E. Gospodarevskaya, L. Segal, Cost-utility analysis of different treatment for post-traumatic stress disorder in sexually abused children. *Child Adolesc Psychiatry Ment Health*. **6** (15) (2012).
55. For the cost-benefit analysis we use the average program cost across the entire treatment group; because not everyone in the treatment group enrolls in WOW, the per-

person program cost of \$2,300 needs to be scaled-down by the participation rate before we can plug it into the ICER's equation. The percentage of study participants from the treatment group enrolled in the WOW program is 65.3%. Note that this percentage varies slightly across the different mental health constructs, as not everyone responded to all questions in the survey (e.g. it's 64.92% for depression, 65.25% for PTSD, etc.).

56. Although anxiety is one of our primary outcomes, we do not include it in the main CBA analysis because we were unable to find utility weights in the literature that mapped well onto the anxiety questions asked in the BASC-3.
57. Institute for Clinical and Economic Review, Final Value Assessment Framework for 2017-2019 (2017).

Acknowledgments: Thank you to our partners at Youth Guidance including Laurel Crown, Gail Day, Ngozi Harris, Rebecca Clarkin, Rebecca Langan, Ryan Crosby, Michelle Adler-Morrison, and the WOW program team. Thanks to Jeffrey Broom, Sarah Dickson, and Jared Sell for their help in accessing the data we analyze here. Thanks also to Kathryn Grant for assistance with the selection of study instruments, and Andrea Tentner for review of ACT and narrative therapy treatment modalities.

Funding: This paper was made possible by the generous support of the Abbvie Foundation, Arnold Ventures, Paul M. Angell Foundation, Edna McConnell Clark Foundation, Logan, MacArthur, Polk, and grant number 2016-JU-FX-0002 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, US Department of Justice. Points of view or opinions in this document are those of the authors and do not necessarily represent the views of the US Department of Justice, National Institutes of Justice or any other funder. Any errors are of course our own.

Competing Interests: Authors declare that there are no competing interests.

Data and materials availability: The data used for this research are owned and controlled by the Chicago Public Schools (CPS); the Education Lab was granted limited, restricted use of this data through a data use agreement between the University of Chicago and CPS that prohibits further sharing of the data. The data is defined in contract as Confidential Information that remains the exclusive property of CPS, and the University is not permitted to share any Confidential Information with a third party such as Science without the prior contractual consent of CPS, even if the data is de-identified. Further, the data is subject to federal privacy protections outlined in the Family Educational Rights and Privacy Act (FERPA) that prohibits disseminating any data below a minimum standard of aggregation. The University retains the right to publish research results, however, including aggregated statistics from our research findings.

Although the University is unable to directly share data from this research with Science, it is possible for researchers to directly access this same data from CPS through a standard process. Researchers may submit an External Data Request Form on the CPS website, which can be found at the following link:

<https://www.cps.edu/about/district-data/conduct-primary-research/>.

Our Master Data Use Agreement with the Chicago Police Department similarly prohibits our resharing of their confidential information, or the dissemination of any information without approval in accordance with the IL Juvenile Court Act. Our contract with CPD states: “Data provision will include juvenile arrest and victimization records which, per the Illinois Juvenile Court Act (705 ILCS 405/5-905-1-f), requires explicit approval from “the judge of juvenile court and the chief executive of the agency that prepared the particular recording.” Judge Michael P. Toomin of the Circuit Court of Cook County and Chicago Police Superintendent David O. Brown both provided written support for the Requestor’s receipt of

juvenile arrest and victimization records. The Requestor acknowledges the confidential nature of the criminal justice information supplied and agrees that disclosure to anyone not directly identified in the approved application in any manner that allows the identification of the individuals whose records have been obtained from the CPD is totally prohibited under any circumstances. Should the Requestor engage in any project involving use of the records by a third party not mentioned in this document, the Requestor will notify the Chicago Police Department. Upon review and approval by CPD, third party use will be documented in a separate agreement signed by representatives of the third party. In order to conceal the identities of persons whose records are supplied to the Requestor, the Requestor agrees to ... refrain from copying any data not absolutely necessary for the use identified in the application or its security and retain such data only so long as is necessary to conduct the program described in the application to CPD.”

The University of Chicago Education Lab is required to permanently destroy identifying information in both administrative records and in any survey or health data collected during the study period upon conclusion of the study. De-identified analytic data sets will be securely stored for a period of up to ten years following project completion, unless data partners request destruction at an earlier date. After ten years following study conclusion, all administrative data obtained for the study (that our study team does not have ownership of) will be permanently deleted from our servers. *De-identified survey data that was collected by our study team during the study period, however, will be uploaded to a public database for future use by researchers.*

BRIEF RESEARCH SUMMARY

U.S. Department of Justice Office of Justice Programs
Office of Juvenile Justice and Delinquency Prevention

Project/Award Number: 2016-JU-FX-0002

Project Title: Evaluating a Cognitive Behavioral Approach for Improving Life Outcomes of
Underserved Young Women: A Randomized Experiment in Chicago

Principal Investigators:

Harold Pollack

Faculty Co-Director, University of
Chicago Crime Lab and Health Lab
Professor, University of Chicago
School of Social Service
Administration
haroldp@uchicago.edu
773-702-4414

Jonathan Guryan

Faculty Co-Director, University of
Chicago Education Lab
Professor, School of Education and
Social Policy at Northwestern
University
j-guryan@northwestern.edu
773-848-9408

Monica Bhatt

Senior Research Director, University of
Chicago Education Lab
mbhatt@uchicago.edu
773-834-9960

Submission Date: December 29, 2022

Project/Grant Period: October 1, 2016 to September 30, 2022

Current IRB Approval/Renewal and Expiration Date: Approved 2/28/2022; Expires 2/27/2022

Working on Womanhood

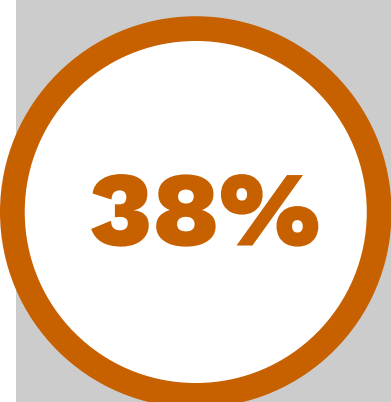
A trauma-informed group counseling program to support Chicago's girls

Why WOW?

Young people across America are facing a surge in substance use, an acute mental health crisis, and an ongoing epidemic of gun violence. These traumas have been exacerbated by the COVID-19 pandemic, particularly for Black and Latino youth. Decades of psychology research note the harms of trauma exposure on mental health, particularly during critical periods of development such as adolescence. As such, today's impacted young people may face long-term negative effects on their overall wellbeing, socio-emotional development, and educational achievement.¹

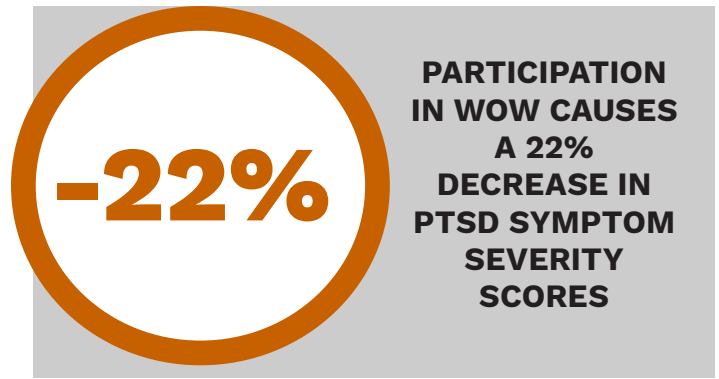
Trauma-related mental health research has traditionally focused on treatments for young men, especially in regard to gun violence prevention, even as the consequences of trauma are disproportionately borne by girls. Young women are more likely than young men to report PTSD symptoms after a traumatic event, suffer from depression, and develop an anxiety disorder. Estimates from our own work suggest that fully **38%** of school-aged girls in Chicago's neighborhood high schools exhibit signs of PTSD – a rate twice that of service members returning from Iraq and Afghanistan.²

Despite astonishing and increasing levels of trauma among school children, there is a lack of evidence about what school and city leaders can do in response. In a review of over 1,400 studies, the influential Blueprints for Healthy Youth Development has identified fewer than ten “model programs” for reducing the harmful effects of trauma exposure; none of them designed to meet the needs of young women specifically.



38%

**of GIRLS IN WOW
SCHOOLS SHOW
SIGNS OF PTSD**



Since 2011, Youth Guidance's Working on Womanhood (WOW) program has delivered in-school, trauma-focused counseling and support to girls across Chicago. Informed by cognitive behavioral therapy, acceptance and commitment therapy, and narrative therapy, WOW clinicians work to empower young women to challenge unhelpful thoughts, resolve conflict, and build resilience, self-esteem, and self-efficacy. Through teaching and reinforcing key concepts, WOW aims to reduce post-traumatic stress, anxiety, and depression symptoms among young girls in a culturally competent way, helping them to thrive in school and achieve their aspirations.

In 2017, the University of Chicago Education Lab partnered with Chicago Public Schools and Youth Guidance to test WOW using a rigorous randomized controlled trial design. This study was implemented in ten high schools citywide during SY2017-2018 and SY2018-2019. By building evidence about how we can support girls and give them tools to process and recover from trauma, we have the potential to mitigate harmful consequences like PTSD, depression, and anxiety, and subsequently improve the behavioral and life outcomes that are critical for our communities to thrive.³

What Have We Learned?

As part of the randomized controlled trial, we monitored the participation of nearly 3,800 students, 85% of whom were Black or Latina. Nearly a third of young women involved had personally witnessed someone being violently assaulted or killed, and almost half had someone close to them die violently and suddenly.

- **Preliminary results suggest that WOW decreases girls' PTSD, depression, and anxiety.** Specifically, we find that participation in WOW causes a 22% decrease in PTSD symptom severity scores, which measures the frequency and intensity of PTSD symptoms, and a 38% decrease in scores that indicate "moderate trauma-related distress." We also find that WOW participation causes a 14% decrease in depression and a 9% decrease in anxiety.
- **WOW is cost-effective and scalable.** At a cost of roughly \$2,300 per participant, WOW far surpasses the standard cost-effectiveness

thresholds used for evaluating medical and public health interventions. This threshold indicates the WOW program is not only effective, but can be replicated at-scale within resource-challenged public schools across the country. Youth Guidance has already expanded the program from Chicago Public Schools to a number of other cities, including Dallas, St. Louis, and Boston.

The evidence generated through the partnership with the UChicago Education Lab, Youth Guidance, and Chicago Public Schools is critical for understanding what works to ensure all girls are able to succeed in school and in life. The WOW evaluation has generated some of the first rigorous evidence about how cities can improve mental health and life outcomes for girls, with the potential to inform policy and practice here in Chicago and in other cities seeking to better support children who have been exposed to trauma.

To learn more about this work, please contact: Monica Bhatt, Senior Research Director, University of Chicago, mbhatt@uchicago.edu.

Endnotes

- 1 Exposure to trauma and its effect on mental health also impacts academic performance and increases the likelihood of high school dropout (Porche et al., 2011). In Chicago, for example, students spend roughly one week out of every month functioning at a lower cognitive level because of proximity to homicides (Sharkey, 2010).
- 2 Schell, T., & Marshall, G. (2008). Survey of Individuals Previously Deployed for OEF/OIF. In Tanielian T. & Jaycox L. (Eds.), *Invisible Wounds of War: Summary and Recommendations for Addressing Psychological and Cognitive Injuries* (pp. 87-116). RAND Corporation.
- 3 Research in other fields suggests that investing in young women may provide benefits beyond just the individuals served. For example, research in development and micro-investment suggests that programs that help men tend to primarily benefit individual men; whereas programs that help women also yield benefits for families and communities (Stotsky, 2006).