# PREVENTING YOUTH VIOLENCE:
# AN EVALUATION OF YOUTH GUIDANCE'S *BECOMING A MAN* PROGRAM

October 5, 2018

EVALUATION TEAM:

University of Chicago Crime Lab
33 N. LaSalle Street, Suite 1600
Chicago IL 60602
(773)-834-4292

Acknowledgements : This project was supported by the Corporation for National and Community Service's Social Innovation Fund through the Edna McConnell Clark Foundation, the University of Chicago's Office of the Provost, Center for Health Administration Studies, and School of Social Service Administration, the City of Chicago, the Chicago Public Schools, the Illinois Criminal Justice Information Authority, the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (P01-HD076816), CDC grant 5U01CE001949-02 to the University of Chicago Center for Youth Violence Prevention, Office of Juvenile Justice and Delinquency Prevention of the U.S. Department of Justice (2012-JU-FX-0019), and grants from the Laura and John Arnold Foundation, the Chicago Community Trust, the Crown Family, the Exelon corporation, the Joyce Foundation, J-PAL, the Reva and David Logan Foundation, the John D. and Catherine T. MacArthur Foundation, the McCormick Foundation, the Polk Bros Foundation, the Smith Richardson Foundation, the Spencer Foundation, the University of Chicago Women's Board, a pre-doctoral fellowship to Sara Heller from the U.S. Department of Education's Institute for Education Sciences, and visiting scholar awards to Jens Ludwig from the Russell Sage Foundation and LIEPP at Sciences Po. For making this work possible we are grateful to the staff of Youth Guidance, World Sport Chicago, the Chicago Public Schools, and to Ellen Alberding, Roseanna Ander, Rebecca Clarkin, Mayor Richard M. Daley, Anthony Ramirez-DiVittorio, Mayor Rahm Emanuel, Wendy Fine, Hon. Curtis Heaston, Michelle Morrison, Robert Tracy, and Anthony Watson. For help accessing administrative data we thank the Chicago Public Schools, the Chicago Police Department, and ICJIA, for providing Illinois Criminal History Record Information through an agreement with the Illinois State Police. For invaluable help with monitoring, data collection, and analysis we thank Nour Abdul-Razzak, Monica Bhatt, Trayvon Braxton, Sam Canas, Brice Cooke, Stephen Coussens, Gretchen Cusick, Jonathan Davis, Jaureese Gaines, Nathan Hess, Stephanie Kirmer, Anindya Kundu, Heather Sophia Lee, Duff Morton, Kyle Pratt, Julia Quinn, Kelsey Reid, Catherine Schwarz, David Showalter, Maitreyi Sistla, Matthew Veldman, Robert Webber, Nathan Weil, David Welgus, John Wolf, and Sabrina Yusuf. The findings and opinions expressed here are those of the authors and do not necessarily reflect those of the Department of Justice, National Institutes of Health, the Centers for Disease Control, the Chicago Police Department, or any other funder or data provider.

## TABLE OF CONTENTS

**Preventing Youth Violence: An Evaluation of Youth Guidance's *Becoming a Man* Program**

## Executive Summary

Improving long-term life outcomes for youth in high-risk environments remains one of our nation's most urgent challenges. Yet little is known about how to effectively and cost-effectively improve outcomes for this population. Chicago nonprofit Youth Guidance's Becoming a Man (BAM) program holds promise for improving academic achievement and reducing violence among disadvantaged youth. BAM offers youth the opportunity to participate in one-hour, once-per-week group sessions held during the school day. The program uses elements of cognitive behavioral therapy (CBT) to help youth recognize their automatic responses and slow down their thinking in high-stakes situations.

This report summarizes preliminary results from four large-scale randomized controlled trials (RCTs) carried out between 2009 and 2015 in Chicago of Youth Guidance's BAM program. This project was in part funded by the Edna McConnell Clark Foundation's Social Innovation Fund (SIF), in partnership with SIF grantee Youth Guidance and an evaluation team from The University of Chicago Crime Lab. The goal of this project was to determine whether providing non-academic supports—the BAM program—to youth can reduce youth violence and improve schooling outcomes for disadvantaged students. Additional research questions focus on the extent to which the BAM program increases social-cognitive skills.

To answer this question, we look at four main sources of data: (1) administrative data collected by the Chicago Public Schools (CPS), Chicago Police Department (CPD), and Illinois State Police (ISP); (2) two waves of in-person surveys administered to subsets of study participants; (3) a mechanisms experiment administered to a subset of study participants; and (4) program provider data from Youth Guidance. We use standard regression techniques to determine effects for intent-to-treat and treatment-on-the-treated.

Results from the various evaluations show that on the whole, the program seems likely to have positive impacts for youth. The main challenge in drawing inferences about the net effects of BAM comes from the variability in results across studies (with two seeming to show beneficial effects, one showing mixed effects, and one seeming to show adverse effects) and from statistical noise and statistical power concerns more generally. When we look at outcomes expressed in their natural units, the statistical significance of the results can be sensitive to how we aggregate information across the studies. For example the estimated effect of BAM participation on an index of school engagement measures (attendance, grade point average and persistence in school) ranges from about 0.03 to 0.08 standard deviations, depending on whether we measure impacts at the end of the first program year, the second program year, the average overall of post-program years, or at the end of the scheduled program period. The high end of this range of estimates is statistically significant but the lower end is not. This range of changes should translate into improved high school graduation rates of between 2.0% and 5.1%, according to our estimates. The effect of BAM participation on violent crime arrests, the other outcome particularly important for any benefit-cost analysis, ranges from -19% up to -37%, where once again the high end of the range is statistically significant, but the low end is not.

A different way to summarize the results across studies and outcomes is to focus on estimating benefit-cost ratios, which aggregate together different outcomes in a way that explicitly up-weights those outcomes with the greatest importance to society. Our estimates for the benefit-cost ratio of the program range from 2:1 to 10:1, depending on how we measure returns to reduced crime and improved graduation. Here again the statistical confidence interval around these estimates often includes zero, although our best estimate is that the program should easily pass a benefit-cost test given the enormous value to society from reducing costly violent crimes and increasing high school graduation rates (and hence long-term future earnings).

We have studied, and will continue to study several candidate explanations for why results differ across studies. The data we have available so far do not seem to be consistent with most of the obvious hypotheses. Better understanding the nature of this variability in impacts is our top priority for future work and central to understanding the potential of BAM to have social impact at large scale. We also note that the results documented in this report are preliminary. The research team plans to release an updated working paper of results in 2018, where we will update our findings and results.

## I. Introduction

Preventing youth violence and improving schooling outcomes for at-risk youth remain two of our nation's most urgent challenges. Unfortunately, as a society, we have made little long-term, large-scale progress in addressing disparities in these outcomes. While we have made dramatic strides in addressing many of the leading public health problems in America over the past 50 years, homicide remains a notable exception. Homicide rates today are almost exactly what they were in 1950, or even in 1900. Nationwide, homicide is by far the leading cause of death for black males 15-19 and is responsible for more deaths than the next nine leading causes combined. We have made similarly little long-term, large-scale progress in addressing disparities in educational outcomes.[1]

While there is widespread agreement about the importance of these problems, there remains great uncertainty about the best way to solve them. For example, the What Works Clearinghouse gives only three dropout prevention program its top rating for strong effects, while the Coalition for Evidence-Based Policy does not list a single program for addressing graduation rates among its "Top Tier" programs. The state of knowledge is not much better in the area of youth violence prevention (see, for example, http://www.colorado.edu/cspv/blueprints/).

What can be done to help address youth violence in Chicago and other cities? What are the important contributing factors to youth violence and negative life outcomes that are currently being unaddressed or inadequately addressed? A growing body of research has shown that various non-academic skills are strongly correlated with a wide range of important life outcomes such as schooling, employment, wages, health, and criminal involvement, even after controlling for cognitive or academic ability (Borghans, et al. 2008, Bowles, Gintis, and Osborne 2001, Cunha and Heckman 2007, Heckman and Rubinstein 2001, Heckman, Stixrud, and Urzua 2006, Moffitt, et al. 2011, Monahan, et al. 2009).

Although the importance of non-cognitive skills is now widely acknowledged, the degree to which it is possible for social policy to intervene and improve these skills remains poorly understood, particularly in the age range when so many of the most socially costly problem behaviors begin to manifest themselves – adolescence. The extent to which these interventions can affect youth outcomes of greatest policy concern, such as violence involvement or schooling attainment, remains understudied. Yet, there is some indication that non-academic skills remain more malleable after the early years of life than do academic skills (Heckman and Kautz 2012) and thus helping youth develop these non-academic skills during adolescence may be beneficial.

The goal of this project was to carry out multiple large-scale randomized effectiveness trials to determine whether providing non-academic supports—Youth Guidance's Becoming a Man program—to youth can reduce youth violence and improve schooling outcomes for

[1] Differences in graduation rates between black and white students have not changed substantially over the past 40 years (Heckman and LaFontaine 2010, Murnane 2013), while the disparity in achievement test scores between rich and poor (90th vs. 10th percentiles of the income distribution) has actually been increasing since 1940 (Reardon 2011). Increasing educational attainment and performance is central to addressing a range of social problems including poverty (Goldin and Katz 2008), health (Lleras-Muney 2005), and particularly crime (Lochner and Moretti 2004).

disadvantaged students. Developed and implemented by the Chicago-based non-profit Youth Guidance (YG), the Becoming a Man (BAM) program is an in-school program designed to help youth develop non-academic skills and more generally to encourage youth to reflect on their decision-making heuristics, or to promote both meta-cognition (to "think about thinking") and deeper emotional connections.

This report summarizes the results of four large-scale randomized controlled trials (RCTs) of the BAM program carried out with 9,804 youth in the Chicago Public Schools system. These youth are disproportionately attending CPS schools located in Chicago's south and west sides, where rates of poverty, high school dropout and violent crime are far above city and national averages. These RCTs were carried out between the years 2009 and 2015, with different samples of middle school and high school students.

This report specifically addresses the following overarching research questions:

- What is the causal effect of the BAM program on schooling and behavioral outcomes?
- Is the BAM program effective among 9th and 10th grade students?
- Do students participating in the BAM program experience increased social-cognitive skills during the course of the two-year intervention?

Results from the various evaluations show that on a whole, the program seems likely to have positive impacts for youth. However, these effects vary across samples and the conclusion about the net effects of the program are sometimes sensitive to exactly how we aggregate information across the studies. For example, the estimated effect of BAM participation on an index of school engagement variables ranges from about 0.03 to 0.08 standard deviations depending on whether we measure impacts at the end of the first program year, the second program year, the average overall of post-program years, or at the end of the scheduled program period. The second year turns out to be different from the end of the post-program period because some of our BAM studies were designed as one-year interventions. The high end of this range of estimates is statistically significant but the lower end is not. We calculate that this range of changes in our school engagement measures would be expected to translate into improved high school graduation rates of between 2.0% and 5.1%. The estimated effect of BAM participation on violent crime arrests, the other outcome that is particularly important for any benefit-cost analysis, ranges from -19% to -37% of the control complier mean (CCM)[2], where once again the high end of the range is statistically significant, but the lower end is not. We estimate that the overall benefit-cost ratio of the program, as calculated from returns to reductions in crime and improvements in predicted high school graduation rates, ranges from between 2:1 to 10:1, suggesting that the program is likely a worthwhile investment for policymakers.

Our analysis of program impacts across all samples of BAM youth suggest the program seems to have a positive impact on average on participating students, though some of the estimates are not statistically significant; what is less clear is for whom and in what contexts the program is most beneficial, differences that may account for varied impact estimates across studies. We have

[2] The control complier mean is the average outcome of those students in the control group who would have participated in BAM had they been randomly assigned to the BAM rather than control group; see Katz, Kling and Liebman 2001.

explored, and will continue to explore several candidate explanations for why results differ across studies. Current hypotheses to explain differences in program impacts include differential effects on various candidate mechanisms of action through which BAM might affect youth outcomes, different characteristics of the student samples in the studies, differences in other supports provided to students in study schools, and variation in program implementation. However, the data we have available as of this writing are not consistent with the most obvious hypotheses to explain the variation in impacts. Better understanding the nature of this variability in impacts is our top priority for future work and central to understanding how to scale BAM most effectively.

Section II describes our conceptual framework for the project. Section III describes the Becoming a Man intervention. Section IV describes our research design, and an overview of our four study samples. Section V describes the data we use to measure outcomes, and Section VI describes the randomization process across studies. Section VII describes our analysis plan, and Section VIII describes missing data. Section IX presents results from our pooled studies of BAM, Section X presents results from our individual studies of BAM, sections XI and XII present results on mechanisms of action, and Section XIII discusses these findings, their implications, and our future work.

## II. Conceptual Framework

### A. Evidence of effectiveness of other social-emotional programs

A prominent strategy to improve the life outcomes of disadvantaged youth is to focus on the non-cognitive factors that are widely thought to contribute to success in school and in life (Tough 2012). Carneiro and Heckman (2003, p. 29) argue that, "manipulating non-cognitive skills is more feasible (less costly) than manipulating cognitive skills." Yet few school systems offer explicit training on how to address and improve non-academic factors, at least after the first few years of school – perhaps because not much good evidence is available about the ability of policy interventions to causally modify these factors, or even what the most important factors are. Many existing programs are built on the assumption that if youth had more developed social and emotional skills, they would find it easier to successfully navigate their environments. In the case of youth in communities with high rates of violence, these skills may prove to be even more valuable. As a result, there are plenty of curricula that focus on training children to strengthen self-control, empathy, anger management, problem solving, and other social skills. Programs that aim to further develop these skills are necessarily complex, sometimes woven into an entire school's culture or extended beyond the classroom to involve family members, peers, and communities.

While previous social emotional learning (SEL) programs seem to have shown mixed results with youth in school settings, the quality of the empirical evidence on prior programs is typically quite limited.[3] Few intervention strategies designed to support social-cognitive skill development

[3] For example, the Positive Adolescent Choices Training (PACT) intervention helps African American youth better interact with each other and reduces school suspensions (Hammond and Yung 1991). On the other hand consider the RCT of the 4Rs program ("reading, writing, respect and resolution"), which provided a 21- to 35-lesson literacy-based SEL curriculum and 25 hours of teacher training and ongoing coaching. Jones, et al. (2011) report 50 different

meet the criteria for top-tier evidence-based programs by organizations specifically devoted to critically assessing the existing research evidence. For example, the influential Blueprints for Healthy Youth Development reviewed over 1,400 studies, and identified fewer than *ten* "model programs" that were found to reduce crime involvement among adolescents. Three of these model programs work with youth already involved with the criminal justice system and are more costly and intensive than the BAM intervention: Multi-Systemic Therapy costs over $7,700 per participant; Multi-Dimensional Treatment Foster Care costs over $8,000 per youth; and Family Functional Therapy costs about $3,500 per youth.[4] By comparison, BAM costs about $1,850 per participant per year. We turn to the cost-effectiveness of BAM and other interventions designed to improve youth outcomes at the end of this report.

### B. Potential of cognitive behavioral therapy-based programs

It is tempting to think these programs are not more effective simply because we have not yet figured out the most effective way to teach these skills. The Crime Lab research team believes that an alternative possibility is that they are built on a flawed assumption. Perhaps youth who are having trouble with school or are engaging in crime are not doing so because they lack self-control or social problem-solving skills. It is hard to argue that teenagers who live in distressed urban neighborhoods and challenging school environments cannot navigate nuanced social situations, as their well-being depends on these skills every day. However, it is plausible that a young person may not always think carefully, and instead may often behave *automatically*.

A growing body of research suggests that *automaticity* plays a key role in the decision-making of all individuals, and that because conscious deliberation is mentally costly, all of us develop a series of automatic responses that are usually adaptive to situations we commonly face. This automaticity is a universal and useful feature of how all individuals address problems and make decisions. For example, automatic behavior might lead us to pull on a door when we should push it just because the door has large handles (Thaler and Sunstein 2008). Automaticity therefore allows us to make quick decisions on a regular basis.

Automaticity is not unique to disadvantaged youth, and is a universal feature of decision-making. Usually, this reliance on automatic behavior is adaptive. But this behavior can also sometimes lead us into trouble—with consequences that may be particular severe for some of the most vulnerable youth growing up in communities where such behaviors may lead to a cascade of increasingly adverse outcomes. For example, for some youth, "fight" might be perceived to be the adaptive out-of-school response when someone tries to steal a small item; the failure to establish a reputation as someone who cannot be pushed around in such neighborhoods can potentially lead to future re-victimization (see, e.g. Anderson 1999 and Papachristos 2009). This response in certain situations can be beneficial and necessary; however, the consequences of this

---

impact estimates (intercepts and slopes for main effects, as well as interactions with baseline covariates) out of which just 4 were significant at 95% (there is about a one in seven chance we'd see that just by chance if these were all independent tests). Meta-analyses like Durlak, et al. (2011) are more positive about SEL programs overall but more than half the studies included there are not RCTs, and results just from RCTs are not reported separately.

[4] Similarly the U.S. Department of Education's What Works Clearinghouse (WWC) only gives three dropout prevention programs its top rating for strong effects (not limiting our focus to SEL programs). The Coalition for Evidence-Based Policy does not list a single program for addressing high school graduation rates among its "Top Tier" of programs.

behavior may be negatively exacerbated in some environments, particularly distressed communities with high rates of poverty and violence whose residents may also experience disproportionate rates of encounters with the criminal justice system. Therefore, above and beyond whatever need there may be for bolstering "social-cognitive" skills, many youth may need help recognizing the situations in which they are most prone to maladaptive automatic behaviors (what psychologists like Kahneman 2011 call "System 1" behavior) and the need to engage in more deliberative, reflective "System 2" decision making. With this observation, the policy challenge becomes finding ways to make young people aware of when and how their automatic responses might get them into trouble so that they will slow down and think more deliberately in high-risk situations.

The Crime Lab's theory that previous discussions about non-academic skill development have misdiagnosed what youth need helps motivate work to evaluate innovative strategies that focus on helping youth "slow down" and better diagnose what situation they are in. It specifically motivates our work in studying the potential of cognitive behavioral therapy-based (CBT) programs, which focus on this idea of automaticity and getting youth to "think about thinking" or engage in "meta-cognition" (Beck 2011).

To add to this body of evidence of CBT programs, we study a structured group mentoring and CBT-based program developed and implemented by the Chicago non-profit Youth Guidance. The goal of this evaluation was to carry out various large-scale RCTs in low-income areas of Chicago to determine whether BAM can reduce violence and improve schooling outcomes for disadvantaged youth.

## III. Description of Becoming a Man

Becoming a Man is a group mentoring program developed by Youth Guidance that uses cognitive behavioral therapy (CBT) to help participants slow down their thinking in high stakes situations. BAM participants have the opportunity to participate in one-hour, once-per-week group sessions held during the school day. The intervention is delivered in small groups—assigned groups of no more than fifteen students and average realized groups of about eight—to help develop relationships. This structure has the added benefit of being a low-cost approach to program delivery. Students are excused from a class to participate, which is a draw for some youth.

The program was developed by Youth Guidance about a decade before the first RCT of the program in academic year (AY) 2009–10. The program was operating in a single Chicago high school and a few elementary schools before being taken to scale for the first RCT in AY2009-10. The BAM curriculum itself has substantially changed since this first RCT. In 2009-10, the curriculum was designed to be one-year long, with about 30 sessions. Since 2013, and in our most recent RCTs, the curriculum is stretched out over two years (up to 45 sessions).[5] In addition, between our 2009-10 study of BAM and our 2013-15 studies of BAM, Youth Guidance made substantial efforts to provide additional training and supervision of counselors to best accommodate implementation with fidelity to the model at large-scale in Chicago. Namely,

[5] In addition, we randomize a subset of students in 2014-15 who we only have data for the first year of a two-year curriculum.

delivery of BAM in AY2013-14 included for the first time the hiring of BAM supervisors (with a staff-to-supervisor ratio of five-to-one), the development of infrastructure support and capacity building roles, a fidelity monitoring dashboard, and additional efforts to develop and manualize the curriculum. In turn, this costs of the program increased from about $1,100 per participant in 2009 to about $1,850 per participant per year in the most recent RCTs.

Table 1 illustrates a few of the key types of activities included in the BAM curriculum and provides a brief description of each selected activity. The program has a program-specific manual and facilitator's guide and is delivered by college-educated men.[6] Youth Guidance prioritizes hiring counselors who are able to keep youth engaged and aims to hire people from neighborhoods similar to those in which they would be working.

The curriculum includes standard elements of CBT, such as a common structure to most sessions that starts with a "check-in." Youth sit in a circle with the counselor, who reflects on how things in his life are going in various domains. The youth then follow suit. This activity is an example of "retrospective / introspective" activities [7], which include various efforts to get youth to talk about the things they are doing well and areas in which they still need to still improve. Youth discuss both their perception of self and their perceptions of peers on these two dimensions.

Another type of activity in the BAM curriculum is "immersive or experiential," of which one example is called the stick. Youth are divided into two groups and lined up facing each other. They are told to put their arms out chest high and extend their index fingers, and the counselor then lays a 10- or 15-foot plastic pipe across everyone's fingers. The group is then told that they must lower the pipe to the floor but their fingers must be touching the pipe at all times. This leads everyone to put upward pressure on the pipe, which makes it go up rather than down. As youth become immersed in the activity, they can lose themselves in the moment and become frustrated, blaming each other rather than recognizing that each of them contributes to the problem—and that they could help solve the problem themselves by trying to coordinate and lead the group.

Other types of activities included in the BAM curriculum are "role-playing" and "stories and discussions." For example, in the $10 role-play activity, students act out a scene in which one of them has borrowed money from another but then never paid it back. The youth act out how they would respond and then the group discusses what happened and why, and what might have led to a better outcome. Stories include the elephant and the rhino, in which two large animals are very persistent in their refusal to make way for the other, to both their detriments.

The program also does some "skill-building." This includes lessons in muscle relaxation, deep breathing, and channeling anger productively. It also includes cognitive thought replacement, a CBT element that asks youth to identify and replace problematic or false beliefs. Finally, the curriculum includes a discussion of different conceptions of masculinity and some general values like the importance of integrity and personal accountability. It also takes youth on field trips to

[6] While not required, Youth Guidance has a preference for training in psychology or social work when selecting program providers.
[7] Activity labels were coded by the research team, not by Youth Guidance, for research purposes.

local colleges to highlight the value of education, and, by putting youth in regular contact with a pro-social adult, has a mentoring component as well.

In addition to the CBT components of the curriculum, BAM counselors also aim to develop additional social-emotional skills through the program including impulse control/emotional self-regulation, social information processing, future orientation, and integrity. Weekly group sessions are structured around six core values to develop these social-emotional skills: integrity, accountability, self-determination, positive anger expression, visionary goal-setting, and respect for womanhood. These values are delivered via a multi-faceted approach using various youth engagement activities. By focusing sessions on CBT and these core values, Youth Guidance hopes to ultimately improve non-cognitive skill development and academic engagement, with the immediate goals of reducing school suspensions and increasing school attendance, the intermediate goals of improving academic achievement and decreasing arrests, and the long-term goals of improving graduation rates and reducing involvement in violence among participants.

BAM is, in short, a program with multiple elements and hence multiple potential mechanisms of action through which it may change youth behavior and outcomes. We return below to the question of what can be surmised about what the most important mechanisms might be.

## IV. Research Design and Student Samples

### A. Overview of all studies

Table 2 below presents a summary table of all studies conducted to date on the Becoming a Man program, with an indicator of whether the study was funded in part by the Social Innovation Fund (SIF). *BAM 1* refers to the original RCT conducted of a one-year BAM curriculum in academic year (AY) 2009-10. *BAM 2* refers to a second study of a two-year curriculum conducted from AY2013-15 in which students were assigned to BAM or a control group in 2013. *BAM 2x2* refers to a third study of the two-year curriculum conducted in AY2013-15 that was part of a larger 2x2 factorial experiment that aimed to measure the synergistic effects of academic and non-academic supports.[8] *BAM Expansion* refers to groups of students who were randomized in 2014, or the second year of the 2013-15 study, but for whom we do not have the second year of participation data from AY2015-16.

We analyze various subsamples depending on the set of schools the students attended and the year in which students were randomized, as depicted in Table 2. Each of these samples are described in greater detail below.

[8] In the 2x2 schools we also independently randomly assigned students to receive high-intensity tutoring delivered by SAGA Innovations, reported on by Cook, et al. (2015). Results from the resulting 2x2 intervention and the implications for how academic and non-academic skills interact will be the topic of a separate paper.

**Table 2: Overview of all studies**[9]

| *Study* | 2009-10 | 2010-11 | 2011-12 | 2012-13 | 2013-14 | 2014-15 | 2015-16 |
|---|---|---|---|---|---|---|---|
| **BAM 1** *(18 schools)* | N= 2,740 | | | | | | |
| **SIF-Funded Schools** *(21 schools in both AY2013-14 and AY2014-15)*[10] | | | | | | | |
| **BAM 2** *(9 schools)* | | | | | N=2,064 | | |
| **BAM 2x2** *(12 schools)* | | | | | N=2,633 | | |
| **BAM Expansion** (*21* schools) | | | | | | N=2,367 | |

We note that all control group students in our studies were allowed to participate in status quo school and community services. Neither Youth Guidance nor the evaluation team thoroughly documented all of the additional services available at the school. However, based on our knowledge of the schools, we do not currently know of similar programs that were being administered at any of the schools during our study years.

## B. Overview of individual studies

### 1. BAM 1: Randomization in AY2009-10

In AY2009-10, the University of Chicago Crime Lab conducted a study of the BAM program in 18 CPS schools in Chicago's distressed south and west sides. In this RCT, 2,740 male youth in 7th through 10th grade were assigned to one of two groups for one academic year—to BAM, or to a control group that received status quo school and community services. It is important to note that the BAM intervention in our 2009 study differed slightly from the intervention studied in our 2013-15 evaluations. The most significant difference is that the BAM curriculum for our BAM 1 study only spanned one year, while the curricula for the remaining BAM evaluations spanned two years.[11] Other differences include modifications to the curriculum itself and different training provided to counselors.

Table 3 shows the sample of youth in BAM 1 represent a population that has high levels of criminal justice involvement and disengagement from school. Youth were between 15 and 16 years old at baseline. About 70% were African-American, and 30% were Hispanic. Youth had an average GPA of 1.7 on a 4-point scale, and about a third of youth had an arrest at baseline.

---

[9] N refers to the number of students who are randomized in the study sample. Shaded cells signify when the intervention occurred. The patterned shading refers to data that we do not have at this time.

[10] Though 21 schools received BAM in AY2013-14 and AY2014-15, only 20 schools implemented BAM *in both* AY2013-14 and AY2014-15. One school implemented BAM in only AY2013-14, and another school implemented BAM in only AY2014-15.

[11] For more information on the 2009 BAM study, please see Heller, et al. (2017).

### 2. BAM 2: Randomization in AY2013-14

In AY2013-14, the University of Chicago Crime Lab began its second study of BAM by randomizing 2,064 male 9th and 10th graders to one of two groups in nine CPS high schools for two academic years—to BAM, or to a control group that received status quo school and community services. We note that one of these nine schools did not continue the BAM curriculum in AY2014-15, and we thus drop this school when looking at impacts during year 2 of the study.[12]

Table 3 indicates students in the BAM 2 study were slightly younger than in BAM 1, and had lower levels of criminal justice involvement and poorer school performance. In this study, youth were about 15 years old at baseline. Reflecting the composition of their neighborhoods, around two-thirds of youth are black and the remainder Hispanic. Youth had an average GPA of 2.1 on a 4-point scale, and about a quarter of youth had an arrest at baseline.

### 3. BAM 2x2: Randomization in AY2013-14

Randomization for the BAM 2x2 students occurred in twelve CPS high schools among 2,633 male 9th and 10th grade students. As these schools were part of a larger 2x2 experiment, the study sample was randomized to one of four groups—to BAM, a high-intensity math tutoring program, both BAM and the high-intensity math tutoring program, or a control group that was offered neither program but received status quo school and community services. To isolate the effects of BAM, we compare all students who did not receive BAM (all students who were randomized to receive math tutoring or to the control group) against all students who did receive BAM (all students who were randomized to receive both programs, or to just receive BAM).[13] Youth randomized to treatment were to receive BAM for two academic years. We drop one school from our BAM 2x2 analyses due to a "broken experiment" in that school; that is, the program providers in the school failed to follow the random assignment and (we fear) incompletely documented which youth actually received the program.[14]

Table 3 describes the population of youth for our BAM 2x2 sample. Over 45% of youth in our sample are African-American, and over 45% are Hispanic. This differs substantially from our

---

[12] One of the schools in our BAM 2 study decided that they did not want the program to continue in AY2014-15. We include this school when looking at impacts for year 1, but drop this school when looking at impacts for year 2.

[13] There may be some concerns with collapsing all students who receive BAM in our analyses, as we may be at risk for underreporting BAM effects across individual study results if the synergistic effects of BAM and the math tutoring program somehow diminish BAM effects. However, based on our preliminary analysis of the full two-by-two experiment, we do not believe this to be true.

[14] We received notification from Youth Guidance leadership post-study that a BAM counselor in one school allowed control students in BAM sessions without documentation in his attendance data. When conducting analysis for this study in fall 2016, we received names from this counselor of additional control students he served in both AY2013-14 and AY2014-15. In addition, there were two other counselors at this school for whom we do not know whether they served control students or not. In our own internal program data, we calculate that this school served four BAM control students out of a total population of 90 students served over both program years. Using the counselor's additional rosters, we find that 36 BAM control students were served out of a total sample of 130 students served over both program years. We do not have similar documentation of control students served by the other two counselors. Due to the failure of randomization at this school we drop it from our BAM 2x2 analyses.

BAM 1 and BAM 2 samples, where about two-thirds of study participants are African-American, and one-third are Hispanic. About 86% are eligible for free lunch. The average GPA for these youth was 2.1 on a 4-point scale, and about 20% of youth had an arrest at baseline. Fewer youth had an arrest at baseline in this sample compared to our BAM 1 and BAM 2 samples.

#### 4. BAM Expansion: Randomization in AY2014-15

During the AY2014-15, we randomized an additional cohort, mostly of incoming $9^{th}$ graders, to receive BAM in both BAM 2x2 and BAM 2 schools. As our study ended in AY2014-15, we do not have data for these youth from their second year of intervention, which would have occurred in AY2015-16. Consequently, we report data for these youth for one year of intervention to date. We return to this issue when pooling the estimates from all of our BAM samples.

Table 3 describes the population of youth for our BAM expansion study sample. Of the 2,367 youth who are in this sample, about 60% of youth are African-American and about 35% are Hispanic. The average age of students is 14.5, and about 88% are eligible for free or reduced lunch. The average GPA for these youth is 2.3 on a 4-point scale, and about 20% had an arrest at baseline. We again note that fewer youth had an arrest at baseline in this sample compared to our BAM 1 and BAM 2 samples, and that students are younger in this sample.

## V. Data

Our study relies on four main sources of data: (1) longitudinal student-level records from administrative data collected by the Chicago Public Schools (CPS), Chicago Police Department (CPD), and Illinois State Police (ISP); (2) two waves of in-person surveys administered to subsets of study participants; (3) a mechanisms experiment administered to a subset of study participants; (4) program provider data from Youth Guidance. Our team also conducted in-person observations in a random sample of BAM sessions to monitor and document program implementation.

### A. Administrative Data

#### 1. Chicago Public Schools (CPS) Data

Our first source of administrative government data come from longitudinal student-level records maintained by CPS. Participants are matched to school administrative records using their unique CPS student ID using STATA. These CPS student records include whether the student has a learning disability; month and year of birth; race/ethnicity; eligibility for free and reduced price lunch; course grades in each subject; enrollment status; absences; and disciplinary actions and suspensions. For all of our BAM studies, we create a summary index of school engagement to both reduce the risk of false positives, and to improve statistical power to detect effects for outcomes within a "family" of outcomes that are expected to move in a similar direction. This index is an average of three Z-scored variables: GPA at the end of the school year, days present during a school year, and enrollment status at the end of the year (i.e. whether a student was enrolled at the end of the school year).[15]

Student participation files are initially linked to CPS administrative data in STATA using a unique CPS student ID. Demographic data, such as birth date, learning disability status, and race, are then added from a CPS enrollment file. We then use this demographic information to connect our CPS data to crime data, as described below.

### 2. Chicago Police Department (CPD) Data

Our second source of administrative government data for these studies are from Chicago Police Department (CPD) and Illinois State Police (ISP) arrest records.

The CPD data that we utilize include information on the identity of the offender, date and location of the crime event, and the criminal charges (for juvenile as well as adult offenders). When recording arrests, CPD uses fingerprint technology to identify individuals. The arrest data should therefore include every CPD arrest of an individual, even if he or she submits an alias at the time of arrest. The data also includes arrests that do and do not result in a conviction. We link CPD data to our study samples using probabilistic matching on first name, last name, gender and date of birth.[16] We use CPD data in all of our BAM studies, with the exception of our BAM 1 study.

In our BAM 1 study, we do not use CPD data, but instead use electronic arrest records ("rap sheets") from the Illinois State Police (ISP), obtained through the Illinois Criminal Justice Information Authority (ICJIA). The ISP records capture arrests in the state going back to 1990 and include arrests of people below the age of majority within the criminal justice system (juvenile arrests), as well as to those who are above the age of majority. Local police departments are required by law to report all juvenile felony arrests to the ISP, and optionally class A and B misdemeanors. ICJIA uses probabilistic matching on name, gender and date of birth to match our study sample to ISP arrest records.[17]

Because previous studies often find more pronounced impacts of policy interventions on violent crimes (particularly impulsive crimes such as assault) than on other crimes (Deming 2011, Evans and Owens 2007, Kling, Ludwig and Katz 2005, Lochner and Moretti 2004, Weiner, Lutz and

---

[15] We do not include standardized test scores in our school engagement index because the Chicago Public Schools by design did not administer standardized tests to all grades during the years of our study. In addition, our index does not include administrative records on school disciplinary actions, as we are not certain of the validity of this data.

[16] In our studies, potential matches to CPD data were reviewed and classified using a machine learning algorithm as well as an additional manual review of borderline cases. The resulting arrest-level data was categorized according to the offense type using a combination of the FBI's Uniform Discipline Code and the accompanying statute description from the Chicago Police Department. The data were then aggregated to the student-level and merged onto the analytic file using the CPS student ID variable.

[17] Once the research team received ICJIA arrest data for the BAM 1 study, the data was then categorized according to the offense type using a combination of FBI's Uniform Discipline Code and the accompanying statute description from the Chicago Police Department. The data was then aggregated to the student-level and merged onto the analytic file using a unique CPS student ID.

Ludwig 2009) and because associated social harms are so varied across crime types, we examine arrests separately for different offense categories. For each arrest incident, we select the most severe charge associated with the incident. In most cases this is a charge recorded at the time of arrest, although occasionally the State's Attorney files a charge more severe than those originally recorded at the police station. We classify crimes as violent, property, drug, and other, as follows:

(i) *Violent crimes* include murder, rape, assault, robbery, threats/harassment, and kidnapping.
(ii) *Property crimes* include larceny, burglary, and auto theft.
(iii) *Drug crimes* include possession or dealing charges.
(iv) *Other crimes* include trespassing, fencing, bribery, animal cruelty, weapons violations, DUIs, disobeying or avoiding law enforcement officers, disorderly conduct, arson, prostitution, criminal neglect, parole violations, underage or public drinking, vandalism, and miscellaneous offenses.

We exclude motor vehicle crimes, including driving with a suspended license, reckless driving, and other driving/traffic related offenses, from our analysis. These are rare in our data.

### B. Mechanisms Survey

A second source of data comes from two waves of in-person surveys carried out for our research team under sub-contract by the Institute for Social Research (ISR) at the University of Michigan. We note that ISR surveys were only administered to students who were randomized during the 2013-14 academic year, and not for students randomized during the 2014-15 academic year. For budget reasons, in both survey waves we selected random sub-samples of youth out of our larger analysis sample for the ISR survey sampling frame. ISR used two-phase sampling in which after interviewing approximately 70% of the survey sample, they selected a random sub-sample of youth in the second stage for more intensive follow-up. Our analyses employ sampling weights that account for this two-phase sampling design.

The first wave of surveys, funded in part by the Corporation for National and Community Service's Social Innovation Fund, was carried out close to the end of our first intervention year from May through June 2014. Only youth in our BAM 2x2 schools received this round of ISR surveys. We selected a sub-sample of 881 youth from this sample; ISR completed surveys with 658 youth with an effective response rate of 88.1%. To achieve the target response rate, ISR completed some surveys in the fall of 2014. Month of survey completion is fairly balanced across randomized groups, as is the overall survey response rate.[18]

Our second wave of in-person survey work, funded by the National Institute of Child Health and Human Development, was carried out in the first post-intervention year from November 2015 through July 2016. Youth who were in the BAM 2x2 schools and the BAM 2 schools received this second round of ISR surveys. We selected a random sub-sample of 1,702 youth out of our

[18] Specifically, for our first wave of surveys, we note that the overall response rate (not accounting for sampling weights) for the four treatment arms was: BAM (77%); BAM + math tutoring program (77%); math tutoring program (73%); and control (80%).

larger analysis sample for the ISR survey sampling frame. ISR completed surveys with 1,238 youth with an effective response rate of 87.0%. Again, the survey response rate is fairly balanced across treatment groups.[19]

For these surveys, we present information by aggregating questions into the key mechanisms that our survey was designed to measure. Survey questions originated from a variety of sources and standard scales used to measure psychometric properties. Please see Table A1 for more information regarding the reliability and sources of our survey measures.

### C. Mechanisms Experiment

An additional source of mechanisms data comes from a mechanisms experiment we conducted with a subset of students randomized into our BAM 2 study in 2013-14. We recruited 490 participants (266 who had been assigned to BAM programming, 224 who had been assigned to the control group) from the nine schools in which 1,551 youth (775 treatment, 776 control) were eligible to participate.

To examine how BAM changes decision-making in confrontational situations where youth are provoked and retaliation is a possibility, and specifically whether BAM causes youth to "slow down," we had participants play a modified version of a real-stakes iterated dictator game. In the experiment, participants were given $10 in one-dollar bills in an envelope. After their "partner" (who was actually a member of the research team) took $6 away from the participant, the participant was then asked how much money they would like to take away from their partner. The purpose of this experiment was to see whether BAM participants would make slower, more deliberate decisions as compared to control group participants.

### D. Program Provider Data

Another source of data comes from Youth Guidance. Data we receive directly from Youth Guidance includes program attendance data, which provides us with program attendance data for each BAM session held. For our 2013-15 evaluations, BAM counselors keep a log of each group session they lead, which documents the lesson and core values that were taught during that session, along with a list of students that attended. In addition, BAM counselors documented each individual counseling session, as well as any field trips or out-of-school activities each student was engaged in. Counselors collected this attendance data via a tablet computer and an internal electronic client information system developed by Youth Guidance. Counselors were mandated to fill out this information.

In addition, for these evaluations we receive measures of fidelity of implementation for each counselor based on assessments by Youth Guidance leadership, specifically of counselor and school quality ratings for each school.

[19] For our second wave of surveys, we note that the overall response rate (not accounting for the sampling weights) for the four treatment arms was: BAM (78%); BAM + math tutoring program (80%); math tutoring program (79%); and control (77%).

### E. In-person Observations

To monitor program implementation during the study period we conducted in-person observations of a random subset of over 90 BAM sessions during AY2014-15. We selected a random subset of sessions, and had a group of graduate and undergraduate research assistants conduct systemic observations of BAM sessions using a standardized observation rubric designed to record information about implementation completeness, quality, and fidelity to the prescribed manual. The BAM rubrics examined constructs related to transitions into and out of BAM sessions, student engagement, activities that occurred during the sessions, counselor leadership, relationships between counselors and students and the overall climate of sessions. These observations were conducted in the second year of our two-year evaluation, from November 2014 through the end of the 2014-15 academic year.

## VI. Randomization, Recruitment, and Retention Procedures

### A. Randomization Procedures

Below, we describe the randomization procedures that occurred in all four RCTs. We note that all student-level randomization is carried out in STATA.

During the summer preceding each study, the evaluation team identified eligible students in CPS schools using CPS administrative records from the previous academic year. In BAM 1, this included students who were in grades 7-10, in BAM 2 and BAM 2x2 this included students who were in grades 9 and 10, and in BAM expansion this mostly included students who were in grade 9. Following the approach used in Heller, et al. (2013), we first excluded those students who we thought were already too disengaged from school to attend regularly enough to benefit from a school-based program. This exclusion criterion was defined as having failed 75% or more of their classes in the previous school year and having missed more than 60% of their enrolled school days in that year. We also excluded students with serious disabilities as designated by the CPS data.[20]

We then calculated a "risk index" that was a function of the number of prior-year course failures and unexcused absences, and being old for grade (interpreted as having been previously held back). Eligible students for the program were then ranked on the basis of this risk index. The research team determined the number of randomized students needed to utilize all available program slots in a school and chose that number of students in descending order on the ranked risk list. The share of eligible students in each study sample varies across schools because of school-by-school variation in both program capacity and school size. In practice, because of the scale of the experiments, in many schools we randomized all students who were not excluded based on their prior year course failures and absences. Essentially one can think of our study

[20] For our BAM 1 study, we use administrative data from the previous school year (AY2008-9) to create our study sample. Specifically, we receive enrollment data using administrative data from the previous school year. For our BAM 2, 2x2, and expansion samples, we receive enrollment data during the summer prior to the intervention, and match these enrollment files with administrative data from the first semester of the previous year (AY2012-13 for study 2 and study 2x2, and AY2013-14 for study expansion) to build our study sample.

samples as pools of youth in distressed Chicago schools in the middle of the distribution for these schools, with both the left (lowest achieving) and right (highest achieving) tails trimmed.

In order to accommodate the varying program capacity within each school, our random assignment algorithm varied the probability of treatment condition assignment.[21] Since our randomization was carried out separately by school and grade for each study, we treat each school-grade combination as separate randomization blocks.[22]

In schools where too few students actually showed up in the fall to randomize, we identified new students entering the school (mostly during the first month of that school year) and randomly assigned them to treatment and control conditions. Specifically, during the school year, members of the evaluation team randomly assigned newly enrolled youth in each school using a spreadsheet pre-populated with treatment and control assignments for each new student added to the study sample. For these students the randomization block is defined by the school and the time period in which the youth was randomized.[23] All of our analyses below control for randomization-block fixed effects.

We include every student we randomized and for whom we have CPS data for in our analysis, including those who were assigned but subsequently left the pool of study schools.

We note that all randomization procedures were carried out in STATA for all four studies. Randomization into treatment and control groups was done by a member of the evaluation team using STATA. Balance among the assignment was verified by regressing treatment assignment against each of the pre-randomization characteristics we used to select our sample, while controlling for block-level fixed effects. Individual and joint significance was checked for each assignment group relative to the control group.

### B. Recruitment and Retention Procedures

After randomization occurred, the evaluation team sent lists of eligible students for the program to BAM counselors at each of the study sites. At the beginning of each study's school year, BAM counselors individually approached these eligible students and invited them to join the program. BAM counselors reviewed informed consent forms and processes in advance with the evaluation team, and followed up with both students and parents to directly answer any questions they had regarding participation in the program. Both parental consent and student assent were obtained from all program participants, and only those who had submitted both forms were able to participate in BAM programming. Students who were randomized into the control group were not approached for consent as they were only tracked through administrative data.

---

[21] Our general rule was to randomize enough people into treatment groups to hit enrollment targets if we achieved a 75% take-up rate, and ideally to have a control group at least as big as the smallest treatment cell. In some schools because of the need to fill treatment slots, our control group was smaller than any of the treatment groups.

[22] We note one exceptions to this procedure in our first BAM study conducted in AY2009-10. In this study, each school, instead of school-grade, is treated as a separate randomization block.

[23] We note that we do not randomize new students entering the school in our first BAM study conducted in AY2009-10.

We find that in our BAM 1 study, 49% of those randomized into the program ever participate in a BAM group session, as compared to 52% in our BAM 2 study, 50% in our BAM 2x2 study (specifically, 50% for those only randomized to only receive BAM, and 51% for those randomized to receive BAM and the math tutoring program), and 31% in our BAM expansion study. We also find that in our BAM 1 study, the average number of BAM sessions attended among all students who were randomized into treatment over the program period was 5.2 sessions in BAM 1, 12.7 sessions in BAM 2, and 14.4 sessions in BAM 2x2 (specifically, 14.4 sessions for those only randomized to receive BAM, and 14.4 sessions for those randomized to receive BAM and the math tutoring program). The average number of BAM sessions attended among students who ever participated is 11 sessions in BAM 1, 24.1 sessions in BAM 2, and 28.6 sessions in BAM 2x2 (specifically, 28.8 sessions for those randomized to receive only BAM and 28.4 sessions for those randomized to receive BAM and the math tutoring program). Lastly, if we define attrition as the difference in year 1 and year 2 participation rates in our BAM studies that employed a two-year curriculum, we find that the BAM 2 sample has a 17% attrition rate, and the BAM 2x2 sample has a 13% attrition rate (specifically, 13% for those randomized only to receive BAM and 13% for those randomized to receive BAM and the math tutoring program).

## VII. Analysis Plan

We illustrate our approach to estimation in a simple regression framework. If $Y$ is some social/behavioral outcome of interest, $S$ is an indicator for assignment to Becoming a Man, $X$ is a set of baseline youth characteristics or pre-randomization outcomes (included in the model to improve statistical power), we would estimate the main intention to treat (ITT) effect as follows.

$$(1)\ Y = \pi_0 + \pi_1 S + \pi_2 X + \varepsilon$$

The main effect of BAM is given by $\pi_1$.We initially estimate equation (1) with ordinary least squares, but for dichotomous dependent variables also re-estimate (2) using non-linear maximum likelihood models like probit and logit (although in practice the average marginal effects from the two approaches tend to be similar).

Since program take-up is never 100%, we calculate effects based on assignment rather than participation (i.e., the ITT). This yields an unbiased ITT estimate regardless of the take-up rate and regardless of whether those youth assigned to treatment who comply (participate) are systematically different from those assigned to treatment who do not comply.

We also estimate the effects of participating in the program, or the effects of treatment on the treated (TOT), by imposing a linear functional form and using random assignment to different treatment conditions as instruments for actual participation (see Bloom 1984 and Angrist, Imbens and Rubin 1996, and for applications see Kling, Ludwig and Katz 2005, Kling, Liebman and Katz 2007, and Ludwig, et al. 2012). Whereas the ITT isolates the causal effect of being offered treatment, TOT estimates isolate the effect of the treatment for the subset of subjects who choose to participate. The TOT estimate is essentially the ratio of two experimental ITT effects: the ITT effect on the outcome of interest ($Y$) divided by the ITT effect on participation rates in the intervention being studied. This method recovers the TOT if assignment to treatment has no effect on outcomes for subjects who do not participate.[24, 25]

For all of our BAM evaluations, baseline demographic characteristics we use in regression models include: age, learning disability status, free/reduced lunch status, race, school-level randomization block, grade level at baseline, GPA at baseline, grades at baseline, school attendance, school disciplinary history (suspensions and incidents), and arrest history.

Our modeling assumptions are as follows: (1) treatment assignment is random and (2) outcomes are only affected through treatment participation. Random assignment is checked through the balance test described previously. We check the second assumption by looking for changes in outcomes among those who were assigned to treatment but did not participate.

We note that all outcomes analyses are conducted in STATA. The final analysis dataset that we use for all four studies is a student study-level dataset where we cluster on students when running pooled study results.

## VIII. Missing Data

We distinguish between two types of missing data in our study: missing data from the baseline variable used on control variables in our outcomes models, and missing data in the outcome variables. In general, youth in study 1 have lower levels of missing baseline information for grades, which is a result of the study team's use of the prior year enrollment file to select youth for randomization. In BAM 2, BAM 2x2, and BAM expansion studies, the number of youth missing grade data is higher due to the use of summer enrollment files, which include new youth who transferred to Chicago Public Schools after the school year ended. See the table below for missingness of attendance and grades by study. In BAM 1 there is some evidence of a statistically significant difference in missing attendance data between treatment and control groups, but otherwise there is no evidence of differences within and across studies. Test of differences is carried out by regressing an indicator for missing data against an indicator for treatment assignment, controlling for randomization blocked fixed effects. We then assess the significance of the treatment assignment coefficient.

[24] If no controls participate in the program, then our instrumental variables estimate identifies the average effect of everyone who is "treated" (participates in the program), the TOT. If some control group members wind up receiving program services, which in a complicated real-world setting like CPS could potentially happen to some small degree, then our IV estimates are still valid, they just estimate a slightly different parameter – the local average treatment effect (LATE) for subjects who participated because they were selected to be in the treatment group but would not have participated if they had been in the control group. This group is called "compliers" in the typology of Angrist, Imbens, and Rubin 1996. We begin our analysis with models using school-level fixed effects.

[25] For example, if treatment assignment results in a 5% increase in outcome Y, and the participation rate is 50%, our TOT estimate becomes 0.05/0.5, or a 10% increase.

**Table 4: Missing Baseline Variables by Study**

| | Missing prior grades | Missing prior attendance |
|---|---|---|
| BAM 1 | 6% | 5% * |
| BAM 2 | 14% | 4% |
| BAM 2x2 | 7% | 3% |
| BAM Expansion | 9% | 4% |

Significance is indicated by *** if p<0.01, by ** if p<0.05, and by * if p<0.1

For youth missing baseline variables, we created new variables with zeros imputed for any missing observations. We use these along with indicator variables flagging students who had a given baseline variable imputed, into all models.

Missing data in the outcome variables increases each of the post-randomization period.

**Table 5: Missing Outcome Variables by Study**

| | Missing grades, year 1 | Missing grades, year 2 | Missing attendance, year 1 | Missing attendance, year 2 |
|---|---|---|---|---|
| BAM 1 | 10% | 33% | 3% | 17% |
| BAM 2 | 27% | 39% | 7% | 17% |
| BAM 2x2 | 15% | 27% | 5% | 14% |
| BAM Expansion | 23%* | 32% | 8% | 16% |

Significance is indicated by *** if p<0.01, by ** if p<0.05, and by * if p<0.1

Using the same test described above for missing baseline data, we see some evidence of a difference in the proportion of treatment and control youth missing grade data in the BAM expansion sample's first year. There is no further evidence of differences within or across studies. By construction there is no missing data in the baseline or outcome arrest variables, nor in the third element of the school index, which measures whether or not youth were still enrolled at the end of a given school year.

For youth missing outcome data for grades, attendance, or both, we impute the treatment or control mean. Heller, et al. (2013) describes this approach as having "the advantage of using all available information and having a straightforward substantive interpretation: it is equivalent to estimating the treatment effect on each individual component of the index (in standardized form) using only observations with non-missing observations, and then averaging the component-specific estimates (Kling, Liebman and Katz 2007, Anderson 2008, Schochet, Burghardt and McConnell 2008)." This assumes that missing grade or attendance data is not related to any observable or unobservable characteristics of the youth, nor to any outcome data. In other words, the data are assumed to be missing completely at random (MCAR).

## IX. Results of Pooled Studies

In what follows here we report the results of analyzing across the different study samples we have examined in Chicago to date.

### A. Outcomes

We begin by pooling together data from all BAM RCTs conducted to date (Table 6) to answer the initial policy question of interest: Does the BAM intervention improve youth outcomes?

To answer the question of what BAM does to affect youth outcomes we use all the information we have available to us about BAM – that is, we pool data from our four BAM RCTs. This raises the question of how best to aggregate information from the different studies. There is no obviously perfect way to do this: for two of our BAM RCTs we have only one year of program data, which captures the entire program period for the BAM 1 study but only the first of two intended program years for the BAM expansion. We err on the side of showing the reader more rather than less of the data, and aggregate results in four reasonable ways:

- The effect measured at the end of the first program year, which will then have the drawback of including data on the BAM 2, BAM 2x2, and BAM expansion measured essentially during the middle of the two-year program period;
- The effect measured at the end of the second program year, which has the drawback of excluding data from the BAM 1 study and BAM expansion;
- The average effect calculated over all program years for which we have data;
- And the effect measured at the end of program completion period, which excludes data from the BAM expansion study.

On balance the analysis from this pooled sample of BAM youth suggests the program seems likely to have a positive impact on participating students: The majority of the estimated impacts in Table 6 are in the direction of improved youth outcomes, although the exact magnitude and level of statistical significance of the impact estimates varies across both outcomes and choices about when and how to measure program effects.

For example, the impact of participation in BAM among youth on school engagement ranges from 0.03 standard deviations to 0.08 standard deviations. The high end of this range is statistically significant, while the lower end of the range is not. We can get some sense of the potential impact on high school graduation rates by using a rubric published by the University of Chicago's Consortium on Chicago School Research which predicts graduation rates using GPA.[26] We expect that these program effect sizes translate into improved high school graduation rates of between 2.0% and 5.1% for participating youth.

The one additional outcome measure that will play a particularly important role in any benefit-cost analysis of the program is violent-crime arrests, which in Table 6 shows a fairly consistent decline in arrests across the four ways of aggregating information from the different RCTs. Specifically, participating in BAM reduces involvement in violent crime by 19% to 37% of the

---

[26] See: https://consortium.uchicago.edu/sites/default/files/publications/07%20What%20Matters%20Final.pdf.

control complier mean (CCM). Here again the high end of this range (in absolute value) is statistically significant while the lower end of the range is not.

The other outcomes show somewhat more mixed patterns of impacts across model specifications. For example for all arrests, three of the four model specifications suggest a decline in total arrests, ranging in magnitude from -7% to -21%, but one model specification (the estimated second-year effect) is now positive, at +13%. For arrests for property and drug crimes most point estimates are slightly positive but very close to zero. For the "other" arrests category, which includes weapons offenses and crimes such as disorderly conduct and trespassing, three of the four estimates are negative and range in size from -16% to -27%, although one of the four estimates is positive, equal to +23%.

We also present results in Table A2 using multiple imputation, which adjusts for the uncertainty involved in making imputations. Multiple imputation uses a Bayesian approach, drawing from the conditional predictive distribution of the missing variable(s) to impute values of the missing data, and iterating the process to analyze $m$ (we use $m = 10$) of these simulated datasets (Little & Rubin 2002; Puma, et al. 2009). We run separate imputations for the treatment and control groups for missing GPA and attendance information, then recalculate the school engagement index within each dataset and re-run our regressions. Imputing outcomes separately for treatment and control groups avoids injecting correlation with the treatment indicator into the imputation (Puma, et al. 2009).[27] Using this imputation approach, we find that those results that were statistically significant without multiple imputation typically remain significant, albeit with a somewhat higher ("less significant") p-value due to a combination of slightly attenuated coefficients and slightly larger standard errors.

Table A3 shows that accounting for multiple comparisons has a similar effect of pushing the p-values slightly higher. For each outcome, we calculate one-stage q-values for the false discovery rate, or FDR-q (the share of significant estimates that are expected to be false positives; see Benjamini and Hochberg 1995), the two-stage FDR-q (the smallest value of which we could reject the null for each outcome using the method from Benjamini, Krieger, and Yekutieli 2006), as well as the family-wise error rate, or FWER (the chance that at least one of our outcomes in the 'family' of outcomes is significant when the null hypothesis of no effect is true; see Westfall and Young 1993). As an example of how these adjustments impact our estimates, we can look at the at the program completion effect on school engagement, where the unadjusted p-value is 0.007 (significant at $p<.01$). After adjustments for multiple comparisons, the p-value remains significant albeit now only at the $p<.05$ level.

### B. Cost-Benefit Analyses Using Pooled Results

In addition to looking at the effects of the program on our outcome variables expressed in their natural units, we present results of benefit-cost estimates that have the effect of aggregating together different outcomes in a way that up-weights those outcomes that are relatively more consequential from society's perspective.

[27] For more information on our multiple imputation approach, please see Appendix D to Heller, et al. 2013.

The costs of BAM are relatively modest. Our best administrative cost estimates are $1,100 and $1,850 per participant per year in BAM 1, and in BAM 2/BAM 2x2/BAM expansion, respectively. To be consistent in our pooled estimates, we assume conservatively that the yearly per participant costs of the program are fixed at $1,850.

Table 7 presents results of our benefit-cost analysis for the program completion and average program effects for our pooled estimates. The table breaks the benefits of the program in two parts: the benefits from the realized crime reduction during the program year (i.e. the broader benefits to society from reduced crime) and—more speculatively—the future benefits from increased graduation. Since costs are more natural to think of in per participant terms (rather than per randomly-assigned youth), the table shows instrumental variables estimates, with monetary estimates of program benefits as dependent variables for youth who participate in the program. We note that our benefit-cost estimates may be conservative as we only calculate crime benefits to potential victims of crime and to the government, and not to youth arrestees who themselves may be redirected from the criminal justice system.

There are some unavoidable sources of uncertainty and measurement error in the data we have available to estimate the dollar-value benefits of BAM, and so our approach is to provide reasonable upper and lower-end estimates (the left and right columns of Table 7). This approach allows us to estimate upper and lower bound estimates for the benefits of BAM to society through both reduced crime as well as increased returns to education among participants and estimate an upper and lower bound for the benefits to BAM on crime and education combined. The difference between our high and low estimates are described below.

To calculate the broader benefits to society from reduced crime, we use the same basic framework as Heller, et al. (2017). For our low-end estimates we use costs of crime derived from jury awards as described in Miller, Cohen and Wiersema (1996). For our high-end estimates we use costs of crime based on contingent valuation surveys of the public's willingness to pay for reductions in crime (Cohen, et al. 2004). These two approaches provide *ex post* and *ex ante* estimates, respectively, the merits of which are described in greater detail in Heller, et al. (2017). The only crime for which we do not follow the framework laid out in Heller, et al. (2017) is homicide. Given the relative infrequency of homicide as well as the wide range in estimates of its cost we are wary of the high variance homicide arrests can impose on our estimates. As such, we fix the cost of homicide at the median estimated value of statistical life (VSL) (Chalfin and McCrary 2017) for both our high and low estimates. Finally, for each arrest we also assign a cost to government which incorporates the cost of booking, arraignment, etc. At the low end, we assume this cost to be $5,770 and at the high end we assume it to be $6,524 (2010 dollars), as described in Heller, et al. (2017).

The bottom panel of Table 7 estimates the potential future benefits from increased high school graduation. To calculate these benefits we use U.S. Census Bureau estimates for lifetime earnings among men by race group and educational attainment. In order to capture both the increase in wages associated with high school graduation and the increased likelihood of employment we calculate the difference in estimated lifetime earnings among by-race high school dropouts and high school graduates as a function of students' likelihood of graduation. We compute this benefit as a discounted sum over a 40-year career beginning five years after

BAM delivery. For our low-end estimates we assume an annual discount rate of 5% on the present value of money, while for our high-end estimates we assume a discount rate of 3%. We also assign a lifetime benefit from increased health outcomes at $13,500 and $44,000 (2010 dollars) as a function of the likelihood of high school graduation based on estimates from Cutler and Lleras-Muney (2008). For each of the low and high-end estimates we assign an undiscounted cost of an additional year of schooling equal to $7,946 (2010 dollars) as a function of their likelihood of graduation, (Illinois State Board of Education 2015).
Combining both the realized crime benefits with the speculative graduation benefits we find suggestive evidence of positive returns from participation in BAM. For our average effect outcome period we find benefit-cost ratios of 2:1 on the low-end and 6:1 on the high-end. For the effect at program completion our estimates suggest even bigger returns from 5:1 at the low-end to 10:1 at the high-end. Although none of these estimates are statistically significant they are consistent in direction and suggest moderate to large returns on investment. We likely do not find statistical significance due to a high degree of sensitivity in the standard errors driven by the measurement of the value of statistical life used to estimate the cost of homicide (consistent with the fact that the estimated gains from improved schooling outcomes alone are significant, but not for the gains from reduced crime or overall). However, the point estimates themselves, which reflect our "best estimate," are all in the direction of favorable benefit-cost ratios for the BAM program. These benefit-cost ratios may be conservative as well, since we do not include direct benefits to youth who are arrested but may be redirected from the criminal justice system (i.e., do not become incarcerated).

The positive overall benefit-cost ratios of the program suggest that BAM may be a worthwhile intervention to improve the life outcomes of at-risk male students. The benefit-cost ratios that we find are comparable to other school-based interventions that have been found effective. The Washington State Institute for Public Policy (WSIPP), for example, estimates that the benefit-cost ratio of school-based mentoring programs to be about 11:1, and of community-based mentoring programs to be about 2.5:1.[28] When looking at the effect of mentoring programs on higher-risk youth in the juvenile justice system, we find WSIPP estimates benefit-cost ratios of approximately 4:1.[29] Thus, the benefit-cost ratios here are comparable to other mentoring programs that have been found to have positive effects.[30]

## X. Results by Study

Though pooling the different studies together yields results that point in the direction of positive impacts, there is variability in study impacts across our RCTs. The variability shown across outcomes, program definitions, and samples in our pooled estimates is important to understand as Becoming a Man continues to grow and operate at a larger scale. To do so, we first show results for each of the four BAM RCTs separately. Tables 8-11 summarize the results of each of the studies separately. As a whole, we find that our BAM 1 and BAM 2 studies show consistently positive effects, while the BAM expansion study shows more mixed effects, and the BAM 2x2 study shows signs of null-to-adverse effects.

[28] See: http://www.wsipp.wa.gov/BenefitCost/Program/764 and http://www.wsipp.wa.gov/BenefitCost/Program/767
[29] See: http://www.wsipp.wa.gov/BenefitCost/Program/369.
[30] We note that the Washington State Institute of Public Policy's cost-benefit framework, though similar, is not the exact methodology used by our research team.

Tables 8 and 9 show that there are large beneficial impacts in our BAM 1 and BAM 2 studies, which are quite similar in magnitude. We find a statistically significant impact on school engagement by 0.14 standard deviations (which we estimate will translate into an increase in graduation rates of 9.7%) and a decrease in violent crime arrests by 45% in our BAM 1 study, and improvements in school engagement and violent crime arrests that vary in impact and statistical significance for our BAM 2 study depending on the model used. We also see reductions in all arrests and "other" arrests. In contrast, the estimated effects in the BAM expansion RCT are mixed: the results are generally not statistically significant but they are large in magnitude, with an adverse effect on school engagement and sizable proportional reductions in every arrest category except for drug offenses (Table 11). And for the BAM 2x2 study (Table 10) there are if anything consistently adverse effects on both school engagement and arrests.

Figures 1-12 present the plausible range of coefficient estimates for each study outcome for our average effect and program completion models using a slightly different methodology—sampling using a bootstrap approach.[31] (Figures for our first-year and second-year models can be found in Appendix Figures A1-A12). These figures show that while statistical significance varies across studies and outcomes, there is general consistency in the magnitude of estimated coefficients in the direction of beneficial impacts. While the results are generally consistent across studies, there are still some outliers—notably for the BAM 2x2 study—that require further examination.

An example of this can be seen with Figures 3 and 4, which present results on arrests to youth for all crimes. Whether we measure outcomes averaged over all program years, or focus on outcomes measured at the end of the program participation period, effects of BAM participation are typically in the range of reductions in arrests of between -9% to -34%. The exception is the BAM 2x2 study, where we see sizable adverse effects – but also a sizable confidence interval. The magnitude is itself somewhat sensitive to when we measure program impacts (comparing across Figures 3 versus 4), which highlights that the estimate itself is being driven in part by a few schools that have large, outlier-type impacts in selected years.

Understanding the source of this variability in estimated program impacts is a top priority for our future work. Initial explorations of our data suggest the answer does not appear to rest with differences in the baseline characteristics of students enrolled in these studies, even though our BAM 1 and BAM 2 have a higher proportion of students that have been arrested at baseline in comparison to our BAM 2x2 and BAM expansion samples. Though difficult to test in our data, additional hypotheses for the source of variability include differences in the schools or time periods in which the RCTs were carried out. We have also, and will continue to, assemble data on program implementation across studies to better understand what role that may have played in affecting program impacts. Lastly, we will continue to analyze how school quality and school supports across our study schools could have affected program impacts, as selection of schools across the four studies was not random. More specifically, the research team is gathering all available data to see whether there is variation in other supports provided to control group

[31] For these figures, we use a bootstrap approach, sampling with replacement to generate a thousand estimates for each model of interest. We take the 2.5th and 97.5th percentile estimates as the confidence interval, and the 50th percentile as the median listed estimate.

students across the schools that could have mitigated the effects of BAM, as the schools in the BAM 2x2 study may have qualitatively been more organized and were in lower-crime neighborhoods in comparison to the schools in the other studies. Our analyses at this point in time do not point to any specific implementation or school quality criteria that can explain the variability across studies.

In the meantime we explore below one candidate hypothesis for which we have data available at this point: differential impacts on mechanisms between studies. For this purpose we focus on contrasting the effects of the BAM 2x2 study and the BAM 2 study, since these are the two RCTs for which we have consistent measures of candidate mediating mechanisms (and also were carried out at the same point in calendar time).

## XI. Mechanisms Survey Results

Tables 12 and 13 present results from wave 2 of our ISR mechanisms survey for the two BAM RCTs that were carried out at the same point in calendar time: the BAM 2 study and the BAM 2x2 study.[32] The purpose of our mechanisms survey was to provide insight into the various candidate mechanisms of action through which BAM might affect youth outcomes. This survey was implemented after year 2 of the study, from November 2015 through July 2016.

We present results by aggregating our questions into indices of eighteen key mechanisms that our ISR surveys were designed to measure:

- automaticity
- self distancing
- grit
- sensation seeking
- peer conflict
- how I think
- conscientiousness
- education and schooling
- social networks
- adult supports
- growth mindset
- subjective expectations
- mental health
- crime victimization

[32] Results from our 2014 wave of ISR surveys can be found in Appendix Table A4. As our first wave of ISR surveys was only carried out with a subset of BAM 2x2 students, we can use information from this wave to further study candidate mechanisms of action, but cannot use this information to look at how differences in effects on candidate mechanisms may be driving variation in program impacts. For our 2014 survey wave, we present results of fifteen indices: automaticity, self distancing, awareness of past action, locus of control, grit, conscientiousness, peer conflict, social networks, education and schooling, and adult supports. We note that the direction of change that would be consistent with a beneficial impact on behavioral outcomes is in the positive direction, with the exception of the peer conflict and self distancing measures, where the desirable effect is in the negative direction. Appendix Table A5 provides a list of the questions that were used in each index for our 2014 ISR survey.

- risky behavior.

Each of these indices is presented in Z-score form. Appendix Table A6 provides a list of the questions that were included in each index for our second wave of surveys. We note that the direction of change that would be consistent with a beneficial impact on behavioral outcomes is in the positive direction, with the exception of the peer conflict, conscientiousness, and crime victimization measures, where the desirable effect is in the negative direction.

Survey results indicate that effects are not statistically significant and have standard errors that are quite sizeable. Our theory had predicted that the automaticity and self-distancing measures should be most strongly affected by BAM, though there is no stronger evidence for them than the others. We also see that there is not a clear pattern that emerges when trying to find differences in candidate mechanisms of action between the BAM 2 and BAM 2x2 studies.

## XII. Mechanisms Experiment Results

It is possible that the automaticity and self-distancing measures do not reveal any BAM impact because they rely on self-reports about past behavior or how the youth would respond in a hypothetical scenario, rather than reflecting actual behavior and decisions made in the moment.

To address this measurement challenge, Table 14 provides results from a mechanisms experiment we conducted with a subset of our BAM 2 study participants in AY2013-14, after the first year of program participation. The purpose of this experiment was to examine how BAM changes decision-making in confrontational situations where youth are provoked and retaliation is a possibility. With this experiment we directly tested this hypothesis of automaticity to see whether BAM causes youth to “slow down” and make more deliberate decisions.

For this experiment we had participants play a modified version of a real-stakes iterated dictator game. Students were informed during lunch periods that a brief study would be conducted giving them the chance to earn about $10. Participants were led through the study by a research assistant (RA) who was blinded to youth treatment status. RAs told participants that they would be communicating over walkie-talkie with another RA who was standing with their “partner.” However, there was no partner; the other RA was actually a confederate who followed a script.[33]

In the first round, participants were given $10 in one-dollar bills in an envelope. Their “partner” was given the chance to take some money away from the participant. The participants heard the confederate say over the walkie-talkie that the “partner” was taking $6 from the participant. The participant was then asked how much money they would like to take from their partner. (So for participating in the decision-making exercise each participant received $4 from the first round plus whatever they took in the second round.)

The automaticity hypothesis predicts that participants who had previously been assigned to BAM would make slower, more deliberate decisions than participants who had been assigned to the

[33] Experimental economists normally, and understandably, seek to avoid use of deception in experiments. However, an overarching concern of the CPS Research Review Board (RRB) one was that we may contribute to antagonism between students if two actual students were playing against each other and taking money from another. Given the RRB’s human subjects concerns, the research team decided that deception was necessary.

control conditions.[34] As we were also interested in testing whether actively trying to reduce automaticity during the decision-making exercise itself could attenuate the BAM-control difference in decision-making, we randomized participants to four different versions of the task:

- A "no delay" condition, in which youth could say how much they wanted to take from their partner as soon as they wished after the partner's take amount was announced.
- A "distraction" condition, which was intended to get all youth (including controls) to do part of what we believe BAM gets youth to initiate on their own—which is to slow down. In this condition, after round 1 the participants were told to first spend 30 seconds completing a word-search puzzle and to then state how much money they wished to take.
- A "reflection" condition, where they were told to first take 30 seconds to rate their partner's action on a scale from -5 (extremely selfish) to +5 (extremely generous) before deciding how much to take from the partner.
- A "rumination" condition that got youth to slow down but then, instead of reflecting and taking a different perspective on the event, they were given an exercise intended to promote unhelpful thinking (rumination). Specifically they were told to take 30 seconds to read over a list of adjectives and to circle the ones that represented their feelings in that moment, with the word list including terms like rude, unfriendly, mean, and unkind.

If our automaticity hypothesis is correct, BAM should get youth to slow down on their own and reflect on what their optimal response would be. We should see this most clearly in condition 1 ("no delay"). We expected to see a smaller BAM-control difference when we externally induced both groups to slow down (as in condition 2), and a still smaller difference when we induced youth in both groups to both slow down and reflect on the nature of their partner's decision (as in condition 3). We also expected condition 4 to attenuate the BAM-control difference by prompting both groups to ruminate on how they feel, which may divert the BAM youth from the tendency to reflect on the situation.

Unfortunately randomization across conditions did not work quite as well as we had hoped, yielding some imbalance in baseline attributes. However within conditions there was baseline balance for BAM versus control. So we can learn about the role of the conditions from the difference-in-difference (how outcomes for BAM versus control differed across conditions).

Table 14 shows that BAM did indeed get youth to slow down before they made a decision. We had the RAs who were working with participants subtly time how long it took youth to respond. The variable is very skewed, so we report results for the log of the time it took youth to respond.

---

[34] Previous research in psychology suggests that *automaticity* plays a large role in people's decision-making (including decisions about dropping out, becoming involved with drugs or gangs, or how to respond to confrontations that could escalate into serious violence). The research team believes that BAM program builds on this idea that because deliberate decision-making and conscious cognition (what psychologists call "system 2" thinking) require effort, people rely heavily on automatic responses that are adaptive to community encountered situations ("system 1" thinking). Problems can arise for youth in disadvantaged neighborhoods where high-stakes situations occur frequently and where being aware of the dangers of automatic thinking can be the difference between life and death. For more information on our automaticity hypothesis, please see our February 2017 paper in the *Quarterly Journal of Economics*, titled "Thinking Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago".

The second row shows that in condition 1 (where our automaticity theory makes a clean prediction that BAM should generate more "slowing down" versus controls) the average control complier took 1.1 seconds to decide. The coefficient on BAM participation in our log-linear specification is 0.62, which implies a statistically significant increase of roughly 86% in the time that youth took to decide.

If our automaticity theory is correct, and BAM causes participants to slow down their thinking and be more reflective, then in conditions 2–4 (which try to even out the difference in those tendencies between BAM and control groups) we should see smaller BAM effects on response times compared to what we see in the first condition. In fact, that is what we find—the effect of BAM was less than half of condition 1 in conditions 2–4. The other rows of the table show that by prompting all youth to do what we believe BAM gets youth to do on their own (slow down and reflect), the distraction, reflection, and even rumination conditions succeeded in narrowing the BAM-control difference in the tendency to slow down and be less automatic when deciding by how much to retaliate.

The automaticity theory does not make any clear prediction about whether BAM youth should actually retaliate less than controls in this iterated dictator game. BAM never tells youth not to fight or retaliate when provoked, since the program recognizes that in the neighborhoods where these youth are growing up there are indeed circumstances in which fighting and an aggressive response may be (unfortunately) necessary and adaptive. The focus of the program instead is to get youth to slow down and reflect on what sort of response is most adaptive for the circumstance they are facing. Consistent with this focus of the program, Table 14 shows that we found no evidence that BAM reduces the retaliation amount. In addition, the fact that we see no change in the average retaliation amount between BAM students and control students argues against the notion that program impacts are driven by improvements in pro-social behaviors as a whole, and instead provides evidence that program impacts are driven by changes in automaticity.

## XIII. Discussion and Conclusion

One key finding from our analysis is that the size and statistical significance of the estimated impacts from the BAM program vary across the four different RCTs we have carried out. As a result the overall estimate for what BAM accomplishes can be somewhat sensitive to how we aggregate information across studies.

The research team will continue to explore causes of the underlying variability between studies. One hypothesis is that the baseline characteristics of students between the studies could explain this variation, as fewer students in the BAM 2x2 and BAM expansion studies had been arrested at baseline in comparison to the BAM 1 and BAM 2 studies. However, initial explorations of our data suggest the answer does not appear to rest with differences in the baseline characteristics of students. A second hypothesis is that differences in implementation between sites are driving variation, specifically in our BAM 2 and BAM 2x2 studies which occurred during the same period of time but show almost opposite program effects. The research team is analyzing implementation data from the BAM 2 and BAM 2x2 studies to see whether differences in implementation between both studies could explain variation, though analyses at this time do not point to any one implementation criteria that can explain variation in program effects. A third candidate hypothesis is that there are differences in the schools and school support systems between the BAM 2 and 2x2 schools, as the BAM 2x2 schools may have been qualitatively more organized and could perhaps have provided additional programming to control group students; these schools also tended to be in lower-crime neighborhoods than the BAM 2 schools. The research team is compiling all available information on school quality and school supports in order to test this hypothesis. A last candidate hypothesis is differences in candidate mechanisms of action through which BAM operates, of which the research team will continue to analyze data from the mechanisms survey. Overall, though the BAM program seems likely to improve youth outcomes when pooling all of our study samples, the puzzle that remains is for whom the program is most effective, in what contexts and why. In any case future work by the research team will seek to disentangle the differential effects between studies, with the ultimate goal of determining the conditions and study participants for which the program seems to work best. The research team plans to release a working paper with these additional analyses and updated results in 2018.

On the whole, our estimates are in the direction of suggesting net benefits to youth from participating in BAM. Our results that pool data from all four RCTs show that the program reduces violent crime by 19-37%, and improves school engagement by 0.03-0.08 standard deviations (which we estimate will translate to improved high school graduation rates of between 2.0% and 5.1%), although the statistical significance of these impacts depends on how we aggregate information from the different studies and at what point in time we measure impacts. We also find that the program has a benefit-cost ratio of between 2:1 and 10:1. While the statistical uncertainty interval around these estimates can also be somewhat large, our best estimate is in the direction of BAM having a favorable benefit-cost ratio.

We are grateful to Youth Guidance, the Edna McConnell Clark Foundation and the Corporation for National and Community Service for their continued partnership and support of this research. We are eager to continue to learn about how these and other strategies work to reduce youth violence and, more broadly, to improve the lives of at-risk youth.

## XIV. References


Anderson, E. (1999). *Code of the Streets*. New York, NY: Norton.

Anderson, M.L. (2008). Multiple inference and gender differences in the effects of early intervention: reevaluation of the Abecedarian Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association,* 103(2008), 1481-95.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444-455.

Beck, J. (2011). *Cognitive Therapy: Basics and Beyond*. New York, NY: The Guilford Press.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, Methodological,* 289-300.

Benjamini, Y., Krieger, A.M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93, 491-507.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation review*, 8(2), 225-246.

Borghans, L., Duckworth, A.L., Heckman, J.J., & Ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources,* 43(4), 972-1059.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, 39(4), 1137-1176.

Carneiro, P. & Heckman, J. (2003). Human Capital Policy. Cambridge, MA: National Bureau of Economic Research, Working Paper No. 9495.

Chalfin, A. & McCrary, J. (2017). Are U.S. Cities Underpoliced?: Theory and Evidence. *The Review of Economics and Statistics*.

Cohen M., Rust, R., Steen, S., & Tidd, S. (2004). Willingness-to-Pay for Crime Control Programs. *Criminology*, 42(1), 89-110.

Cook, P., Dodge, K., Farkas, G., Fryer, R., Guryan, J., Ludwig, J., Mayer, S. (2015). Not Too Late: Improving Academic Outcomes for Disadvantaged Youth. *Northwestern University Institute for Policy Research Working Paper Series.*

Cunha, F. & Heckman, J.J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31-47.

Cutler, D.M. & Lleras-Muney, A. (2008). Chapter 2: Education and Health: Evaluating

Theories and Evidence," in *Making Americans Healthier: Social and Economic Policy as Health Policy*. New York: Russell Sage Foundation.

Deming, D. J. (2011). Better Schools, Less Crime?. *The Quarterly Journal of Economics*, 126 (4), 2063–2115.

Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Weissberg, R. P., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-32.

Evans, W.N., & Owens, E.G. (2007). COPS and Crime. *Journal of Public Economics*, 91(1-2), 181-201.

Goldin, C. & Katz, L.F. (2008). *The Race between Education and Technology.* Cambridge, MA: Harvard University Press.

Hammond, W. R. & Yung, B. (1991). Preventing Violence in At-Risk African-American Youth. *Journal of Health Care for the Poor and Underserved*, 2(3), 359-373.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451-464.

Heckman, J. J., & LaFontaine, P. A. (2010). The American High School Graduation Rate: Trends and Levels. *Review of Economics and Statistics*, 92(2), 244-262.

Heckman, J. J., & Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2), 145-149.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411-482.

Heller, S. B., Pollack, H. A., Ander, R., & Ludwig, J. (2013). Preventing youth violence and dropout: A randomized field experiment. Cambridge, MA: National Bureau of Economic Research, Working Paper No. 19014.

Heller, S.B., Shah, A.K., Guryan, J., Ludwig, J., Mullainathan, S. & Pollack, H. (2017). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *Quarterly Journal of Economics*, 132(1), 1-54.

Illinois State Board of Education, "City of Chicago Sd 299 Per Student Spending," (https://illinoisreportcard.com/District.aspx?source=Environment&source2=PerStudentSpending&Districtid=15016299025: Illinois Report Card, 2014-2015, 2015).

Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, 82(2), 533-554.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Katz, L.F., Kling, J.R., and Liebman, J.B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics*, 116(2), 607-654.

Kling, J.R., Liebman, J.B., & Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1), 83-119.

Kling, J. R., Ludwig, J. & Katz, L.F. (2005). Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. *Quarterly Journal of Economics,* 120(1), 87-130.

Little, R.J. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data, 2nd Edition.* Wiley Series in Probability and Statistics.

Lleras-Muney, A. (2005). The Relationship between Education and Adult Mortality in the United States. *Review of Economic Studies*, 72(1), 189-221.

Lochner, L. & Moretti, E. (2004). The Effect Of Education On Crime: Evidence From Prison Inmates, Arrests, And Self-Reports. *American Economic Review*, 94(1), 155-189.

Ludwig, J., Duncan, G.J., Gennetian, L.A., Katz, L.F., Kessler, R.C., Kling, J.R., Sanbonmatsu, L. (2012). Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science*, 337(6101) 1505-1510.

Miller, T., Cohen, A., & Wiersema, B. (1996). *Victim Costs and Consequences: A New Look*. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698.

Monahan, K.C., Steinberg, L., Cauffman, E., & Mulvey, E.P. (2009). Trajectories of antisocial behavior and psychosocial maturity from adolescence to young adulthood. *Developmental psychology*, 45(6), 1654.

Murnane, R. J. (2013). U.S. high school graduation rates: Patterns and explanations. Cambridge, MA: National Bureau of Economic Research, Working Paper No. 18701.

Papachristos, A.V. (2009). Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide. *American Journal of Sociology*, 115(1): 74-128.

Puma, M.J., Olsen, R.B., Bell, S.H., & Price, C. (2009). What to Do when Data are Missing in Group Randomized Controlled Trials. NCEE 2009-0049. National Center for Education Evaluation and Regional Assistance, 131.

Reardon, S. F. (2011). “The widening academic achievement gap between the rich and the poor: New evidence and possible explanations.” In *Whither Opportunity? Rising Inequality, Schools, and Children’s Life Chances*, Eds. Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation Press. pp. 91-116.

Schochet, P.Z., Burghardt, J., McConnell, S. (2008). Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review*, 98(5), 1864-86.

Thaler, R.H. & Sunstein, C.R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. New York, NY: Penguin Books.

Tough, P. (2012). *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. New York: Houghton Mifflin Harcourt.

Weiner, D.A., Lutz, B.F. & Ludwig, J. (2009). The Effects of School Desegregation on Crime. Cambridge, MA: National Bureau of Economic Research, Working Paper No. 15380.

Westfall, P. & Young, S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. New York: John Wiley and Sons.

**Table 1: Select BAM Activities**

| Activity Category | Example Activities |
|---|---|
| Reflective/ Introspective | **Check-Ins:** Students talk to each other about what they are doing well and areas where they still need to improve. Students must listen patiently while someone else discusses their attributes. |
| Immersive/ Experiential | **The Fist:** Students are told to get an object from a partner. Many try to use force. The counselor asks questions to highlight how their partners were willing to give up the object if they calmly requested it. |
| | **Plates:** Students reflect on what it has taken to successfully complete group missions and write those attributes on a plate. The plates are placed on the floor, and students must cross the floor by using the plates. However, if no one is standing on a plate, then it is removed (making the task more difficult). |
| | **Trust Walk:** Students follow group leaders around the school silently and without disrupting the school. They are told that with freedom comes responsibility. |
| | **Focus Mitt Drill:** Students punch focus mitts for an extended period. |
| | **Human Knot:** Students stand in a circle and grab the hands of someone standing across from them. They must then untangle themselves without letting go. |
| Role Playing | **$10 Role Play:** Students role play a student borrowing money and then never paying it back. |
| | **High School Day:** Students do a role-play where a student and administrator have a confrontation. They act out the conflict with "out of control" and "in control" anger expressions. |
| | **Our Story Of What Happened:** Students imagine a conflict and discuss why the conflict came about. They examine thinking distortions that might have made the conflict worse. |
| Stories & Discussion | **Rudy:** Students watch and discuss the movie Rudy. Before beginning the movie, the counselor holds up two dollars and asks who wants the money. Even as students raise their hand, he keeps asking who wants it until someone simply takes it from him. He explains that we often overlook opportunities, but the student who took the money saw it as an opportunity and took a chance. |
| | **The Boy Who Cried Wolf:** Students listen to and discuss the story where one day a boy pretends that he is being attacked by a wolf. He is amused by how his town responds to this prank. So when he feels bored on another day, he does it again. And again. He promises to stop playing around, but when he feels bored he can't help but do it again. In the end, when he is actually attacked by a wolf, no one responds to his pleas for help. |
| | **Miracle:** Students watch and discuss the film Miracle about the U.S. men's hockey team. |
| Skill-building | **Cognitive Thought Replacement:** Students learn how to recognize negative thoughts that arise and how to replace them. It is not necessary to replace negative thoughts with positive thoughts, but rather to instead focus on what can be done to control the situation that is leading to the negative thought. |
| | **Manhood Questions and Rites of Passage:** Students discuss the key moments when boys become men and various rites of passage that exist. |
| | **Positive Anger Expression:** Students are taught about how to express anger in a controlled way. |

Table 3: Becoming a Man Studies – Baseline Characteristics

| | Study 1 | | Study 2 | | Study 2x2 | | Study Expansion | |
|---|---|---|---|---|---|---|---|---|
| | Control | Treatment | Control | Treatment | Control | Treatment | Control | Treatment |
| Number of Students | 1267 | 1473 | 1048 | 1016 | 1453 | 1180 | 1282 | 1085 |
| Demographics | | | | | | | | |
| Black | 0.720 | 0.688 | 0.698 | 0.683 | 0.456 | 0.479 | 0.632 | 0.610 |
| Hispanic | 0.276 | 0.307 | 0.275 | 0.300 | 0.493 | 0.464 | 0.325 | 0.357 |
| English language learner | | | 0.050 | 0.053 | 0.116 | 0.116 | 0.079 | 0.105 |
| Age | 15.700 | 15.512 | 14.845 | 14.910 | 14.802 | 14.811 | 14.484 | 14.457 |
| Free lunch recipient | 0.899 | 0.910 | 0.845 | 0.834 | 0.866 | 0.869 | 0.885 | 0.895 |
| Learning disability | 0.198 | 0.186 | 0.168 | 0.164 | 0.178 | 0.162 | 0.178 | 0.163 |
| Schooling | | | | | | | | |
| Grade 9 | 0.455 | 0.450 | 0.597 | 0.543 | 0.552 | 0.580 | 0.864 | 0.899 |
| Grade 10 | 0.493 | 0.445 | 0.395 | 0.448 | 0.434 | 0.410 | 0.121 | 0.089 |
| GPA | 1.679 | 1.734 | 2.111 | 2.157 | 2.120 | 2.068 | 2.317 | 2.260 |
| Crime | | | | | | | | |
| Any arrests at baseline | 0.369 | 0.346 | 0.230 | 0.232 | 0.182 | 0.188 | 0.186 | 0.186 |
| Number of baseline arrests for: | | | | | | | | |
| Violent offenses | 0.353 | 0.348 | 0.185 | 0.184 | 0.138 | 0.132 | 0.139 | 0.153 |
| Property offenses | 0.206 | 0.189 | 0.138 | 0.129 | 0.081 | 0.090 | 0.089 | 0.102 |
| Drug offenses | 0.168 | 0.177 | 0.111 | 0.144 | 0.069 | 0.064 | 0.073 | 0.061 |
| Other offenses | 0.449 | 0.470 | 0.290 | 0.321 | 0.208 | 0.238 | 0.234 | 0.234 |

Note: P-value on F-test of treatment-control comparison for all baseline characteristics: BAM 1: p=0.668; BAM 2: p=0.499; BAM 2x2: p=0.882; BAM Expansion: p=0.717. Joint significance tests for equality of all baseline characteristics use only non-missing data. Grade level measured at start of study. Asterisks indicate statistical significance of pairwise treatment-control comparison for a given baseline characteristic controlling for randomization block fixed effects with heteroscedasticity-robust standard errors. Data from Chicago Public Schools administrative data and Chicago Police Department arrest records. Means calculated using non-missing observations for each variable. Pre-program arrests are arrests prior to start of program school year. GPA is measured on a 0-4 scale. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 6: Becoming a Man Pooled Studies - Effect on Youth Outcomes**

| Model | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **School Engagement Index** | | | | **All Arrests** | | | | **Violent Arrests** | | | |
| First-year effect | 9307 | 0 | 0.0128 | 0.0301 | 0.2607 | 0.509 | -0.0446** | -0.1053** | 0.4977 | 0.1122 | -0.0199*** | -0.0470*** | 0.1279 |
| | | | (0.0122) | (0.0286) | | | (0.0224) | (0.0526) | | | (0.0076) | (0.0179) | |
| Second-year effect | 4174 | 0 | 0.0175 | 0.0524 | 0.2921 | 0.4348 | 0.0106 | 0.0317 | 0.2370 | 0.0793 | -0.0051 | -0.0154 | 0.0824 |
| | | | (0.0181) | (0.0532) | | | (0.0334) | (0.0989) | | | (0.0106) | (0.0313) | |
| Average effect | 9307 | -0.0031 | 0.0152 | 0.0358 | 0.2281 | 0.5052 | -0.0341 | -0.0803 | 0.4871 | 0.1091 | -0.0163** | -0.0385** | 0.1214 |
| | | | (0.0115) | (0.0267) | | | (0.0209) | (0.0489) | | | (0.0071) | (0.0165) | |
| Measured at program completion | 6914 | 0 | 0.0359*** | 0.0785*** | 0.1723 | 0.5306 | -0.0152 | -0.0331 | 0.4849 | 0.1109 | -0.0131 | -0.0286 | 0.1157 |
| | | | (0.0139) | (0.0300) | | | (0.0273) | (0.0591) | | | (0.0092) | (0.0200) | |
| | | **Property Arrests** | | | | **Drug Arrests** | | | | **Other Arrests** | | | |
| First-year effect | 9307 | 0.064 | 0.0023 | 0.0054 | 0.0465 | 0.0989 | -0.0001 | -0.0001 | 0.0914 | 0.2339 | -0.0269** | -0.0636** | 0.232 |
| | | | (0.0063) | (0.0148) | | | (0.0092) | (0.0216) | | | (0.0132) | (0.0309) | |
| Second-year effect | 4174 | 0.0506 | 0.0035 | 0.0105 | 0.0270 | 0.1035 | 0.0072 | 0.0215 | 0.0612 | 0.2013 | 0.0050 | 0.0151 | 0.0663 |
| | | | (0.0086) | (0.0254) | | | (0.0143) | (0.0425) | | | (0.0190) | (0.0563) | |
| Average effect | 9307 | 0.0619 | 0.0035 | 0.0082 | 0.0499 | 0.1020 | 0.0023 | 0.0054 | 0.0914 | 0.2322 | -0.0236* | -0.0555* | 0.2244 |
| | | | (0.0058) | (0.0135) | | | (0.0082) | (0.0192) | | | (0.0122) | (0.0284) | |
| Measured at program completion | 6914 | 0.06 | 0.0069 | 0.0151 | 0.0562 | 0.1206 | 0.0073 | 0.0159 | 0.0968 | 0.2390 | -0.0163 | -0.0355 | 0.2161 |
| | | | (0.0072) | (0.0156) | | | (0.0112) | (0.0242) | | | (0.0158) | (0.0342) | |

Note: Table presents four different methods of measuring the effect of BAM across a pooled sample of all four studies. The first-year effect measures outcomes from the first year of all studies. The second-year effect measures outcomes from the second year of Study 2 and Study 2x2. The average effect measures outcomes from the first year of Study 1 and Study Expansion and an average of outcomes from the first and second years of Study 2 and Study 2x2. The program completion effect measures outcomes from the first year of Study 1 and the second year of Study 2 and Study 2x2. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses clustered on individuals to account for students who were randomized into more than one study sample. * p<0.10, ** p<0.05, *** p<0.01.

**Table 7: Becoming a Man Pooled Studies: Estimated Benefits Per Participant**

| | *Average Effect (N=9,307)* | | *Program Completion Effect (N=6,914)* | |
|---|---|---|---|---|
| | Low-end Estimate | High-end Estimate | Low-end Estimate | High-end Estimate |
| | *From Crime Reduction* | | | |
| Savings to Potential Victims | -1712 | 2916 | 2253 | 6447 |
| | (8972.583) | (9193.374) | (12186.42) | (12407.65) |
| Savings to Government | 451 | 510 | 242 | 274 |
| | (280.8456) | (317.5454) | (336.618) | (380.6058) |
| Subtotal | -1261 | 3426 | 2495 | 6721 |
| | (8986.673) | (9241.91) | (12205.25) | (12462.48) |
| | *From Increased High School Graduation* | | | |
| Earnings Increase to Participant | 5102** | 7918** | 7854*** | 12234*** |
| | (2499.338) | (3926.671) | (2752.063) | (4331.411) |
| Cost of Additional Schooling | -162 | -162 | -263** | -263** |
| | (105.5172) | (105.5172) | (118.1357) | (118.1357) |
| Subtotal | 4940** | 7756** | 7591*** | 11971*** |
| | (2405.924) | (3831.292) | (2646.654) | (4223.899) |
| | *Total* | | | |
| | 3679 | 11182 | 10086 | 18692 |
| | (9339.146) | (10118.75) | (12485.36) | (13227.49) |
| Costs per participant | $1,850 | $1,850 | $1,850 | $1,850 |
| Benefits/Costs | 2/1 | 6/1 | 5/1 | 10/1 |

Note: Table presents instrumental variable estimates for the benefits per participant of the Becoming A Man program among the full pooled sample of students from the BAM 1, BAM 2, BAM 2x2, and BAM Expansion studies. Benefits to society from crime reduction are based on observed arrests among study population while earnings benefits from education are a function of predicted high school graduation. Estimates are presented in 2017 dollars with standard errors in parentheses. All models include standard baseline covariates, randomization block-fixed effects, and student-clustered standard errors. See text for full description of benefit-cost assumptions and sources. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

**Table 8: Becoming a Man Study 1 - Effect on Youth Outcomes**

| Model | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **School Engagement Index** | | | | **All Arrests** | | | | **Violent Arrests** | | |
| First-year effect | 2740 | 0 | 0.0584*** | 0.1400*** | 0.2184 | 0.6993 | -0.0744 | -0.1784 | 0.6640 | 0.1665 | -0.0346** | -0.0829** | 0.1860 |
| | | | (0.0214) | (0.0508) | | | (0.0472) | (0.1123) | | | (0.0166) | (0.0395) | |
| | | | **Property Arrests** | | | | **Drug Arrests** | | | | **Other Arrests** | | |
| First-year effect | 2740 | 0.0766 | 0.0075 | 0.0180 | 0.0599 | 0.1507 | -0.0001 | -0.0003 | 0.1003 | 0.3054 | -0.0472* | -0.1132* | 0.3179 |
| | | | (0.0128) | (0.0305) | | | (0.0183) | (0.0434) | | | (0.0278) | (0.0663) | |

Note: Table presents estimates for the effect of BAM in Study 1. The first-year effect measures outcomes from the first year; participation for the IV is attending at least one session during the first year. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

**Table 9: Becoming a Man Study 2 - Effect on Youth Outcomes**

| Model | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **School Engagement Index** | | | | **All Arrests** | | | | **Violent Arrests** | | | |
| First-year effect | 2064 | 0 | 0.0148 | 0.0297 | 0.2050 | 0.5954 | -0.0921* | -0.1844* | 0.6455 | 0.1193 | -0.0223 | -0.0447 | 0.1294 |
| | | | (0.0245) | (0.0482) | | | (0.0497) | (0.0984) | | | (0.0158) | (0.0313) | |
| Second-year effect | 1872 | 0 | 0.0543** | 0.1716** | 0.1887 | 0.6197 | -0.0550 | -0.1740 | 0.5040 | 0.1165 | -0.0280 | -0.0884 | 0.1817 |
| | | | (0.0269) | (0.0833) | | | (0.0581) | (0.1812) | | | (0.0191) | (0.0599) | |
| Average effect | 2064 | -0.0140 | 0.0349* | 0.0692* | 0.1241 | 0.5825 | -0.0704 | -0.1394* | 0.6185 | 0.1150 | -0.0244* | -0.0483* | 0.1374 |
| | | | (0.0211) | (0.0409) | | | (0.0429) | (0.0841) | | | (0.0131) | (0.0256) | |
| Measured at program completion | 1872 | 0 | 0.0543** | 0.1110** | 0.0939 | 0.6197 | -0.0550 | -0.1126 | 0.6628 | 0.1165 | -0.0280 | -0.0572 | 0.1598 |
| | | | (0.0269) | (0.0541) | | | (0.0581) | (0.1174) | | | (0.0191) | (0.0386) | |
| | | **Property Arrests** | | | | **Drug Arrests** | | | | **Other Arrests** | | | |
| First-year effect | 2064 | 0.0725 | -0.0059 | -0.0119 | 0.0710 | 0.1269 | -0.0178 | -0.0356 | 0.1734 | 0.2767 | -0.0460 | -0.0921 | 0.2717 |
| | | | (0.0126) | (0.0248) | | | (0.0236) | (0.0467) | | | (0.0289) | (0.0572) | |
| Second-year effect | 1872 | 0.0716 | -0.0034 | -0.0108 | 0.0608 | 0.1474 | 0.0017 | 0.0053 | 0.0781 | 0.2842 | -0.0253 | -0.0801 | 0.1834 |
| | | | (0.0141) | (0.0440) | | | (0.0250) | (0.0778) | | | (0.0327) | (0.1018) | |
| Average effect | 2064 | 0.0687 | -0.0031 | -0.0062 | 0.0747 | 0.1307 | -0.0081 | -0.0160 | 0.1550 | 0.2681 | -0.0348 | -0.0689 | 0.2515 |
| | | | (0.0101) | (0.0197) | | | (0.0184) | (0.0359) | | | (0.0238) | (0.0466) | |
| Measured at program completion | 1872 | 0.0716 | -0.0034 | -0.0070 | 0.0899 | 0.1474 | 0.0017 | 0.0034 | 0.1494 | 0.2842 | -0.0253 | -0.0518 | 0.2636 |
| | | | (0.0141) | (0.0285) | | | (0.0250) | (0.0504) | | | (0.0327) | (0.0659) | |

Note: Table presents estimates for the effect of BAM in Study 2. The first-year effect measures outcomes from the first year; participation for the IV is attending at least one session during the first year. The second-year effect measures outcomes from the second year; participation for the IV is attending at least one session during the second year. The average effect measures the mean of outcomes from the first year and second years; for Johnson the average effect is just the first year; participation for the IV is attending at least one session during either year. The program completion effect measures outcomes from the second year; participation for the IV is attending at least one session during either year. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

**Table 10: Becoming a Man Study 2x2 - Effect on Youth Outcomes**

| Model | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **School Engagement Index** | | | | **All Arrests** | | | | **Violent Arrests** | | | |
| First-year effect | 2302 | 0 | -0.0033 | -0.0070 | 0.2792 | 0.3089 | 0.0305 | 0.0641 | 0.1754 | 0.0687 | -0.0124 | -0.0260 | 0.0668 |
| | | | (0.0255) | (0.0529) | | | (0.0362) | (0.0751) | | | (0.0122) | (0.0253) | |
| Second-year effect | 2302 | 0 | -0.0144 | -0.0412 | 0.3755 | 0.3012 | 0.0663* | 0.1902* | 0.0192 | 0.0525 | 0.0116 | 0.0334 | 0.0098 |
| | | | (0.0241) | (0.0686) | | | (0.0386) | (0.1095) | | | (0.0116) | (0.0328) | |
| Average effect | 2302 | 0 | -0.0089 | -0.0185 | 0.2256 | 0.3050 | 0.0484 | 0.1011* | 0.1776 | 0.0606 | -0.0004 | -0.0008 | 0.0452 |
| | | | (0.0217) | (0.0447) | | | (0.0297) | (0.0612) | | | (0.0093) | (0.0191) | |
| Measured at program completion | 2302 | 0 | -0.0144 | -0.0300 | 0.1754 | 0.3012 | 0.0663* | 0.1385* | 0.1717 | 0.0525 | 0.0116 | 0.0243 | 0.0240 |
| | | | (0.0241) | (0.0498) | | | (0.0386) | (0.0795) | | | (0.0116) | (0.0239) | |
| | | **Property Arrests** | | | | **Drug Arrests** | | | | **Other Arrests** | | | |
| First-year effect | 2302 | 0.0448 | 0.0091 | 0.0191 | 0.0070 | 0.0548 | 0.0071 | 0.0149 | 0.0374 | 0.1405 | 0.0267 | 0.0561 | 0.0642 |
| | | | (0.0107) | (0.0223) | | | (0.0151) | (0.0313) | | | (0.0212) | (0.0440) | |
| Second-year effect | 2302 | 0.0355 | 0.0097 | 0.0278 | -0.0012 | 0.0718 | 0.0146 | 0.0418 | 0.0389 | 0.1413 | 0.0304 | 0.0872 | -0.0283 |
| | | | (0.0101) | (0.0285) | | | (0.0170) | (0.0482) | | | (0.0220) | (0.0624) | |
| Average effect | 2302 | 0.0402 | 0.0094 | 0.0196 | 0.0217 | 0.0633 | 0.0108 | 0.0226 | 0.0502 | 0.1409 | 0.0285* | 0.0596* | 0.0605 |
| | | | (0.0081) | (0.0166) | | | (0.0124) | (0.0255) | | | (0.0173) | (0.0357) | |
| Measured at program completion | 2302 | 0.0355 | 0.0097 | 0.0202 | 0.0323 | 0.0718 | 0.0146 | 0.0304 | 0.0612 | 0.1413 | 0.0304 | 0.0635 | 0.0542 |
| | | | (0.0101) | (0.0207) | | | (0.0170) | (0.0351) | | | (0.0220) | (0.0453) | |

Note: Table presents estimates for the effect of BAM in Study 2x2. The first-year effect measures outcomes from the first year; participation for the IV is attending at least one session during the first year. The second-year effect measures outcomes from the second year; participation for the IV is attending at least one session during the second year. The average effect measures the mean of outcomes from the first year and second years; participation for the IV is attending at least one session during either year. The program completion effect measures outcomes from the second year; participation for the IV is attending at least one session during either year. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

**Table 11: Becoming a Man Study Expansion - Effect on Youth Outcomes**

| Model | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **School Engagement Index** | | | | **All Arrests** | | | | **Violent Arrests** | | | |
| First-year effect | 2201 | 0 | -0.0522** | -0.1742* | 0.4375 | 0.4489 | -0.0409 | -0.1364 | 0.4491 | 0.0956 | -0.0101 | -0.0336 | 0.1197 |
| | | | (0.0266) | (0.0892) | | | (0.0418) | (0.1370) | | | (0.0141) | (0.0462) | |
| | | **Property Arrests** | | | | **Drug Arrests** | | | | **Other Arrests** | | | |
| First-year effect | 2201 | 0.0640 | -0.0050 | -0.0168 | 0.0438 | 0.0673 | -0.0004 | -0.0015 | 0.0577 | 0.2219 | -0.0253 | -0.0845 | 0.2280 |
| | | | (0.0140) | (0.0460) | | | (0.0145) | (0.0477) | | | (0.0243) | (0.0796) | |

Note: Table presents estimates for the effect of BAM in Study Expansion. The first-year effect measures outcomes from the first year; participation for the IV is attending at least one session during the first year. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors clustered on individuals in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 12: Wave 2: Estimated Effects on Outcomes from ISR Survey, BAM 2 Study

| | N | Control Mean | Intention to Treat | Effect of Participation (IV) | Control Complier Mean | FDR Q-Value |
|---|---|---|---|---|---|---|
| Indices | | | | | | |
| Automaticity | 317 | -0.050 | 0.117 | 0.163 | -0.147 | 0.561 |
| | | | (0.088) | (0.123) | | |
| Self Distancing | 317 | -0.026 | -0.083 | -0.115 | 0.114 | 0.561 |
| | | | (0.055) | (0.076) | | |
| Grit | 317 | 0.133 | -0.075 | -0.105 | 0.165 | 0.615 |
| | | | (0.066) | (0.092) | | |
| Sensation Seeking | 317 | 0.070 | -0.06 | -0.084 | 0.135 | 0.720 |
| | | | (0.07) | (0.097) | | |
| Peer Conflict Vignette 1 (-) | 317 | 0.132 | -0.143 | -0.199 | 0.025 | 0.561 |
| | | | (0.094) | (0.131) | | |
| Peer Conflict Vignette 2 (-) | 317 | 0.043 | 0.004 | 0.006 | 0.060 | 0.995 |
| | | | (0.081) | (0.113) | | |
| Peer Conflict Vignette 3 (-) | 317 | 0.013 | 0.013 | 0.019 | -0.200 | 0.995 |
| | | | (0.09) | (0.126) | | |
| Peer Conflict Vignette 4 (-) | 317 | 0.073 | 0.07 | 0.097 | -0.070 | 0.720 |
| | | | (0.084) | (0.117) | | |
| How I Think | 317 | 0.051 | -0.105 | -0.146 | 0.154 | 0.561 |
| | | | (0.06) | (0.084) | | |
| Conscientiousness (-) | 317 | -0.167 | 0.027 | 0.038 | -0.160 | 0.995 |
| | | | (0.076) | (0.106) | | |
| Education and Schooling | 317 | -0.037 | 0.013 | 0.019 | 0.021 | 0.995 |
| | | | (0.06) | (0.083) | | |
| Social Networks | 317 | -0.017 | -0.012 | -0.016 | 0.100 | 0.995 |
| | | | (0.073) | (0.102) | | |
| Adult Supports | 317 | -0.013 | -0.031 | -0.043 | 0.067 | 0.995 |
| | | | (0.079) | (0.11) | | |
| Growth Mindset | 316 | -0.050 | 0.001 | 0.001 | -0.067 | 0.995 |
| | | | (0.086) | (0.119) | | |
| Subjective Expectations | 314 | -0.033 | -0.008 | -0.011 | -0.061 | 0.995 |
| | | | (0.069) | (0.095) | | |
| Mental Health | 317 | 0.036 | -0.086 | -0.12 | 0.155 | 0.561 |
| | | | (0.061) | (0.085) | | |
| Crime Victimization (-) | 315 | 0.038 | -0.091 | -0.126 | -0.037 | 0.639 |
| | | | (0.086) | (0.118) | | |
| Risky Behavior | 316 | 0.000 | 0.069 | 0.096 | 0.011 | 0.561 |
| | | | (0.054) | (0.075) | | |

Note: Data are from survey designed by research team and given to a randomly selected subsample of youth, proportional to overall treatment and control group size randomized into Study 2 during 2013. Unless otherwise noted with (-), the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * = p-value $<$ 0.1, ** = p-value $<$ 0.05, *** = p-value $<$ 0.01.

Table 13: Wave 2: Estimated Effects on Outcomes from ISR Survey, BAM 2x2 Study

| | N | Control Mean | Intention to Treat | Effect of Participation (IV) | Control Complier Mean | FDR Q-Value |
|---|---|---|---|---|---|---|
| Indices | | | | | | |
| Automaticity | 782 | 0.017 | 0.06 | 0.093 | -0.006 | 0.834 |
| | | | (0.05) | (0.078) | | |
| Self Distancing | 781 | 0.008 | -0.026 | -0.041 | 0.027 | 0.834 |
| | | | (0.034) | (0.053) | | |
| Grit | 781 | -0.046 | 0.033 | 0.051 | -0.003 | 0.834 |
| | | | (0.042) | (0.066) | | |
| Sensation Seeking | 781 | -0.025 | 0.036 | 0.056 | -0.074 | 0.834 |
| | | | (0.042) | (0.065) | | |
| Peer Conflict Vignette 1 (-) | 781 | -0.046 | -0.042 | -0.066 | -0.025 | 0.834 |
| | | | (0.052) | (0.081) | | |
| Peer Conflict Vignette 2 (-) | 781 | -0.015 | -0.012 | -0.019 | 0.001 | 0.863 |
| | | | (0.048) | (0.075) | | |
| Peer Conflict Vignette 3 (-) | 781 | -0.004 | -0.051 | -0.079 | 0.019 | 0.834 |
| | | | (0.054) | (0.084) | | |
| Peer Conflict Vignette 4 (-) | 781 | -0.025 | -0.004 | -0.007 | 0.011 | 0.931 |
| | | | (0.05) | (0.078) | | |
| How I Think | 781 | -0.017 | 0.039 | 0.061 | -0.059 | 0.834 |
| | | | (0.04) | (0.062) | | |
| Conscientiousness (-) | 781 | 0.058 | -0.034 | -0.053 | 0.039 | 0.834 |
| | | | (0.044) | (0.069) | | |
| Education and Schooling | 781 | 0.004 | -0.009 | -0.015 | 0.088 | 0.863 |
| | | | (0.032) | (0.049) | | |
| Social Networks | 781 | -0.010 | -0.023 | -0.036 | 0.063 | 0.834 |
| | | | (0.042) | (0.065) | | |
| Adult Supports | 781 | 0.002 | -0.029 | -0.045 | 0.057 | 0.834 |
| | | | (0.047) | (0.074) | | |
| Growth Mindset | 781 | 0.018 | -0.065 | -0.101 | 0.071 | 0.834 |
| | | | (0.051) | (0.08) | | |
| Subjective Expectations | 779 | 0.008 | 0.022 | 0.034 | 0.081 | 0.834 |
| | | | (0.042) | (0.066) | | |
| Mental Health | 781 | -0.011 | -0.01 | -0.015 | 0.039 | 0.863 |
| | | | (0.037) | (0.058) | | |
| Crime Victimization (-) | 781 | -0.013 | -0.012 | -0.019 | -0.106 | 0.863 |
| | | | (0.054) | (0.084) | | |
| Risky Behavior | 781 | 0.011 | 0.022 | 0.034 | 0.061 | 0.834 |
| | | | (0.033) | (0.052) | | |

Note: Data are from survey designed by research team and given to a randomly selected subsample of youth, proportional to overall treatment and control group size randomized into Study 2x2 during 2013. Unless otherwise noted with (-), the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * = p-value $< 0.1$, ** = p-value $< 0.05$, *** = p-value $< 0.01$.

**Table 14. Effect of BAM Participation on Decision-Making Time and Retaliation in Iterated Dictator Game, BAM Study 2**

| | Log time to make decisions (seconds) | | Take amount ($) | |
|---|---|---|---|---|
| | Control Complier Mean | Effect of BAM participation (IV) | Control Complier Mean | Effect of BAM participation (IV) |
| **All Conditions Pooled** | 1.000 | 0.2956** | 7.092 | 0.2076 |
| (n = 490) | | (.1273) | | (.2237) |
| **Condition 1** | 1.075 | 0.6224*** | 7.234 | -0.4527 |
| No delay (n = 117) | | (.2396) | | (.4282) |
| **Condition 2** | 0.889 | 0.0784 | 6.719 | 0.9061** |
| Delay (n = 126) | | (.2254) | | (.4082) |
| **Condition 3** | 1.011 | 0.1940 | 7.064 | 0.2599 |
| Delay plus reflection (n = 120) | | (.237) | | (.4777) |
| **Condition 4** | 0.804 | 0.1776 | 7.423 | -0.0506 |
| Delay plus rumination (n = 127) | | (.2411) | | (.4378) |

Notes: Table presents results from administering iterated dictator game to sub-sample of youth in BAM study 2. Sample sizes listed for retaliation decision (take amount); decision time was measured for all youth in condition 1 but just for sub-sample of youth in conditions 2-4. Sample sizes for those conditions are 60, 63, and 62 respectively. Baseline covariates and randomization block fixed effects included in all models. Heteroskedasticity-robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

## Effect of BAM on School Engagement Index Outcomes in Each Study
## Average and Program Completion Effects

Figure 1: School Engagement Index (Average Effect)

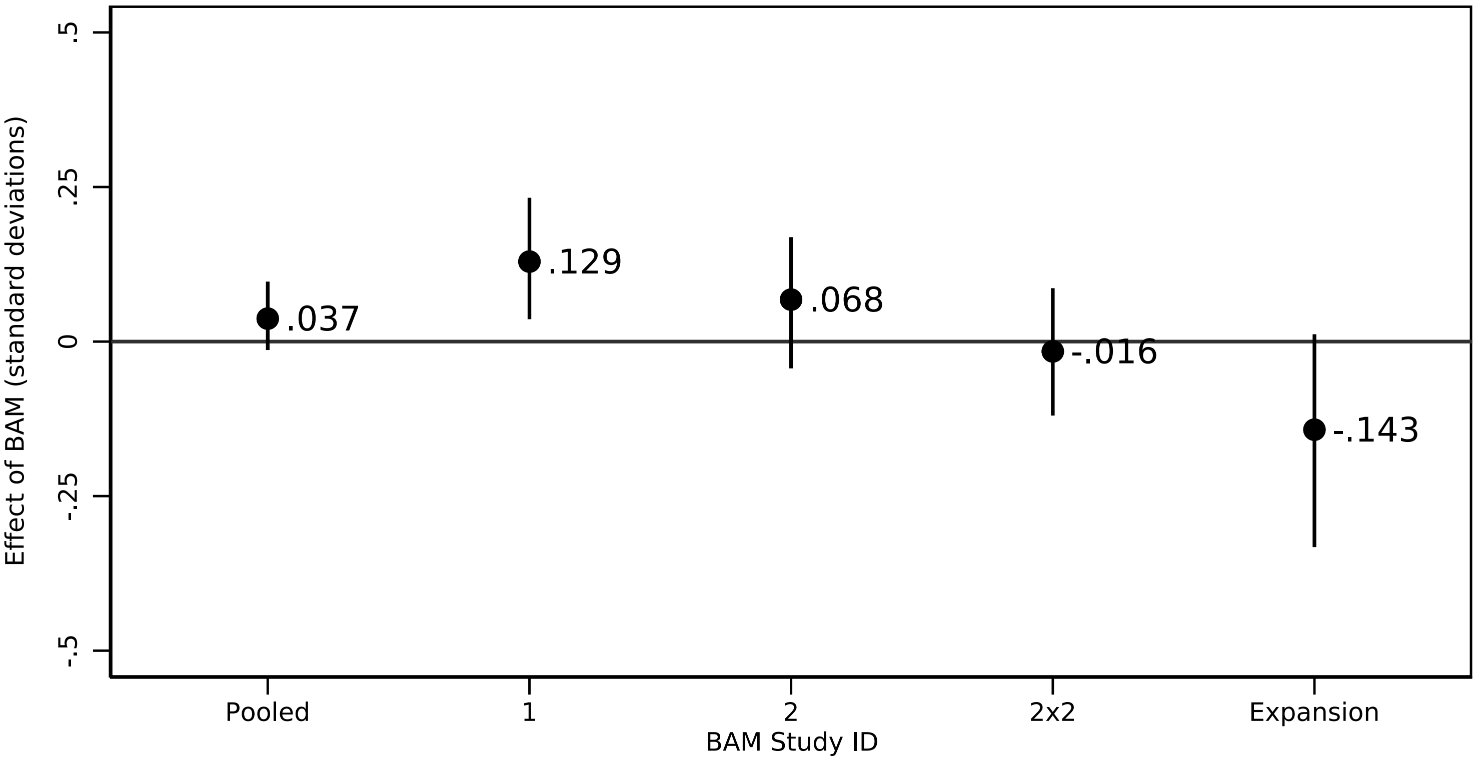
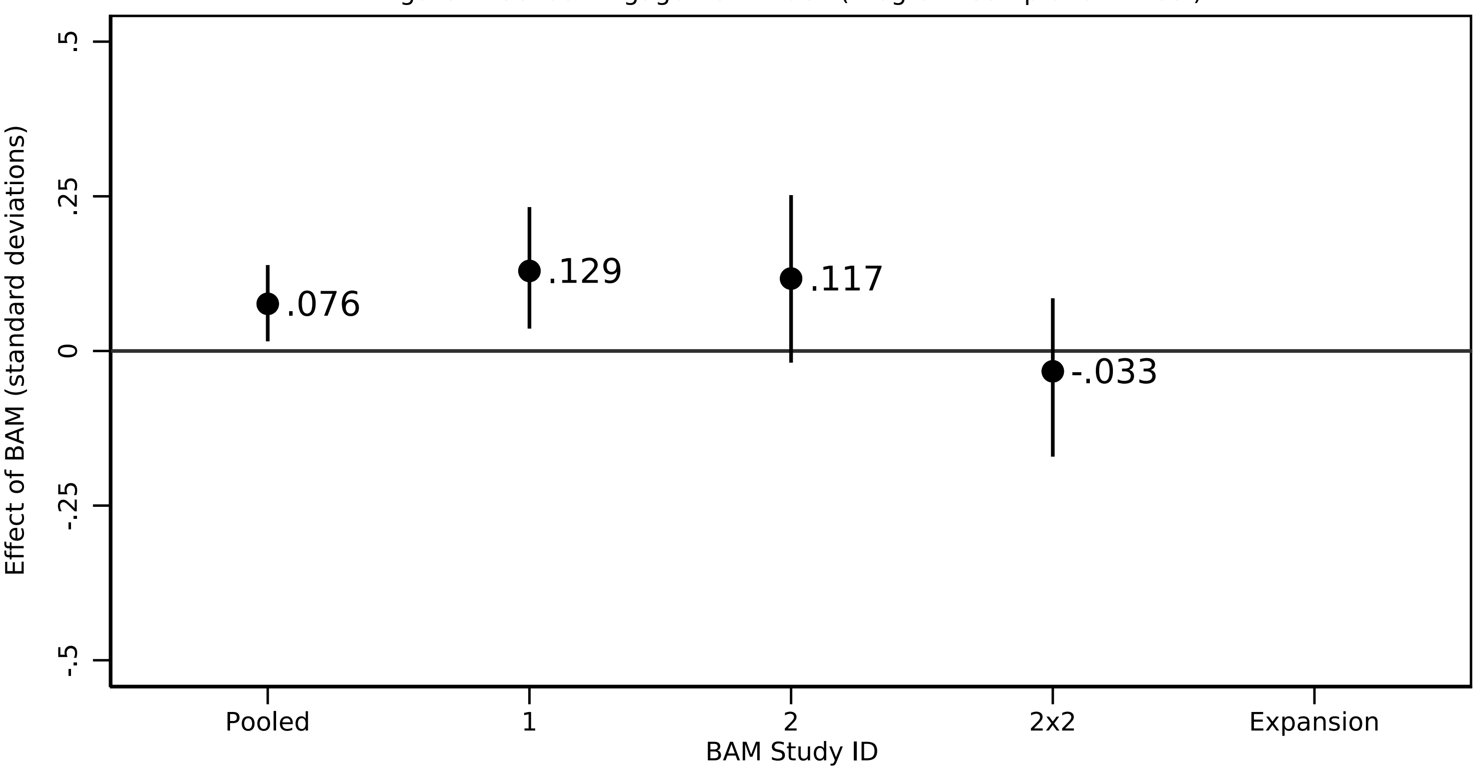



Figure 2: School Engagement Index (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in standard deviationss, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

Effect of BAM on All Arrests Outcomes in Each Study
Average and Program Completion Effects

Figure 3: All Arrests (Average Effect)

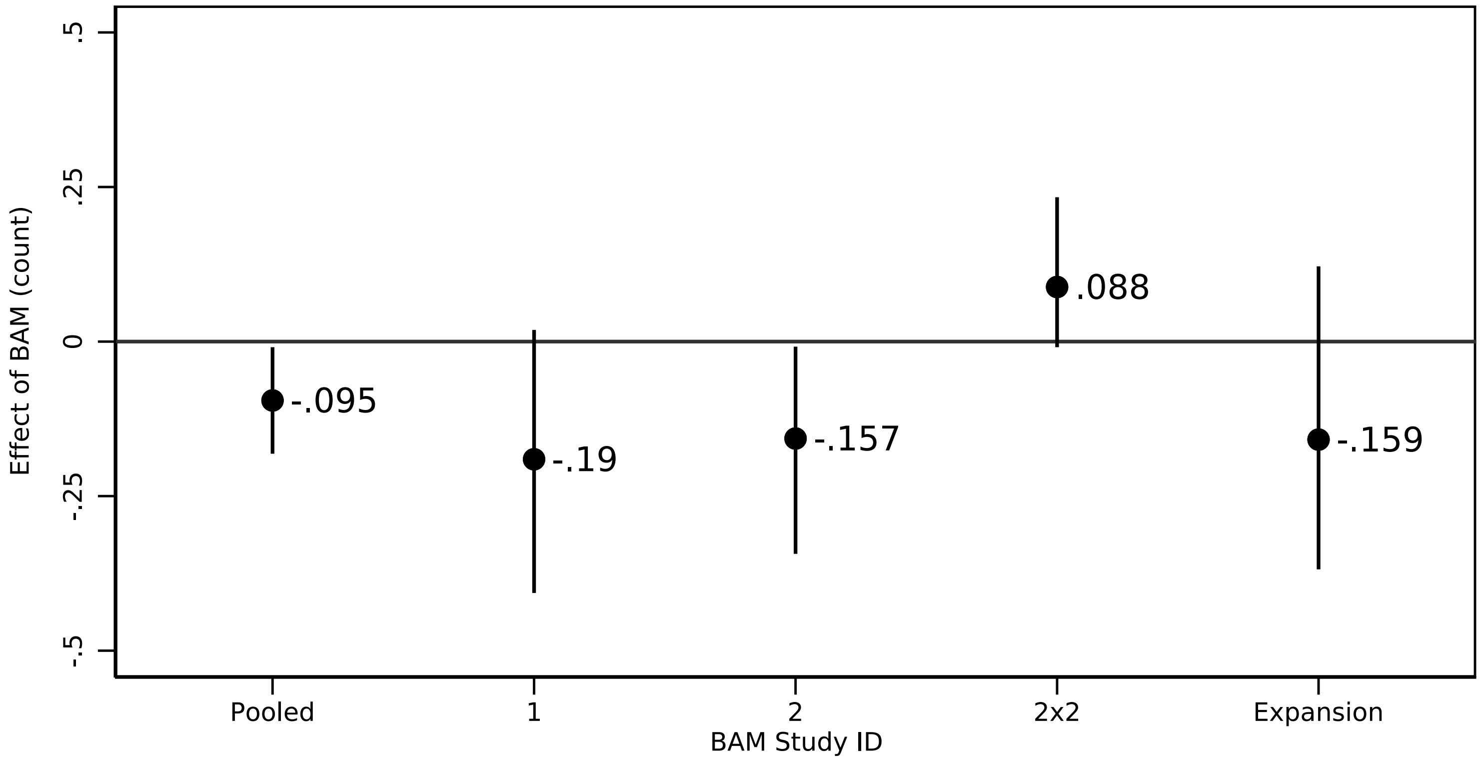
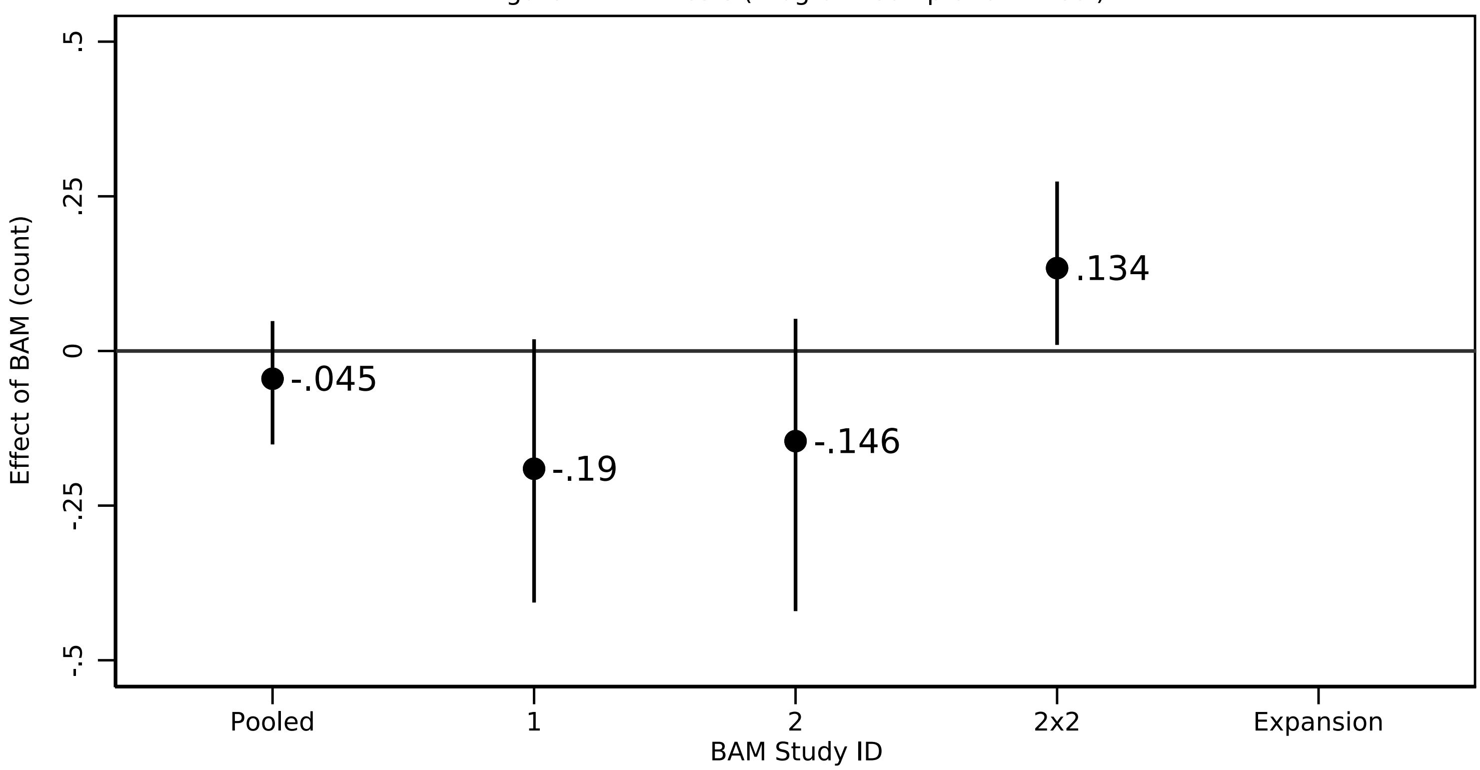



Figure 4: All Arrests (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

Effect of BAM on Violent Arrests Outcomes in Each Study
Average and Program Completion Effects

Figure 5: Violent Arrests (Average Effect)

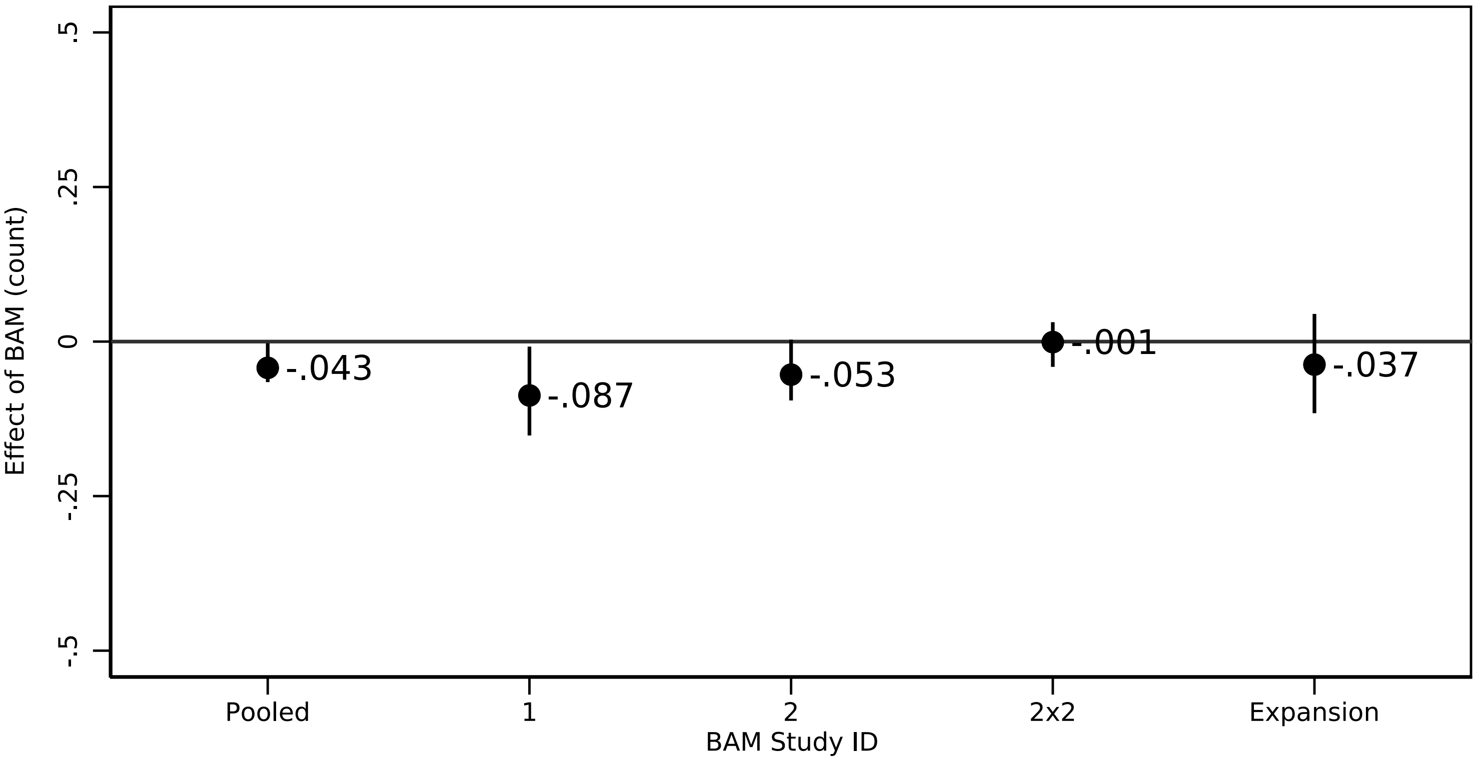
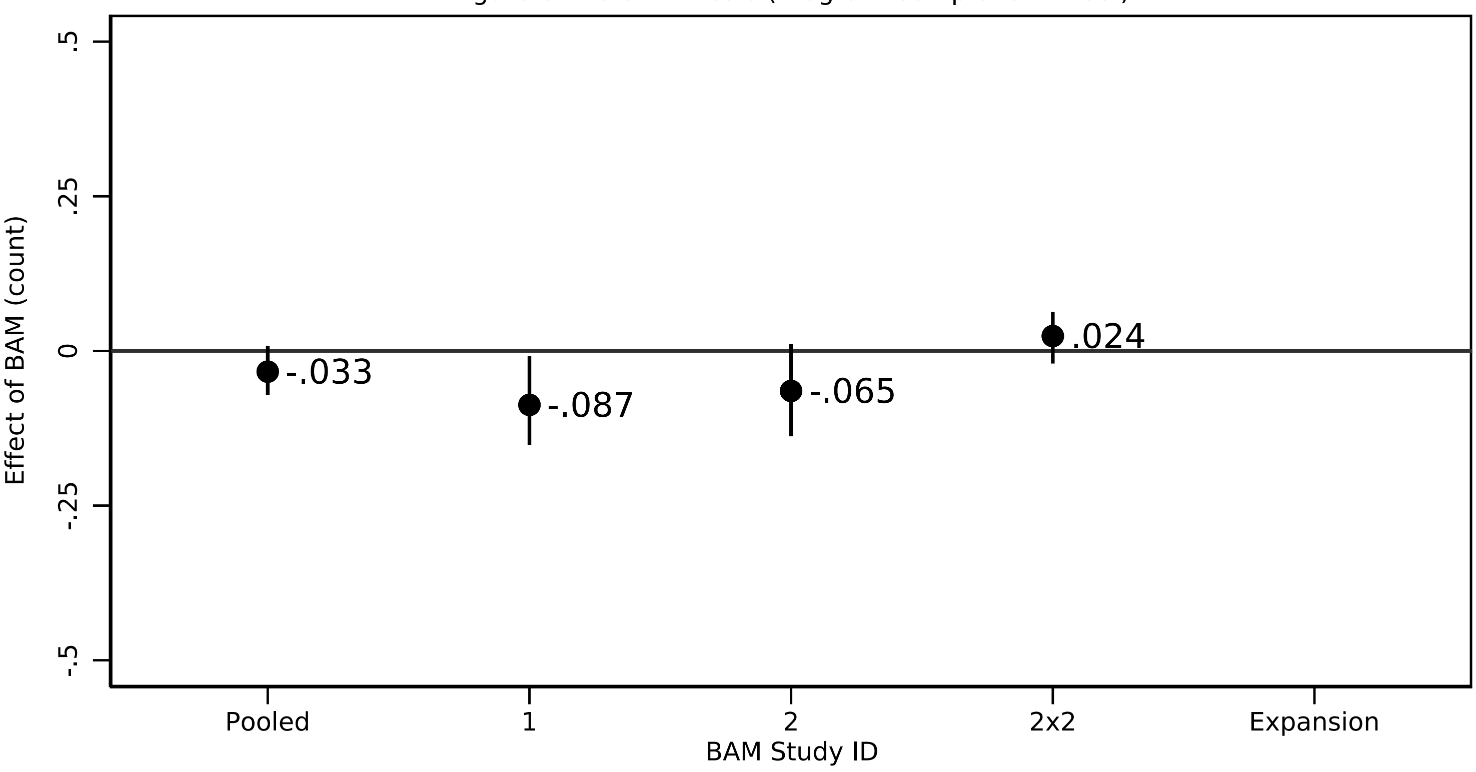



Figure 6: Violent Arrests (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

Effect of BAM on Property Arrests Outcomes in Each Study
Average and Program Completion Effects

Figure 7: Property Arrests (Average Effect)

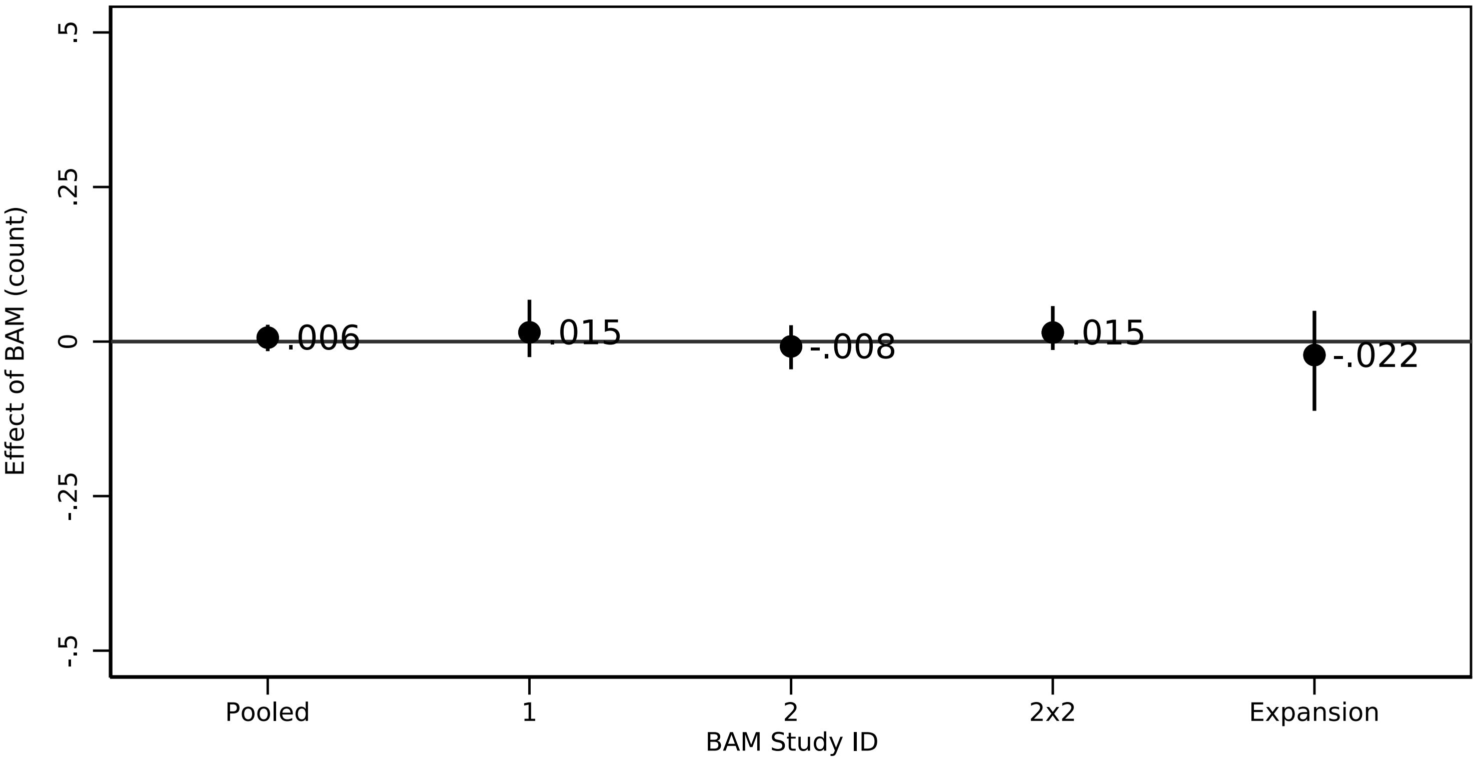
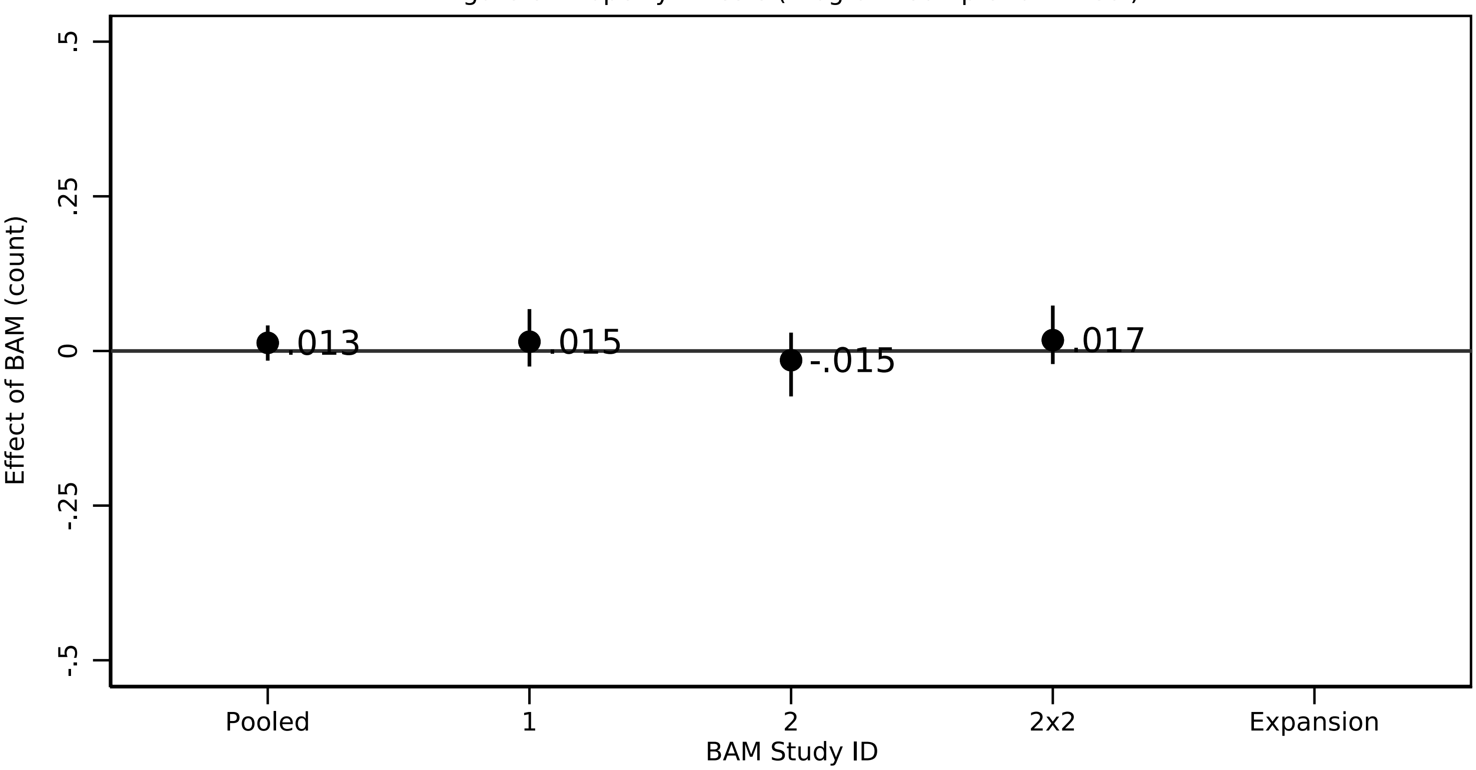



Figure 8: Property Arrests (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

Effect of BAM on Drug Arrests Outcomes in Each Study
Average and Program Completion Effects

Figure 9: Drug Arrests (Average Effect)

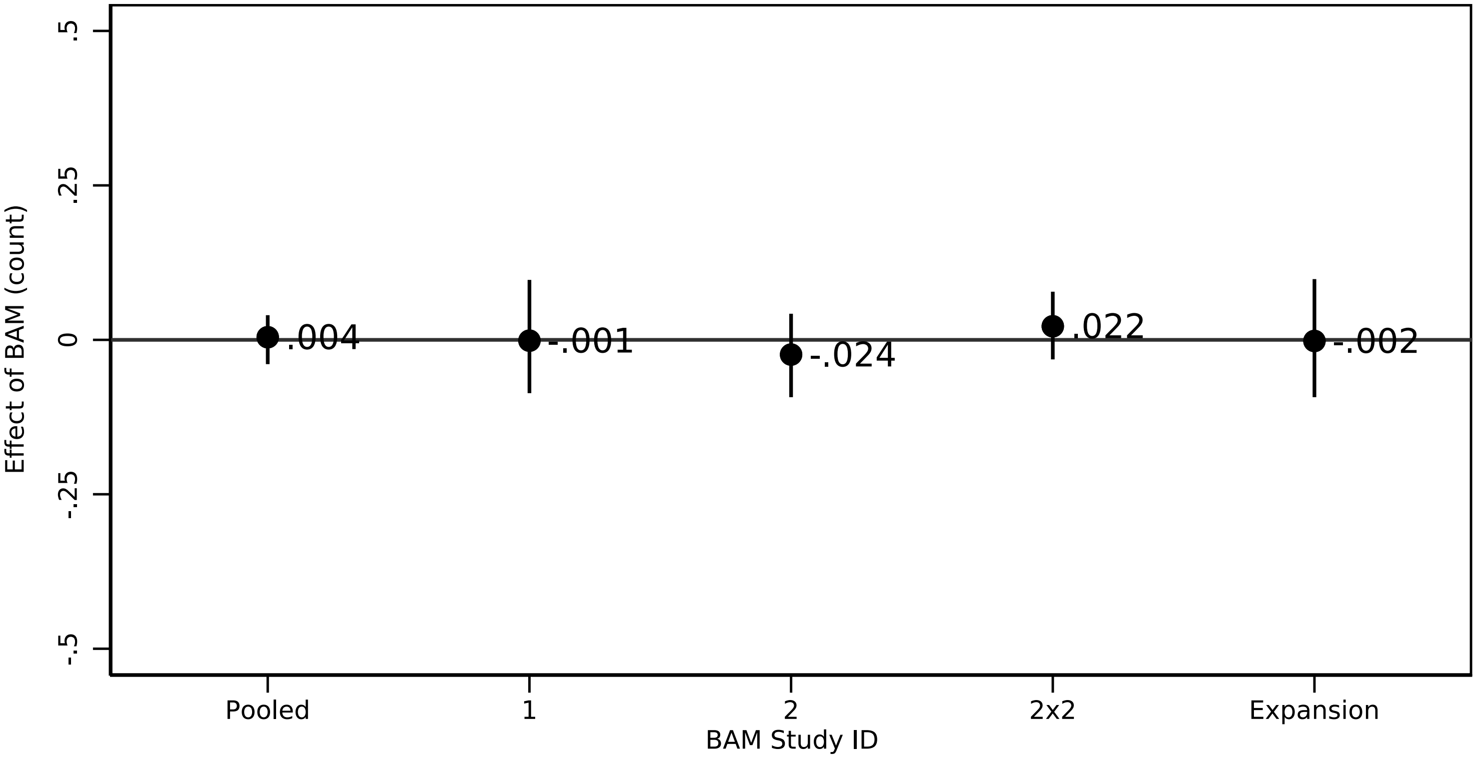
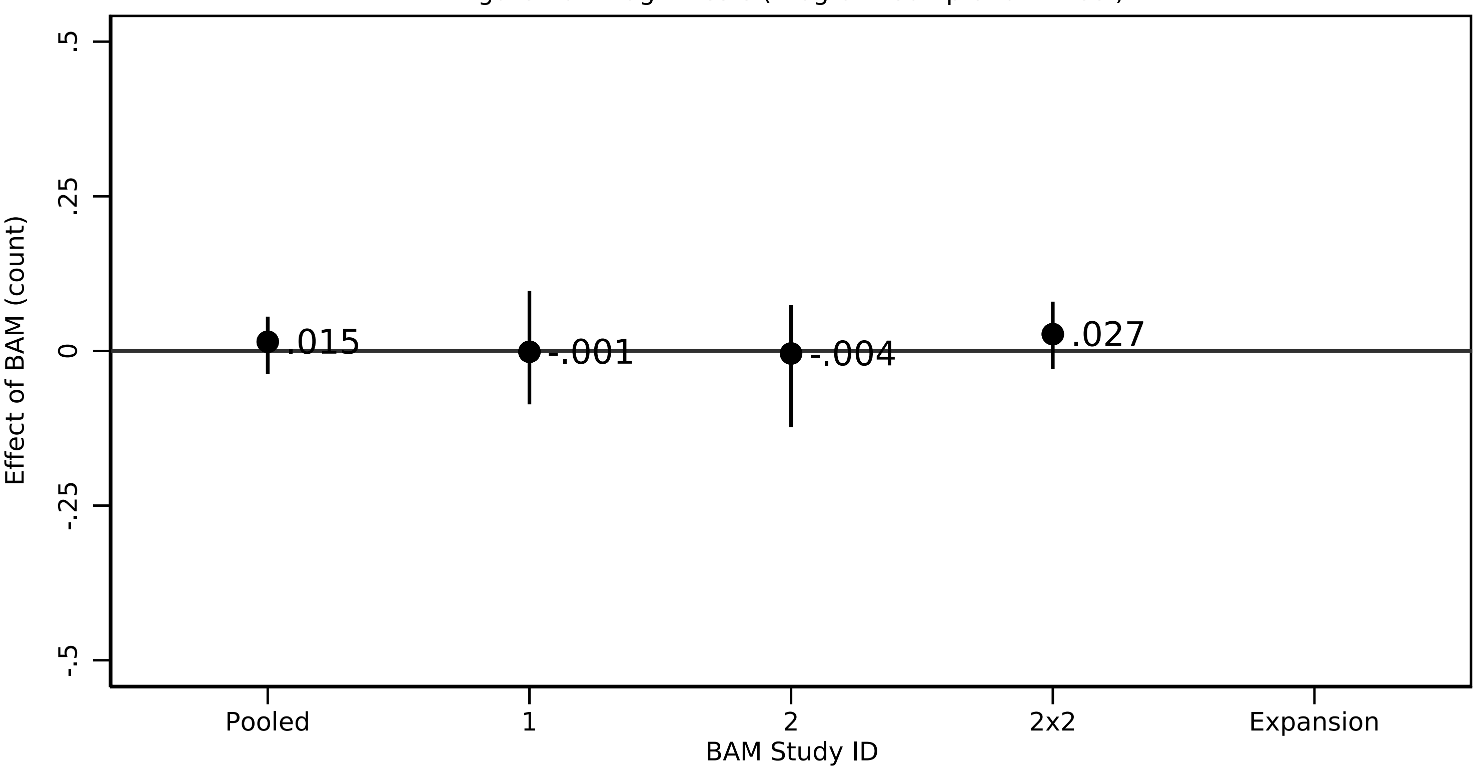



Figure 10: Drug Arrests (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

Effect of BAM on Other Arrests Outcomes in Each Study
Average and Program Completion Effects

Figure 11: Other Arrests (Average Effect)

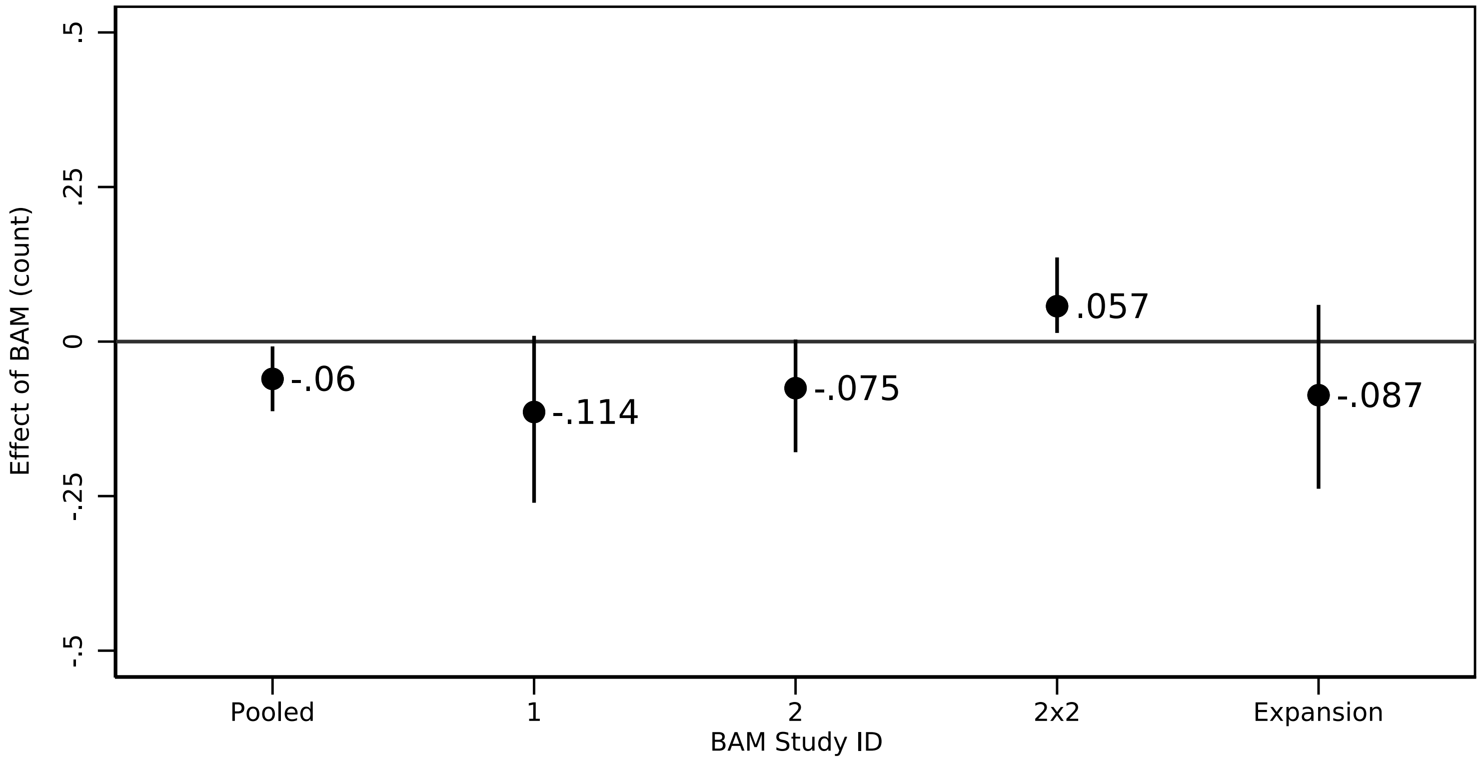
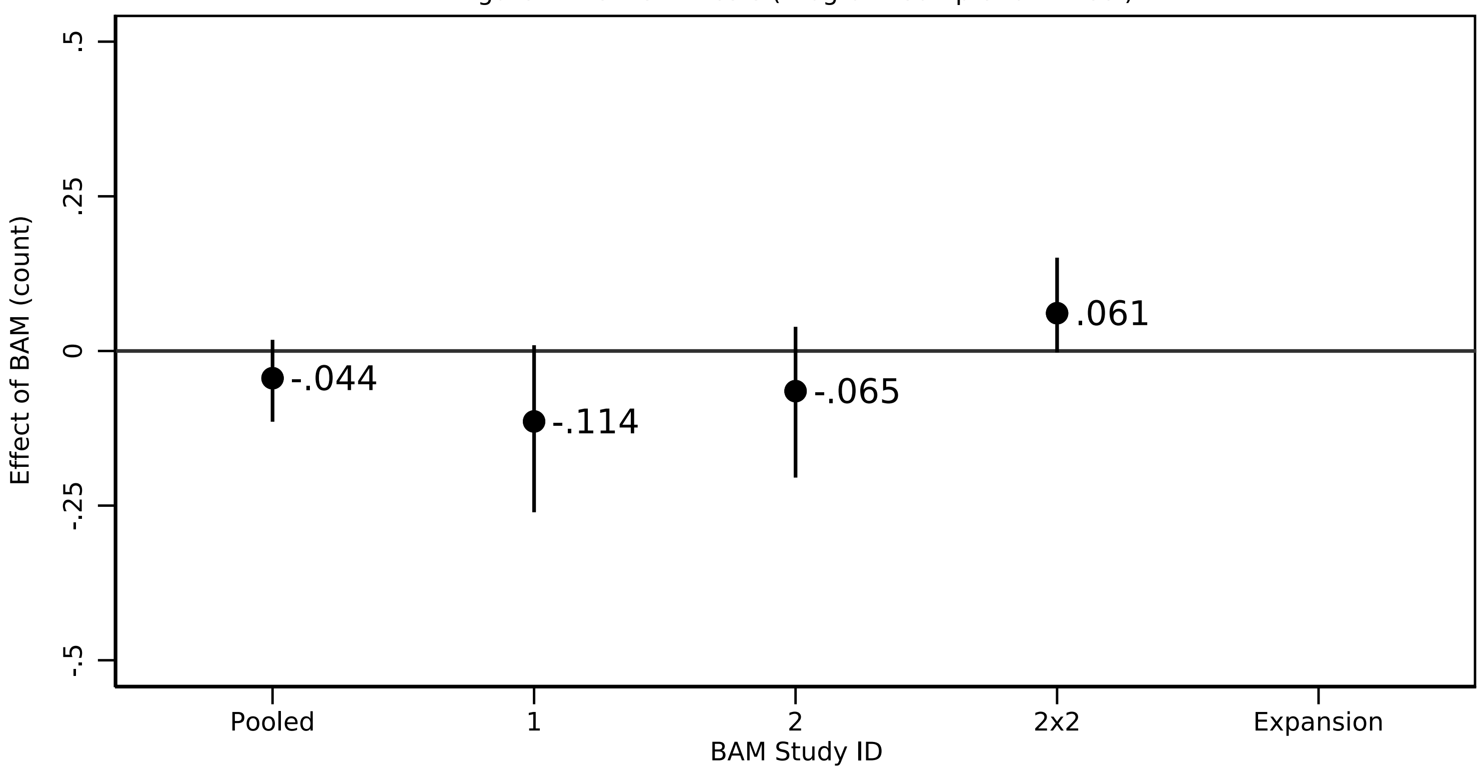



Figure 12: Other Arrests (Program Completion Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The average effect measures outcomes from Year1 of Study 1 and Study Expansion and outcomes from Years 1 and 2 of Study 2 and Study 2x2 (see text). The program completion effect measures outcomes from year 1 of Study 1 and year 2 of Study 2 and Study 2x2. Study Expansion is omitted from the program completion effect since data are only available for the first year but the program was designed as a two-year curriculum (see text).

# APPENDIX

**Table A1: Reliability and Validity Information Regarding ISR Surveys, Waves 1 and 2**

Below, we present some information regarding the reliability and sources of our ISR survey questions. Specifically, we present raw and standardized Cronbach's alpha scores for each of our mechanisms indices. We also present the sources of questions used in the survey. The vast majority of our survey questions come from surveys that have been validated; please see source information for each scale's validity.

**Reliability and Sources for ISR Survey Wave 1**

| Mechanisms Index | Raw Cronbach's alpha | Standardized Cronbach's alpha | Question Source (see below for more information) |
|---|---|---|---|
| Automaticity | 0.49 | 0.50 | Phillips & Springer, 1992; *original questions* |
| Self Distancing | 0.22 | 0.21 | *Original questions* |
| Awareness of past action | 0.23 | 0.34 | *Original questions* |
| Locus of control | 0.53 | 0.53 | National Center for Education Statistics (NCES), National Education Longitudinal Study of 1998 |
| Grit | 0.66 | 0.67 | Duckworth & Quinn, 2009 |
| Conscientiousness | 0.43 | 0.47 | MacCann et al., 2009 |
| Peer conflict vignette 1 | 0.46 | 0.42 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 2 | 0.34 | 0.36 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 3 | 0.37 | 0.35 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 4 | 0.56 | 0.53 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 5 | 0.56 | 0.48 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 6 | 0.50 | 0.47 | Fast Track Project – Adolescent Stories |
| Social Networks | 0.38 | 0.55 | US Department of Housing and Urban Development, 2002; National Center for Education Statistics (NCES), National Education Longitudinal Study of 1998; *original questions* |
| Education and Schooling | 0.66 | 0.68 | US Department of Housing and Urban Development, 2002; National Center for Education Statistics (NCES), National Education Longitudinal Study of 1998; National Center for Education Statistics, Early Childhood Longitudinal Program; *original questions* |
| Adult Supports | 0.60 | 0.64 | US Department of Housing and Urban Development, 2002; National Center for Education Statistics (NCES), National Education Longitudinal Study of 1998; National Center for Education Statistics, Early Childhood Longitudinal Program; *original questions* |

**Reliability and Sources for ISR Survey Wave 2**

| Mechanisms Index | Raw Cronbach's alpha | Standardized Cronbach's alpha | Question Source (see below for more information) |
|---|---|---|---|
| Automaticity | 0.51 | 0.53 | Phillips & Springer, 1992; *original questions* |
| Self Distancing | 0.52 | 0.52 | White, Kross, and Duckworth, 2015; *original questions* |
| Non-cognitive Grit | 0.68 | 0.70 | Duckworth & Quinn, 2009 |
| Sensation Seeking | 0.52 | 0.53 | *Original questions* |
| Peer conflict vignette 1 | 0.53 | 0.53 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 2 | 0.31 | 0.31 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 3 | 0.56 | 0.56 | Fast Track Project – Adolescent Stories |
| Peer conflict vignette 4 | 0.40 | 0.40 | Fast Track Project – Adolescent Stories |
| How I Think | 0.64 | 0.64 | Nas, Brugman, & Koops, 2008 |
| Conscientiousness | 0.78 | 0.79 | John and Srivastava, 1999 |
| Education and Schooling | 0.53 | 0.73 | National Center for Education Statistics (NCES), National Education Longitudinal Study of 1998; National Center for Education Statistics, Early Childhood Longitudinal Program; CPS School Connection Survey; March et al. 2006; *original questions* |
| Social Networks | 0.52 | 0.59 | US Department of Housing and Urban Development, 2002; National Center for Education Statistics (NCES). National Education Longitudinal Study of 1998; *original questions* |
| Adult Supports | 0.44 | 0.48 | US Department of Housing and Urban Development, 2002 |
| Growth Mindset | 0.80 | 0.81 | Dweck, 2006; Chiu et al., 1997 |
| Subjective Expectations | 0.36 | 0.37 | Bureau of Labor Statistics, 1997 |
| Mental Health | 0.59 | 0.69 | Kessler, 2002 |
| Crime Victimization | 0.71 | 0.72 | National Longitudinal Study of Adolescent to Adult Health (1994) |
| Risky Behavior | 0.76 | 0.81 | US Department of Housing and Urban Development, 2002; National Comorbidity Survey, 2005; Centers for Disease Control and Prevention, 2009; National Longitudinal Study of Adolescent to Adult Health, 1994; *original questions* |

## Source Information for ISR Survey Questions


1. American Institutes for Research. (2007). Chicago Public Schools' Student Connection Survey. Retrieved from: http://www.air.org/sites/default/files/downloads/report/CFL_S ample_Score_Report_1690_ northside_learning_center_high.pdf

2. Bureau of Labor Statistics. (1997). *National Longitudinal Survey 1997 (NLSY97).* [Measure instrument]. Retrieved from http://www.bls.gov/nls/nlsy97.htm

3. Centers for Disease Control and Prevention. (2009). Youth Risk Behavior Survey. Available at: www.cdc.gov/yrbss.

4. Chiu et al. (1997). Implicit Theories and Conceptions of Morality. *Journal of Personality and Social Psychology* 73(5): 923-940.

5. Duckworth, A.L, & Quinn, P.D. (2009). Development and validation of the Short Grit Scale (GritS). Journal of Personality Assessment, 91, 166-174. http://www.sas. upenn.edu/~duckwort/images/Duckworth%20and%20Quinn.pdf

6. Dweck, C. (2006). Growth Mindset. [Measurement Instrument]. Retrieved from http://mindsetonline.com/testyourmindset/step1.php

7. Fast Track Project. (1998). *Adolescent Stories*. [Measurement instrument]. Retrieved from http://www.fasttrackproject.org/techrept/a/ast/adolp.pdf

8. John, O.P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. "Promise and Paradox: Measuring Students' Noncognitive Skills and the Impact of Schooling" Retrieved from https://www.ocf.berkeley.edu/~johnlab/bfi.htm

9. Kessler, R.C. (2002). *Kessler K-6 Psychological Distress Scale*. [Measurement instrument] Retrieved from http://www.hcp.med.harvard.edu/ncs/ftpdir/k6/K6+self%20admin-3-05-%20FINAL.pdf

10. MacCann, C. et al. (2009). Empirical identification of the major facets of Conscientiousness, Learning and Individual Differences, doi: 10.1016/j.lindif.2009.03.007

11. March et al. (2006). *Percieved self-efficacy scale*. [Measurement instrument]

12. Nas, C. N., Brugman, D. and Koops, W. (2008). Cognitive Disorders and the How I Think Questionnaire. *European Journal of Psychological Assessment* 24(3): 181-189. Retrieved from http://www.researchgate.net/profile/Daniel_Brugman2/publication/228328799 _Measuring_SelfServing_Cognitive_Distortions_with_the_How_I_Think_Questionnaire/links/0912f4ff77 68e8d259000000.pdf

13. National Center for Education Statistics (NCES). Early Childhood Longitudinal Program (ECLS). Retrieved from https://nces.ed.gov/ecls/kinderinstruments.asp

14. National Center for Education Statistics (NCES). National Education Longitudinal Study of 1998 (NELS: 88) Student Questionnaires. Retrieved from: https://nces.ed.gov/surveys/ nels88/questionnaires.asp

15. National Comorbidity Survey. (2005). *NCS-A*. Retrieved from http://www.hcp.med.harvard.edu/ncs/index.php

16. National Longitudinal Study of Adolescent to Adult Health (Add Health). (1994). Add Health. [Measurement instrument]. Retrieved from http://www.cpc.unc.edu/projects /addhealth/data/publicdata

17. Phillips, J. L. and Springer, F.J. (1992). *Conflict Resolution-Individual Protective Factors Index*. [Measurement instrument]. Retrieved from https://www.cyfernetsearch.org/sites /default/files/PsychometricsFiles/Phillips-Conflict%20Resolution%20%28Grades%207-11%29.pdf

18. US Department of Housing and Urban Development. (2002). *Moving to Opportunity Interim Evaluation-Youth*. [Measurement instrument]. Retrieved from http://www.nber.org/mtopublic/instruments/interim_youth.pdf

19. White, R. E., Kross, E. and Duckworth, A.L. (2015). Spontaneous Self-Distancing and Adaptive Self-Reflection Across Adolescence. *Child Development*, 1467-8624. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/cdev.12370/abstract.

**Table A2: BAM Sensitivity Analysis: Two methods of missing data imputation for the School Engagement Index**

**Pooled**

| | First-year effect (N = 9307) | | | | Second-year effect (N = 4174) | | | | Average effect (N = 9307) | | | | Program completion effect (N = 6914) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM |
| Main Results | 0 | 0.0128 | 0.0301 | 0.261 | 0 | 0.0175 | 0.0524 | 0.292 | -0.003 | 0.0152 | 0.0358 | 0.228 | 0 | 0.0359*** | 0.0785*** | 0.172 |
| | | (0.0122) | (0.0286) | | | (0.0181) | (0.0532) | | | (0.0115) | (0.0267) | | | (0.0139) | (0.0300) | |
| Multiple Imputation | 0 | 0.0087 | 0.0205 | 0.228 | 0 | 0.0131 | 0.0392 | 0.171 | -0.002 | 0.0100 | 0.0235 | 0.212 | 0 | 0.0315** | 0.0687** | 0.206 |
| | | (0.0128) | (0.0301) | | | (0.0208) | (0.0613) | | | (0.0124) | (0.0288) | | | (0.0155) | (0.0336) | |

**Study 1**

| | First-year effect (N = 2740) | | | |
|---|---|---|---|---|
| | CM | ITT | IV | CCM |
| Main Results | 0 | 0.0584*** | 0.1400*** | 0.218 |
| | | (0.0214) | (0.0508) | |
| Multiple Imputation | 0 | 0.0518** | 0.1242** | 0.238 |
| | | (0.0228) | (0.0540) | |

**Study 2**

| | First-year effect (N = 2064) | | | | Second-year effect (N = 1872) | | | | Average effect (N = 2064) | | | | Program completion effect (N = 1872) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM |
| Main Results | 0 | 0.0148 | 0.0297 | 0.205 | 0 | 0.0543** | 0.1716** | 0.189 | -0.014 | 0.0349* | 0.0692* | 0.124 | 0 | 0.0543** | 0.1110** | 0.094 |
| | | (0.0245) | (0.0482) | | | (0.0269) | (0.0833) | | | (0.0211) | (0.0409) | | | (0.0269) | (0.0541) | |
| Multiple Imputation | 0 | 0.0134 | 0.0268 | 0.216 | 0 | 0.0681** | 0.2151** | -0.004 | -0.009 | 0.0365 | 0.0723 | 0.159 | 0 | 0.0681** | 0.1393** | 0.114 |
| | | (0.0256) | (0.0505) | | | (0.0306) | (0.0947) | | | (0.0244) | (0.0476) | | | (0.0306) | (0.0616) | |

**Study 2x2**

| | First-year effect (N = 2302) | | | | Second-year effect (N = 2302) | | | | Average effect (N = 2302) | | | | Program completion effect (N = 2302) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM | CM | ITT | IV | CCM |
| Main Results | 0 | -0.0033 | -0.0070 | 0.279 | 0 | -0.0144 | -0.0412 | 0.376 | 0 | -0.0089 | -0.0185 | 0.226 | 0 | -0.0144 | -0.0300 | 0.175 |
| | | (0.0255) | (0.0529) | | | (0.0241) | (0.0686) | | | (0.0217) | (0.0447) | | | (0.0241) | (0.0498) | |
| Multiple Imputation | 0 | -0.0054 | -0.0114 | 0.284 | 0 | -0.0343 | -0.0985 | 0.310 | 0 | -0.0199 | -0.0415 | 0.266 | 0 | -0.0343 | -0.0717 | 0.239 |
| | | (0.0274) | (0.0569) | | | (0.0269) | (0.0772) | | | (0.0242) | (0.0501) | | | (0.0269) | (0.0558) | |

**Study Expansion**

| | First-year effect (N = 2201) | | | |
|---|---|---|---|---|
| | CM | ITT | IV | CCM |
| Main Results | 0 | -0.0522** | -0.1742* | 0.438 |
| | | (0.0266) | (0.0892) | |
| Multiple Imputation | 0 | -0.0561** | -0.1872** | 0.304 |
| | | (0.0283) | (0.0951) | |

Note: School engagement index is a composite average of z-scores for GPA, attendance, and persistance. "Main Results" models standardize on non-missing observations and impute missings elements at the group mean. "Multiple Imputation" models impute missing elements through a 10-iteration pmm chained model with matching done using k=10 nearest neighbor at the study-x-treatment group level. All models include standard baseline covariates and randomization block-fixed effects. Study 1, Study 2, Study 2x2 models include heteroskedacity-robust standard errors. Pooled and Study Expansion models include student-clustered standard errors.

*** p < 0.01, ** p < 0.5, * p < 0.1.

**Table A3: Becoming A Man Program Effects with Multiple Hypothesis Testing Adjustments**

| | Control Mean | Intention to Treat Effect | Unadjusted p-value | FWER adjusted p, Families = a) Schooling Index + 4 Crime Categories | FWER adjusted p, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime |
|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Pooled BAM Studies (Program Years: 2009-2010, 2013-2015)** | | | | | | | | | |
| **First-year effect (N = 9307)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0128 | 0.192 | 0.496 | 0.227 | 0.320 | 0.192 | 0.238 | 0.238 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.112 | -0.0199 | 0.035 | 0.176 | 0.141 | 0.177 | 0.177 | 0.215 | 0.215 |
| Property Offenses | 0.064 | 0.0023 | 0.784 | 0.954 | 0.954 | 0.980 | 0.980 | 0.645 | 0.645 |
| Drug Offenses | 0.099 | -0.0001 | 0.987 | 0.987 | 0.987 | 0.988 | 0.988 | 0.653 | 0.653 |
| Other Offenses | 0.234 | -0.0269 | 0.074 | 0.279 | 0.221 | 0.185 | 0.179 | 0.215 | 0.215 |
| All Offenses | 0.509 | -0.0446 | 0.107 | | 0.261 | | 0.179 | | 0.215 |
| **Second-year effect (N = 4174)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0175 | 0.256 | 0.792 | 0.339 | 0.806 | 0.257 | 1.000 | 0.345 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.079 | -0.0051 | 0.570 | 0.873 | 0.925 | 0.806 | 0.932 | 1.000 | 1.000 |
| Property Offenses | 0.051 | 0.0035 | 0.644 | 0.873 | 0.925 | 0.806 | 0.932 | 1.000 | 1.000 |
| Drug Offenses | 0.103 | 0.0072 | 0.608 | 0.873 | 0.925 | 0.806 | 0.932 | 1.000 | 1.000 |
| Other Offenses | 0.201 | 0.0050 | 0.931 | 0.931 | 0.931 | 0.932 | 0.932 | 1.000 | 1.000 |
| All Offenses | 0.435 | 0.0106 | 0.843 | | 0.931 | | 0.932 | | 1.000 |
| **Average effect (N = 9307)** | | | | | | | | | |
| School Engagement | -0.003 | 0.0152 | 0.108 | 0.318 | 0.145 | 0.180 | 0.108 | 0.220 | 0.121 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.109 | -0.0163 | 0.063 | 0.294 | 0.235 | 0.180 | 0.232 | 0.220 | 0.301 |
| Property Offenses | 0.062 | 0.0035 | 0.579 | 0.811 | 0.811 | 0.724 | 0.724 | 0.407 | 0.487 |
| Drug Offenses | 0.102 | 0.0023 | 0.810 | 0.811 | 0.811 | 0.811 | 0.811 | 0.480 | 0.487 |
| Other Offenses | 0.232 | -0.0236 | 0.093 | 0.318 | 0.267 | 0.180 | 0.232 | 0.220 | 0.301 |
| All Offenses | 0.505 | -0.0341 | 0.196 | | 0.434 | | 0.328 | | 0.301 |
| **Program completion effect (N = 6914)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0359 | 0.007 | 0.043 | 0.017 | 0.034 | 0.007 | 0.036 | 0.007 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.111 | -0.0131 | 0.345 | 0.506 | 0.659 | 0.505 | 0.631 | 0.677 | 1.000 |
| Property Offenses | 0.060 | 0.0069 | 0.215 | 0.506 | 0.619 | 0.505 | 0.631 | 0.677 | 1.000 |
| Drug Offenses | 0.121 | 0.0073 | 0.470 | 0.506 | 0.659 | 0.505 | 0.631 | 0.677 | 1.000 |
| Other Offenses | 0.239 | -0.0163 | 0.505 | 0.506 | 0.659 | 0.505 | 0.631 | 0.677 | 1.000 |
| All Offenses | 0.531 | -0.0152 | 0.929 | | 0.928 | | 0.929 | | 1.000 |

Note: BAM Study 1 and BAM Study 2 results on next page

| | Control Mean | Intention to Treat Effect | Unadjusted p-value | FWER adjusted p, Families = a) Schooling Index + 4 Crime Categories | FWER adjusted p, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime |
|---|---|---|---|---|---|---|---|---|---|
| **Panel B: BAM Study 1 (Program Years: 2009-2010)** | | | | | | | | | |
| **First-year effect (N = 2740)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0584 | 0.006 | 0.034 | 0.009 | 0.033 | 0.007 | 0.034 | 0.007 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.167 | -0.0346 | 0.037 | 0.140 | 0.149 | 0.094 | 0.187 | 0.081 | 0.230 |
| Property Offenses | 0.077 | 0.0075 | 0.558 | 0.807 | 0.807 | 0.698 | 0.698 | 0.388 | 0.388 |
| Drug Offenses | 0.151 | -0.0001 | 0.994 | 0.994 | 0.994 | 0.995 | 0.995 | 0.661 | 0.661 |
| Other Offenses | 0.305 | -0.0472 | 0.090 | 0.241 | 0.263 | 0.150 | 0.192 | 0.103 | 0.230 |
| All Offenses | 0.699 | -0.0744 | 0.115 | | 0.278 | | 0.192 | | 0.230 |
| **Panel C: BAM Study 2 (Program Years: 2013-2015)** | | | | | | | | | |
| **First-year effect (N = 2064)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0148 | 0.545 | 0.641 | 0.590 | 0.636 | 0.545 | 0.651 | 1.000 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.119 | -0.0223 | 0.158 | 0.532 | 0.409 | 0.395 | 0.263 | 0.651 | 0.357 |
| Property Offenses | 0.072 | -0.0059 | 0.636 | 0.641 | 0.641 | 0.636 | 0.636 | 0.651 | 0.357 |
| Drug Offenses | 0.127 | -0.0178 | 0.452 | 0.641 | 0.641 | 0.636 | 0.565 | 0.651 | 0.357 |
| Other Offenses | 0.277 | -0.0460 | 0.112 | 0.475 | 0.378 | 0.395 | 0.263 | 0.651 | 0.357 |
| All Offenses | 0.595 | -0.0921 | 0.064 | | 0.244 | | 0.263 | | 0.357 |
| **Second-year effect (N = 1872)** | | | | | | | | | |
| School Engagement | 0.000 | 0.0543 | 0.044 | 0.244 | 0.099 | 0.221 | 0.045 | 0.284 | 0.047 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.117 | -0.0280 | 0.144 | 0.455 | 0.465 | 0.360 | 0.719 | 0.403 | 1.000 |
| Property Offenses | 0.072 | -0.0034 | 0.809 | 0.947 | 0.947 | 0.947 | 0.947 | 1.000 | 1.000 |
| Drug Offenses | 0.147 | 0.0017 | 0.947 | 0.947 | 0.947 | 0.947 | 0.947 | 1.000 | 1.000 |
| Other Offenses | 0.284 | -0.0253 | 0.438 | 0.817 | 0.817 | 0.731 | 0.731 | 0.780 | 1.000 |
| All Offenses | 0.620 | -0.0550 | 0.343 | | 0.731 | | 0.731 | | 1.000 |
| **Average effect (N = 2064)** | | | | | | | | | |
| School Engagement | -0.014 | 0.0349 | 0.097 | 0.364 | 0.160 | 0.241 | 0.098 | 0.318 | 0.108 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.115 | -0.0244 | 0.062 | 0.309 | 0.230 | 0.241 | 0.241 | 0.318 | 0.318 |
| Property Offenses | 0.069 | -0.0031 | 0.758 | 0.758 | 0.758 | 0.758 | 0.758 | 0.435 | 0.435 |
| Drug Offenses | 0.131 | -0.0081 | 0.660 | 0.758 | 0.758 | 0.758 | 0.758 | 0.435 | 0.435 |
| Other Offenses | 0.268 | -0.0348 | 0.144 | 0.364 | 0.364 | 0.241 | 0.241 | 0.318 | 0.318 |
| All Offenses | 0.583 | -0.0704 | 0.101 | | 0.284 | | 0.241 | | 0.318 |

Note: BAM Study 2x2 and BAM Study Expansion results on next page

| | Control Mean | Intention to Treat Effect | Unadjusted p-value | FWER adjusted p, Families = a) Schooling Index + 4 Crime Categories | FWER adjusted p, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime | FDR q value, Families = a) Schooling Index + 4 Crime Categories | FDR q value, Families = a) Schooling Index, b) 4 Crime Categories + Total Crime |
|---|---|---|---|---|---|---|---|---|---|
| **Panel D: BAM Study 2x2 (Program Years: 2013-2015)** | | | | | | | | | |
| **First-year effect (N = 2302)** | | | | | | | | | |
| School Engagement | 0.000 | -0.0033 | 0.896 | 0.903 | 0.903 | 0.896 | 0.896 | 1.000 | 1.000 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.069 | -0.0124 | 0.311 | 0.792 | 0.616 | 0.663 | 0.500 | 1.000 | 1.000 |
| Property Offenses | 0.045 | 0.0091 | 0.397 | 0.798 | 0.616 | 0.663 | 0.500 | 1.000 | 1.000 |
| Drug Offenses | 0.055 | 0.0071 | 0.639 | 0.880 | 0.647 | 0.799 | 0.639 | 1.000 | 1.000 |
| Other Offenses | 0.141 | 0.0267 | 0.208 | 0.703 | 0.616 | 0.663 | 0.500 | 1.000 | 1.000 |
| All Offenses | 0.309 | 0.0305 | 0.400 | | 0.616 | | 0.500 | | 1.000 |
| **Second-year effect (N = 2302)** | | | | | | | | | |
| School Engagement | 0.000 | -0.0144 | 0.552 | 0.610 | 0.610 | 0.552 | 0.552 | 0.967 | 1.000 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.052 | 0.0116 | 0.316 | 0.610 | 0.401 | 0.492 | 0.394 | 0.967 | 0.648 |
| Property Offenses | 0.035 | 0.0097 | 0.337 | 0.610 | 0.401 | 0.492 | 0.394 | 0.967 | 0.648 |
| Drug Offenses | 0.072 | 0.0146 | 0.393 | 0.610 | 0.401 | 0.492 | 0.394 | 0.967 | 0.648 |
| Other Offenses | 0.141 | 0.0304 | 0.167 | 0.610 | 0.401 | 0.492 | 0.394 | 0.967 | 0.648 |
| All Offenses | 0.301 | 0.0663 | 0.086 | | 0.312 | | 0.394 | | 0.648 |
| **Average effect (N = 2302)** | | | | | | | | | |
| School Engagement | 0.000 | -0.0089 | 0.683 | 0.909 | 0.712 | 0.854 | 0.683 | 1.000 | 1.000 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.061 | -0.0004 | 0.969 | 0.969 | 0.969 | 0.969 | 0.969 | 1.000 | 0.692 |
| Property Offenses | 0.040 | 0.0094 | 0.245 | 0.702 | 0.580 | 0.614 | 0.409 | 0.983 | 0.350 |
| Drug Offenses | 0.063 | 0.0108 | 0.382 | 0.786 | 0.624 | 0.637 | 0.478 | 1.000 | 0.402 |
| Other Offenses | 0.141 | 0.0285 | 0.099 | 0.427 | 0.325 | 0.496 | 0.259 | 0.983 | 0.350 |
| All Offenses | 0.305 | 0.0484 | 0.104 | | 0.325 | | 0.259 | | 0.350 |
| **Panel E: BAM Study Expansion (Program Years: 2014-2015)** | | | | | | | | | |
| **First-year effect (N = 2201)** | | | | | | | | | |
| School Engagement | 0.000 | -0.0522 | 0.050 | 0.242 | 0.074 | 0.249 | 0.050 | 0.330 | 0.053 |
| *Arrests per youth per year* | | | | | | | | | |
| Violent Offenses | 0.096 | -0.0101 | 0.473 | 0.858 | 0.858 | 0.789 | 0.789 | 1.000 | 1.000 |
| Property Offenses | 0.064 | -0.0050 | 0.720 | 0.924 | 0.924 | 0.901 | 0.901 | 1.000 | 1.000 |
| Drug Offenses | 0.067 | -0.0004 | 0.976 | 0.976 | 0.976 | 0.977 | 0.977 | 1.000 | 1.000 |
| Other Offenses | 0.222 | -0.0253 | 0.297 | 0.751 | 0.749 | 0.743 | 0.789 | 1.000 | 1.000 |
| All Offenses | 0.449 | -0.0409 | 0.328 | | 0.749 | | 0.789 | | 1.000 |

Note: Baseline covariates and randomization block fixed effects included in all models. Standard errors are clustered on individuals to account for students randomized into multiple studies. School engagement index is eqaul to an unweighted average of days present, GPA, and enrollment status at end of school year, all normalized to Z-score form using control group distributions. The first-year effect captures the program years: year 1 for all studies. The second-year effect captures the program years: year 2 for Study 2; year 2 for Study 2x2. The average effect captures the program years: year 1 for Study 1; an average of year 1 and year 2 for Study 2; an average of year 1 and year 2 for Study 2x2; year 1 for Study Expansion. The program completion effect captures the program years: year 1 for Study 1; year 2 for Study 2; year 2 for Study 2x2. The FWER p-value is calculated using the boostrap re-samping technique from Westfall and young (1993). The FDR one-stage q-value is calculated using thr procedure from benjamini and Hochberg (1995). The two-stage FDR q-value is calculated using the procedure from Benjamini, Krieger, and Yekutieli (2006). We calculate these values using two definitions of our 'family' of outcomes, first defining the family as our schooling variable plus the measures of arrests for different specific offense categories (violent, property, drug, other), exluding total arrests since it is a linear combination of the other four crime-type-specific measures; and then again defining two separate families of outcomes, using schooling as itw own family and then a separate family of all our arrest measures (four sub-categories and the total sum).

Table A4: Wave 1: Estimated Effects on Outcomes from ISR Survey, BAM 2x2 Study

| | N | Control Mean | Intention to Treat | Effect of Participation (IV) | Control Complier Mean | FDR Q-Value |
|---|---|---|---|---|---|---|
| Indices | | | | | | |
| Automaticity | 545 | -0.019 | -0.006 | -0.01 | 0.065 | 0.886 |
| | | | (0.045) | (0.067) | | |
| Self Distancing (-) | 513 | 0.001 | -0.058 | -0.086 | 0.031 | 0.659 |
| | | | (0.057) | (0.084) | | |
| Awareness of Past Action | 544 | -0.000 | -0.132* | -0.198* | 0.155 | 0.177 |
| | | | (0.069) | (0.103) | | |
| Locus of Control | 545 | -0.001 | -0.132*** | -0.198*** | 0.099 | 0.060 |
| | | | (0.051) | (0.077) | | |
| Grit | 545 | -0.000 | -0.082* | -0.123* | 0.064 | 0.267 |
| | | | (0.052) | (0.078) | | |
| Conscientiousness | 545 | -0.000 | -0.068 | -0.102 | 0.061 | 0.629 |
| | | | (0.06) | (0.091) | | |
| Peer Conflict Vignette 1 (-) | 545 | -0.001 | 0.038 | 0.057 | -0.075 | 0.739 |
| | | | (0.046) | (0.069) | | |
| Peer Conflict Vignette 2 (-) | 545 | 0.001 | 0.029 | 0.043 | -0.064 | 0.761 |
| | | | (0.045) | (0.067) | | |
| Peer Conflict Vignette 3 (-) | 545 | -0.000 | 0.124*** | 0.185*** | -0.082 | 0.060 |
| | | | (0.044) | (0.067) | | |
| Peer Conflict Vignette 4 (-) | 545 | -0.004 | 0.024 | 0.036 | 0.010 | 0.859 |
| | | | (0.048) | (0.072) | | |
| Peer Conflict Vignette 5 (-) | 544 | 0.000 | 0.094** | 0.141** | -0.070 | 0.175 |
| | | | (0.048) | (0.072) | | |
| Peer Conflict Vignette 6 (-) | 544 | -0.000 | -0.011 | -0.016 | 0.043 | 0.859 |
| | | | (0.044) | (0.067) | | |
| Social Networks | 544 | -0.004 | -0.033 | -0.05 | 0.065 | 0.761 |
| | | | (0.051) | (0.078) | | |
| Education and Schooling | 544 | 0.000 | -0.085** | -0.128** | 0.150 | 0.175 |
| | | | (0.041) | (0.063) | | |
| Adult Supports | 544 | -0.001 | -0.02 | -0.03 | 0.007 | 0.859 |
| | | | (0.083) | (0.125) | | |

Note: Data are from survey designed by research team and given to a randomly selected subsample of youth, proportional to overall treatment and control group size randomized into Study 2x2 during 2013. Unless otherwise noted with (-), the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroscedasticity-robust standard errors in parentheses. * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

## Table A5 – ISR Survey Questions, Wave 1

| QUESTIONS | RESPONSES |
| --- | --- |
| **AUTOMATICITY** | |
| It is important to think before you act. | *Yes; No* |
| To make a good decision, it is important to think. | *Yes; No* |
| Think of the last time an adult blamed you for doing something wrong. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| How much more carefully did you think about what to do? | *I spent a lot less time thinking about it this year than I would have last year; I spent a little less time thinking about it this year than I would have last year; I the same amount of time thinking about it this year as I would have last year; I spent a little more time thinking about it than I would have last year; I spent a lot more time thinking about it this year than I would have last year.* |
| Think of the last time you hit another person or nearly hit another person. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| Think of the last time an adult blamed you for doing something wrong. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| Think of the last time you encountered a roadblock that might have prevented you from doing something you set out to do. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| **SELF-DISTANCING** | |
| Please think about the last time you got really, really mad at someone else. How close were you to the person you got really mad at? Is this someone you feel very close to, or not so close to? | *Not close at all; really not close; not very close; somewhat close; moderately close; very close; extremely close* |
| As you are remembering this last time you got very mad at someone, how intense are the feelings you are experiencing? | *Not intense at all; really not intense; not very intense; somewhat intense; moderately intense; very intense; extremely intense* |

| | |
|---|---|
| As you are remembering this last time that you got very mad at someone, does it feel like you are right back there seeing the event through your own eyes, or does it feel like you are watching the event unfold through the eyes of someone else | *Exactly like I'm watching the event through my own eyes; a lot more like I'm watching the event through my own eyes; somewhat more like I'm watching the event through my own eyes; a mix of watching the event through my own eyes and through someone else's eyes; somewhat more like I'm watching the event through someone else's eyes; a lot more like I'm watching the event through someone else's eyes; exactly like I'm watching the event through someone else's eyes* |
| Now imagine that you faced that situation or a similar one again in the future. Please talk about how you would feel in the moment and how you would deal with the situation. | |
| How did you feel when you woke up yesterday morning? | *Rested; calm; excited; happy; hopeful; alert; worried; angry; scared; worried; upset; tired* |
| How sure are you that you felt that way? | *I'm completely guessing; I'm sort of guessing; I'm somewhat sure; I'm very sure* |
| **AWARENESS OF PAST ACTIONS** | |
| How often do you get into arguments? | *Rarely (1-3/year); occasionally (1-2/month); regularly (daily or 1-2/week)* |
| How sure are you that this is the number of arguments you get into? | *Not sure at all; somewhat sure; very sure; completely sure* |
| Have you ever fought anybody else? | *Yes; No* |
| How often? | *Rarely (1-3/year); occasionally (1-2/month); regularly (daily or 1-2/week)* |
| **LOCUS OF CONTROL** | |
| I don't have enough control over the direction my life is taking. | *Strongly agree; agree; disagree; strongly disagree* |
| Every time I try to get ahead, something or somebody stops me. | *Strongly agree; agree; disagree; strongly disagree* |
| In my life, good luck is more important than hard work for success. | *Strongly agree; agree; disagree; strongly disagree* |
| My plans hardly ever work out, so planning only makes me unhappy. | *Strongly agree; agree; disagree; strongly disagree* |
| When I make plans, I am almost certain I can make them work. | *Strongly agree; agree; disagree; strongly disagree* |
| Chance and luck are very important for what happens in my life | *Strongly agree; agree; disagree; strongly disagree* |

| GRIT | |
|---|---|
| New ideas and projects sometimes distract me from previous ones. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| Setbacks don't discourage me. I bounce back from disappointments faster than most people. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I have been obsessed with a certain idea or project for a short time but later lost interest. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I am a hard worker. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I often set a goal but later choose to follow a different one. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I have difficulty keeping my focus on projects that take more than a few months to complete. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I finish whatever I begin. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I am diligent (hard working and careful). | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| **CONSCIENTIOUSNESS** | |
| I am always prepared. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I continue until everything is perfect. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I leave a mess in my room. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| **PEER CONFLICT** | |
| **Vignette 1:** Let's imagine that you are talking with a girl in the hallway at school. You kind of like this person and seem to be getting along with her. You are just about to ask her to get together after school when another kid yells, "Fire!" and laughs. Everybody runs outside. It turns out to be a false alarm. But, you lose sight of the girl and don't get to ask her to get together. | |

| | |
|---|---|
| How likely is it that this happened to you because the kid who yelled “Fire!” was being mean to you or was playing a joke specifically on you so you wouldn’t get to talk to the girl? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that this happened to you because of some other reason that does not involve the other kid being mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| How worried would you be that you wouldn’t be able to find the girl if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? | *You’d want the kid who yelled “Fire!”: to like you OR; to respect you* |
| What would you do or say to the kid who yelled “Fire!” if this happened? | *Say “why did you do that?” OR; say “What IS your problem?!”* |
| **Vignette 2:** Imagine that you are walking down the street in a hurry to get to a friend’s house, and a police car slowly pulls up next to you. The policeman gets out of the car and says, “Hey, you. We just got a report from a gas station owner nearby who says that his store has been robbed. I want to talk to you about it.” | |
| How likely is it that the policeman is being mean to you or is thinking that you robbed the store? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the policeman questioned you because he thought you could help out with important information about the robbery? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How worried would you be that you would be arrested or taken to the police station if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? You’d want the policeman: | *To like you OR; to respect you* |
| What would you do or say to the policeman if this happened? | *Say “I don’t know anything about it” OR; say “it wasn’t me; mind your own business”* |
| **Vignette 3:** Imagine you are given a huge homework assignment by a particularly tough teacher. You work hard on it, complete it, and bring it to school in a book bag. When it comes time to turn it in, you look in the book bag, and it’s not there! You say to the teacher, “My homework is missing.” The teacher yells out in an angry voice, “Your homework is missing? Where is your homework?” | |

| | |
|---|---|
| How likely is it that the teacher said this to you because she doesn't trust you and was being mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the teacher thought someone else had taken your homework and that in fact you had completed the assignment? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| How worried would you be that you would have to do the assignment over if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? You'd want the teacher: | *To like you OR; to respect you* |
| What would you do or say to the teacher if this happened? | *Say "I put it in my bag. Someone must have taken it" OR; say "Someone must have taken it. I'm NOT doing it over!"* |
| **Vignette 4:** Imagine that you are sitting at your desk at school before class starts and another kid runs down the aisle past your desk. Your books get knocked off the desk onto the floor, making a mess. | |
| How likely is it that the other kid knocked over your books on purpose to be mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the other kid did not see your books and knocked them over by accident? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| How worried would you be that your stuff would be ruined if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? You'd want the kid: | *To like you OR; to respect you* |
| What would you do if this happened? | *Tell the kid to pick the books up OR; say "You BETTER pick them up" to the other kid* |
| **Vignette 5:** Imagine that some illegal drugs are found at your school, but you know absolutely nothing about it. The school principal sends a letter home to all the parents in the entire school, telling them that there is a drug problem at your school. That night at your home, just as you are about to go out, your parent reads the letter and yells out to you "Get in here. I have something to talk about with you." | |
| How likely is it that your parent believes that you are involved in the drug problem at school? | *Not at all likely; unlikely; unsure; likely; very likely* |

| | |
|---|---|
| How likely is it that your parent believes that you are <u>not involved</u> in this drug problem and just wants to talk about what's going on at school? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| How worried would you be that your parent was going to get upset with you if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? You'd want your parent: | *To like you OR; to respect you* |
| What would you do or say to your parent if this happened? | *Say "I'm not involved with drugs or with the people who are" OR; say "Get off my back!"* |
| **Vignette 6:** Imagine that you are at a park near your house, and you see a bunch of kids talking in a circle about 15 feet away. You yell out, "Hey, everybody!" These kids keep on talking and don't say anything to you. | |
| How likely is it that the other kids failed to answer you because they don't like you and were being mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the other kids did not hear you or did not answer for some other acceptable reason? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| How embarrassed would you be if this happened? | *Not at all; a little; somewhat; worried; very worried* |
| How would you want this situation to turn out in the end? You'd want the kids: | *To like you OR; to respect you* |
| What would you do or say to the other kids if this happened? | *Just go over and start talking OR; say "Don't talk to me then!"* |
| **SOCIAL NETWORKS** | |
| About how many close friends do you have these days? These are people you feel at ease or hang out with, can talk to about private matters, or call on for help. Would you say that you have no close friends, one or two, three to five, six to ten, or more than ten? | *[Record response]* |
| Among the close friends you hang out with, how important is it to: | |
| Attend classes regularly. | *Not at all important; not very important; somewhat important; very important* |
| Get good grades. | *Not at all important; not very important; somewhat important; very important* |

| | |
|---|---|
| Study. | *Not at all important; not very important; somewhat important; very important* |
| Continue their education past high school. | *Not at all important; not very important; somewhat important; very important* |
| In the last X month, have you stopped hanging around with anyone because you thought spending time with them was likely to put you in a situation that could lead to trouble? | *No, Yes* |
| In the last X months, have you **started** hanging around with anyone because you thought spending time with them was likely to keep you out of situations that could lead to trouble? | *No, Yes* |
| **EDUCATION AND SCHOOLING** | |
| Thinking about [your school/when you were last in school], in general, how much do you agree with the statement: Disruptions by other students [get/got] in the way of my learning. | *Strongly agree; agree; disagree; strongly disagree* |
| Overall about how much total time do you spend on homework each week, both in and out of school? | *Less than an hour; 1 to 5 hours; 6 to 10 hours; 11 to 14 hours; 15 hours or more* |
| When homework is assigned, how much of it do you usually complete? | *None of it; not very much of it; half of it; most of it; all of it* |
| As things stand now, how far in school do you think you will get? | *Won't finish high school; will graduate from high school, but won't go any further; will go to vocational, trade, or business school after high school; will attend college; will graduate from college; will attend a higher level of school after graduating from college* |
| As things stand now, how far in school do you want to go? | *Don't want to finish high school; want to graduate from high school, but not go any further; want to go to vocational, trade, or business school after high school; want to attend college; want to graduate from college; want to attend a higher level of school after graduating from college* |
| How important are good grades to you? | *Not important; somewhat important; important; very important* |
| This school year, how often did you feel safe at your school? | *Never; sometimes; often; always* |
| How true is each about you? | |
| I like math. | *Not at all true; a little bit true; mostly true; very true* |
| I get good grades in math. | *Not at all true; a little bit true; mostly true; very true* |
| In general, I like school. | *Not at all true; a little bit true; mostly true; very true* |

| | |
|---|---|
| I like reading. | *Not at all true; a little bit true; mostly true; very true* |
| I get good grades in English. | *Not at all true; a little bit true; mostly true; very true* |
| **ADULT SUPPORTS** | |
| How many adults do you have in your life who you feel comfortable talking to about personal problems? | *[Record response]* |
| How many adults do you have in your life who care a lot about how you turn out and who will help you if you get into trouble? | *[Record response]* |
| Who are the adult(s) who you go to first to talk about personal problems or who will help you if you get into trouble? | *Mother; father; stepparent; brothers or sisters; other relatives; teachers; coach; guidance counselor; advisor or school principal; other leaders in the community; no one* |
| Of all of the people you know personally, think about the person you admire the most. How would you describe this person? | *Honest; popular; dresses well; intelligent; makes a lot of money; has an important job; has a college degree; good at sports* |
| What is your relationship to that person? | *Friend; mother/father; relative; boyfriend/girlfriend; other* |
| What adult do you talk to when you need help with school work? | *Parent; adult relative; adult at school; other adult; no one* |

## Table A6 – ISR Survey Questions, Wave 2

| QUESTIONS | RESPONSES |
|---|---|
| **AUTOMATICITY** | |
| It is important to think before you act. | *Yes; No* |
| To make a good decision, it is important to think. | *Yes; No* |
| Think of the last time an adult blamed you for doing something wrong. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| Think of the last time you encountered a roadblock that might have prevented you from doing something you set out to do. How much do you agree with the statement "This year I handled the situation very differently than how I would have handled it last year?" | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| **SELF-DISTANCING** | |
| No matter how well two people get along, sometimes there are times when they get very mad at each other, so mad that they feel like they are going to explode. They might get annoyed about something the other person does, get into fights because they are in bad moods, or argue with each other. Take a few minutes right now to think about a time when you got very mad at someone. Try to remember a specific fight or argument that happened not too long ago and that still makes you mad when you think about it. | |
| When you thought about the fight a few moments ago, how much did it feel real or imagined? | *Very real; mostly real; somewhat real; neither real or imagined; somewhat imagined; mostly imagined; very imagined* |
| When you thought about the fight a few moments ago, how long ago did it feel like the fight happened? | *Right now; like just yesterday; a little while ago; a slightly longer while ago; a moderately long time ago; a long time ago; a very long time ago* |
| Now, please rate your current emotional state using these prompts: | |
| Thinking about the event still makes me feel upset (for example, angry, sad, hurt, rejected). | *Completely agree; agree; somewhat agree; neither agree nor disagree; somewhat disagree; disagree; completely disagree* |
| When I thought about the fight, I realized something that makes me think differently about why I felt the way I did. | *Completely agree; agree; somewhat agree; neither agree nor disagree; somewhat disagree; disagree; completely disagree* |

| | |
|---|---|
| When I thought about the fight, I realized something that made the fight bother me less. | *Completely agree; agree; somewhat agree; neither agree nor disagree; somewhat disagree; disagree; completely disagree* |
| When I thought about this fight, I still blamed the other person. | *Completely agree; agree; somewhat agree; neither agree nor disagree; somewhat disagree; disagree; completely disagree* |
| When I thought about this fight, I realized something that makes me forgive the person I fought with. | *Completely agree; agree; somewhat agree; neither agree nor disagree; somewhat disagree; disagree; completely disagree* |
| Think again about the last time you got in a fight. How much were you responsible? | *Completely my fault; almost all my fault; mostly my fault; somewhat my fault; a little my fault; barely my fault; not at all my fault* |
| How much was another person to blame? | *Completely to blame; almost all to blame; mostly to blame; somewhat to blame; a little to blame; barely to blame; not at all to blame* |
| How much could you change the way the whole thing went down? | *Could completely change it; could change almost all of it; could mostly change it; could somewhat change it; could change it a little; could barely change it; could not change it at all* |
| How long ago did this fight happen? | *[Response recorded]* |
| **GRIT** | |
| New ideas and projects sometimes distract me from previous ones. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| Setbacks don’t discourage me. I bounce back from disappointments faster than most people. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I have been obsessed with a certain idea or project for a short time but later lost interest. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I am a hard worker. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I often set a goal but later choose to follow a different one. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I have difficulty keeping my focus on projects that take more than a few months to complete. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| I finish whatever I begin. | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |

| | |
|---|---|
| I am diligent (hard working and careful). | *Very much like me; mostly like me; somewhat like me; not much like me; not like me at all* |
| **SENSATION SEEKING** | |
| I like to have new and exciting experiences even if they are a little frightening. | *True, False* |
| I like doing things just for the thrill of it. | *True, False* |
| I sometimes like to do things that are a little frightening. | *True, False* |
| I'll try anything once. | *True, False* |
| I sometimes do "crazy" things just for fun. | *True, False* |
| I like wild and "crazy" parties. | *True, False* |
| **NON- COGNITIVE PEER CONFLICT** | |
| **Vignette 1:** Imagine that you are walking down the street in a hurry to get to a friend's house, and a police car slowly pulls up next to you. The policeman gets out of the car and says, "Hey, you. We just got a report from a gas station owner nearby who says that his store has been robbed. I want to talk to you about it." | |
| How likely is it that the policeman is being mean to you or is thinking that you robbed the store? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the policeman questioned you because he thought you could help out with important information about the robbery? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| What would you do or say to the policeman if this happened? | *Say "I don't know anything about it"; OR say "It wasn't me; mind your own business."* |
| **Vignette 2:** Imagine you are given a huge homework assignment by a particularly tough teacher. You work hard on it, complete it, and bring it to school in a book bag. When it comes time to turn it in, you look in the book bag, and it's not there! You say to the teacher, "My homework is missing." The teacher yells out in an angry voice, "Your homework is missing? Where is your homework?" | |
| How likely is it that the teacher said this to you because she doesn't trust you and was being mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the teacher thought someone else had taken your homework and that in fact you had completed the assignment? | *Not at all likely; unlikely; unsure; likely; very likely* |

| | |
|---|---|
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| What would you do or say to the teacher if this happened? | *Say "I put it in my bag. Someone must have taken it" OR 2- Say "Someone must have taken it. I'm NOT doing it over!"* |
| **Vignette 3:** Imagine that you are sitting at your desk at school before class starts and another kid runs down the aisle past your desk. Your books get knocked off the desk onto the floor, making a mess. | |
| How likely is it that the other kid knocked over your books on purpose to be mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the other kid did not see your books and knocked them over by accident? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| What would you do if this happened? | *Tell the kid to pick the books up; Say "You BETTER pick them up" to the other kid.* |
| **Vignette 4:** Imagine that you are at a park near your house, and you see a bunch of kids talking in a circle about 15 feet away. You yell out, "Hey, everybody!" These kids keep on talking and don't say anything to you. | |
| How likely is it that the other kids failed to answer you because they don't like you and were being mean to you? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How likely is it that the other kids did not hear you or did not answer for some other acceptable reason? | *Not at all likely; unlikely; unsure; likely; very likely* |
| How angry would you be if this happened? | *Not at all; a little; somewhat; angry; very angry* |
| What would you do or say to the other kids if this happened? | *Just go over and start talking; Say "Don't talk to me then!"* |
| **HOW I THINK** | |
| How much would you say you agree or disagree with the following statement:<br>When I get mad, I don't care who I hurt. [egocentric bias] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| When I lose my temper, it's because people try to make me mad [blaming others] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| If I made a mistake, it's because I got mixed up in the wrong crowd [blaming others] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| Only a coward would ever walk away from a fight [mislabeling] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |

| | |
|---|---|
| You can't trust people because they will always lie to you [assuming the worst] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| It's no use trying to stay out of fights [assuming the worst] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| People are always trying to hassle me [assuming the worst] Do you… | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| **CONSCIENTIOUSNESS** | |
| I see myself as someone who does things carefully and completely. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who can be somewhat careless. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who is a reliable worker. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who tends to be disorganized. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who tends to be lazy. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who keeps working until things are done. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who does things efficiently (quickly and correctly). | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who makes plans and sticks to them. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| I see myself as someone who is easily distracted; has trouble paying attention. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| **EDUCATION AND SCHOOLING** | |
| As things stand now, how far in school do you *want* to go? | *Don't want to finish high school; Want to graduate from high school, but not go any further; Want to go to vocational, trade, or business school after high school; Want to attend college; Want to graduate from college; Want to attend a higher level of school after graduating from college* |
| How far in school you *expect* to go? | *Expect to graduate from high school, but not go any further; Expect to go to vocational, trade, or business school after high school; Expect to attend college; Expect to graduate from college; Expect to attend a higher level of school after graduating from college* |

| | |
|---|---|
| Adults in my school really care about me. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| Adults in my school are often too busy to give students extra help. | *Strongly disagree; disagree; neither agree nor disagree; agree; strongly agree* |
| How often have you talked to an adult at school about something outside of school that is important to you? | *Never; 1 or 2 times; 3 or 4 times; 4 or more times* |
| Overall about how much total time do you spend on homework each week, both in and out of school? | *Less than an hour; 1 to 5 hours; 6 to 10 hours; 11 to 14 hours; 15 hours or more* |
| How important are good grades to you? | *Not Important; somewhat Important; important; very important* |
| This school year, how often do you feel safe at your school? | *Never; sometimes; often; always* |
| In general, I like school. | *Not at all true; a little bit true; mostly true; very true* |
| I like reading. | *Not at all true; a little bit true; mostly true; very true* |
| I get good grades in English. | *Not at all true; a little bit true; mostly true; very true* |
| I like math. | *Not at all true; a little bit true; mostly true; very true* |
| I get good grades in math. | *Not at all true; a little bit true; mostly true; very true* |
| How many days of school did you miss over the past four weeks, that is over the last 20 school days? | *[Response recorded]* |
| How often do you cut or skip classes? | *Never or almost never; sometimes, but less than once a week; not every day, but at least once a week; daily* |
| How many times were you late for school over the past four weeks, that is over the last 20 school days? | *[Response recorded]* |
| I'm certain I can master the skills being taught in math. | *Strongly disagree; disagree; agree; strongly agree* |
| I'm certain I can master the skills being taught in English. | *Strongly disagree; disagree; agree; strongly agree* |
| **SOCIAL NETWORKS** | |
| About how many CLOSE FRIENDS do you have these days? These are people you feel at ease or hang out with, can talk to about private matters, or call on for help. Would you say that you have no close friends, one or two, three to five, six to ten, or more than ten? | *[Response recorded]* |

| | |
|---|---|
| Among the close friends you hang out with, how important is it to: attend classes regularly? | *Not at all important; not very important; somewhat important; very important* |
| Among the close friends you hang out with, how important is it to: get good grades? | *Not at all important; not very important; somewhat important; very important* |
| Among the close friends you hang out with, how important is it to: study? | *Not at all important; not very important; somewhat important; very important* |
| Among the close friends you hang out with, how important is it to: continue their education past high school? | *Not at all important; not very important; somewhat important; very important* |
| Since the beginning of Grade X, have you stopped hanging around with anyone because you thought spending time with them was likely to put you in a situation that could lead to trouble? | *Yes; No* |
| Since the beginning of Grade X, have you started hanging around with anyone because you thought spending time with them was likely to keep you out of situations that could lead to trouble? | *Yes; No* |
| Since the beginning of Grade X, have you started hanging around with anyone because you thought they would help you do better in school? | *Yes; No* |
| **ADULT SUPPORTS** | |
| How many adults do you have in your life who you feel comfortable talking to about personal problems? | *[Response recorded]* |
| How many adults do you have in your life who care a lot about how you turn out and who will help you if you get into trouble? | *[Response recorded]* |
| Do you live with your father or someone you consider to be like a father figure to you? | *[Response recorded]* |
| Have you seen your father in the past month? | *[Response recorded]* |
| **GROWTH MINDSET** | |
| You can learn new things, but you can't really change your basic intelligence. | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |
| Your intelligence is something about you that you can't change very much. | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |

| | |
|---|---|
| You have a certain amount of intelligence and you really can't do much to change it. | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |
| A person's moral character is something very basic about them and it can't be changed much. | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |
| Whether a person is responsible and sincere or not is deeply ingrained in their personality. It cannot be changed much." | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |
| There is not much that can be done to change a person's moral traits (e.g., conscientiousness, uprightness, and honesty). | *Strongly disagree; disagree; somewhat disagree; somewhat agree; agree; strongly agree* |
| **SUBJECTIVE EXPECTATIONS** | |
| What is the percent chance that you will be a student in a regular school one year from now? | *[Response recorded]* |
| What is the percent chance that you will get someone pregnant within the next year? | *[Response recorded]* |
| What is the percent chance that you will be arrested, whether rightly or wrongly, at least once in the next year? | *[Response recorded]* |
| What is the percent chance that you will have received a high school diploma by the time you turn 20? | *[Response recorded]* |
| **MENTAL HEALTH** | |
| How much of the time during the past month have you felt:<br>1. So sad that nothing could cheer you up?<br>2. Nervous?<br>3. Restless or fidgety?<br>4. Hopeless?<br>5. That everything was an effort?<br>6. Worthless?<br>7. Calm and peaceful? | *All of the time; most of the time; some of the time; a little of the time; none of the time* |
| The last six questions asked about feelings that might have occurred during the past 30 days. Taking them altogether, how often did this feelings occur? | *More often in the past 30 days than is usual for you, about the same as usual, or less often than usual?* |
| During the past 30 days, how many days out of 30 were you totally unable to work or carry out your normal activities because of these feelings? | *[Response recorded]* |

| | |
|---|---|
| Not counting the days you reported in response to the question above, how many days in the past 30 were you able to do only *half or less* of what you would normally have been able to do, because of these feelings? | *[Response recorded]* |
| During the past 30 days, how many times did you see a doctor or other health professional about these feelings? | *[Response recorded]* |
| During the past 30 days, how often have physical health problems been the main cause of these feelings? | *All of the time; most of the time; some of the time; a little of the time; none of the time* |
| **CRIME VICTIMIZATION** | |
| During the past 12 month, how often did the following thing happen: someone pulled a knife or gun on you? | *Never; once; more than once; refused; don't know; not applicable* |
| During the past 12 month, how often did the following thing happen: you got into a physical fight? | *Never; once; more than once; refused; don't know; not applicable* |
| During the past 12 month, how often did the following thing happen: you were jumped: you were beaten up and something was stole from you? | *Never; once; more than once; refused; don't know; not applicable* |
| **RISKY/DELINQUENT BEHAVIOR** | |
| During your life, how many days have you had at least one drink of alcohol? | *0 days; 1 or 2 days; 3 to 9 days; 10 to 19 days; 20 to 39 days; 40 to 99 days; 100 or more days* |
| During the past 30 days, on how many days did you have at least one drink of alcohol? | *0 days; 1 or 2 days; 3 to 5 days; 6 to 9 days; 10 to 19 days; 20 to 29 days; All 30 days* |
| During your life, how many times have you used marijuana? | *0 days; 1 or 2 days; 3 to 9 days; 10 to 19 days; 20 to 39 days; 40 to 99 days; 100 or more days* |
| During the past 30 days, how many times did you use marijuana? | *0 days; 1 or 2 days; 3 to 5 days; 6 to 9 days; 10 to 19 days; 20 to 29 days; All 30 days* |
| During your life, how many times have you tried any other sort of illegal drug, or sniffed glue, or inhaled any paints or sprays to get high, or used any prescription drugs? | *0 days; 1 or 2 days; 3 to 9 days; 10 to 19 days; 20 to 39 days; 40 to 99 days; 100 or more days* |
| Do any of your brothers, sisters, cousins, or friends belong to a gang? | *[Response recorded]* |
| Do you belong to a gang? | *[Response recorded]* |
| Have you ever sold marijuana or any other drug to your friends? | *[Response recorded]* |
| How about to people you didn't know? | *[Response recorded]* |

| | |
|---|---|
| Have you ever had sexual intercourse? | *[Response recorded]* |
| During the past 3 months with how many people did you have sexual intercourse? | *[Response recorded]* |
| The last time you had sexual intercourse, did you or your partner use a condom? | *[Response recorded]* |
| How many times have you gotten someone pregnant? | *[Response recorded]* |
| In the past 12 months, how many times did you get in a physical fight in which you were so badly injured that you were treated by a doctor or a nurse? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| In the past 12 months, how often did you hurt someone badly enough in a physical fight that he or she needed to be treated by a doctor or nurse? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| During the past 30 days, on how many days did you carry a weapon – such as a gun, knife, or club – to school? | *None; 1 day; 2 or 3 days; 4 or 5 days; 6 or more days; refused; don't know; not applicable* |
| In the past 12 months, how often did you: paint graffiti or signs on someone else's property or in a public place? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| In the past 12 months, how often did you: deliberately damage property that didn't belong to you? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| In the past 12 months, how often did you: take something from a store without paying for it? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| In the past 12 months, how often did you: drive a car without owner's permission? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |
| In the past 12 months, how often did you: break into someone's home in order to steal? | *Never; 1 or 2 times; 3 or 4 times; 5 or more times; refused; don't know; not applicable* |

Effect of BAM on School Engagement Index Outcomes in Each Study
First and Second Year Effects

Figure A1: School Engagement Index (First-Year Effect)

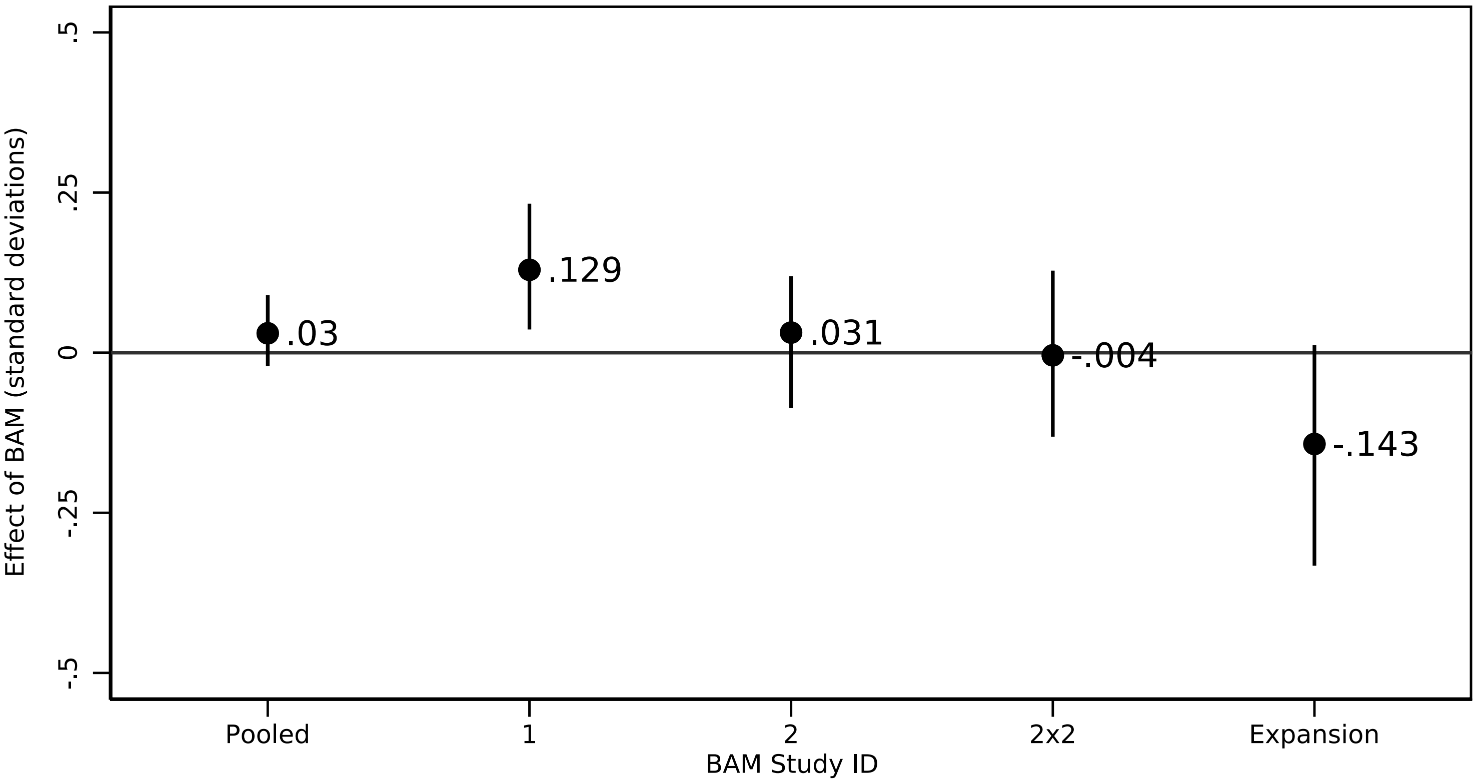
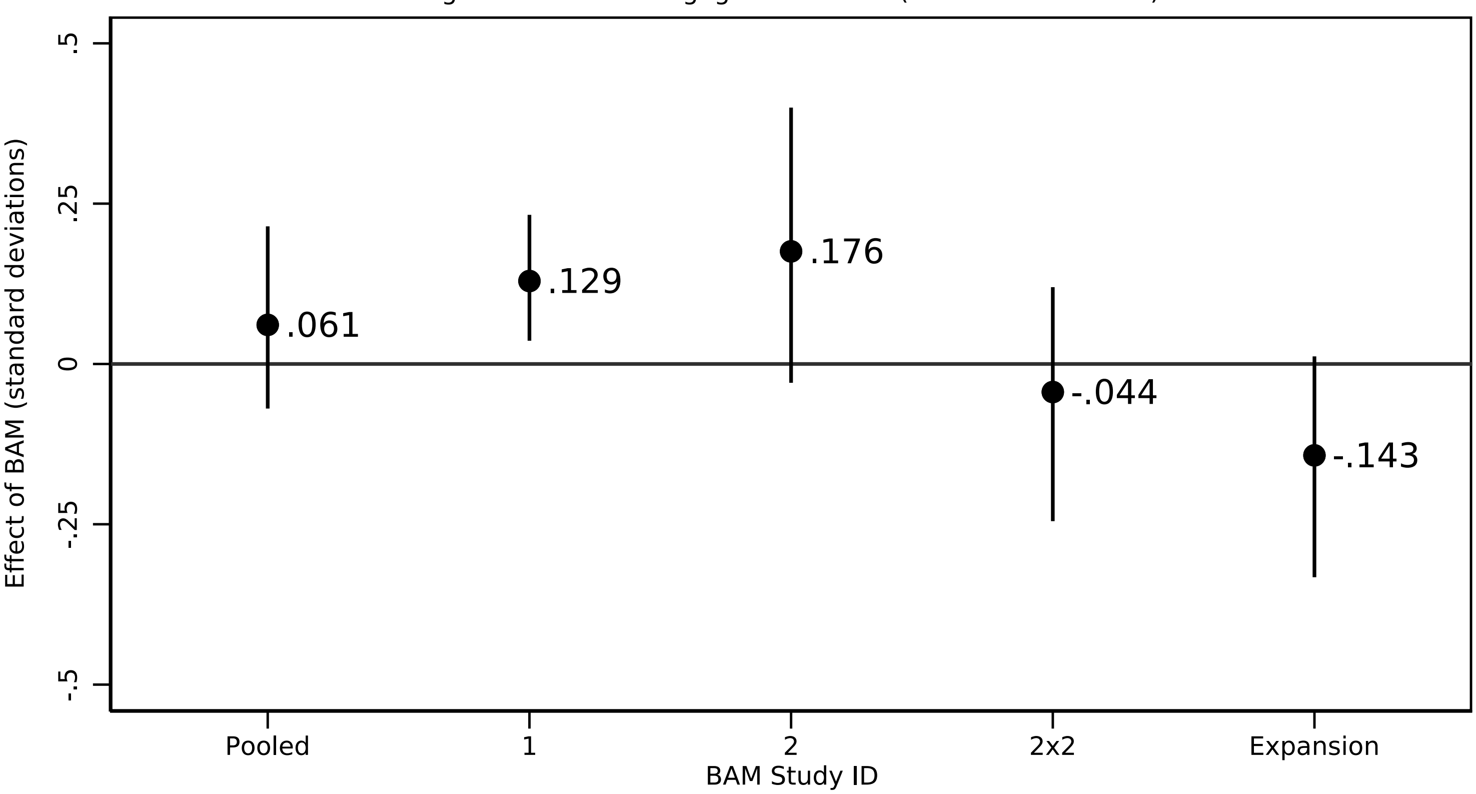



Figure A2: School Engagement Index (Second-Year Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in standard deviations, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).

Effect of BAM on All Arrests Outcomes in Each Study
First and Second Year Effects

Figure A3: All Arrests (First-Year Effect)

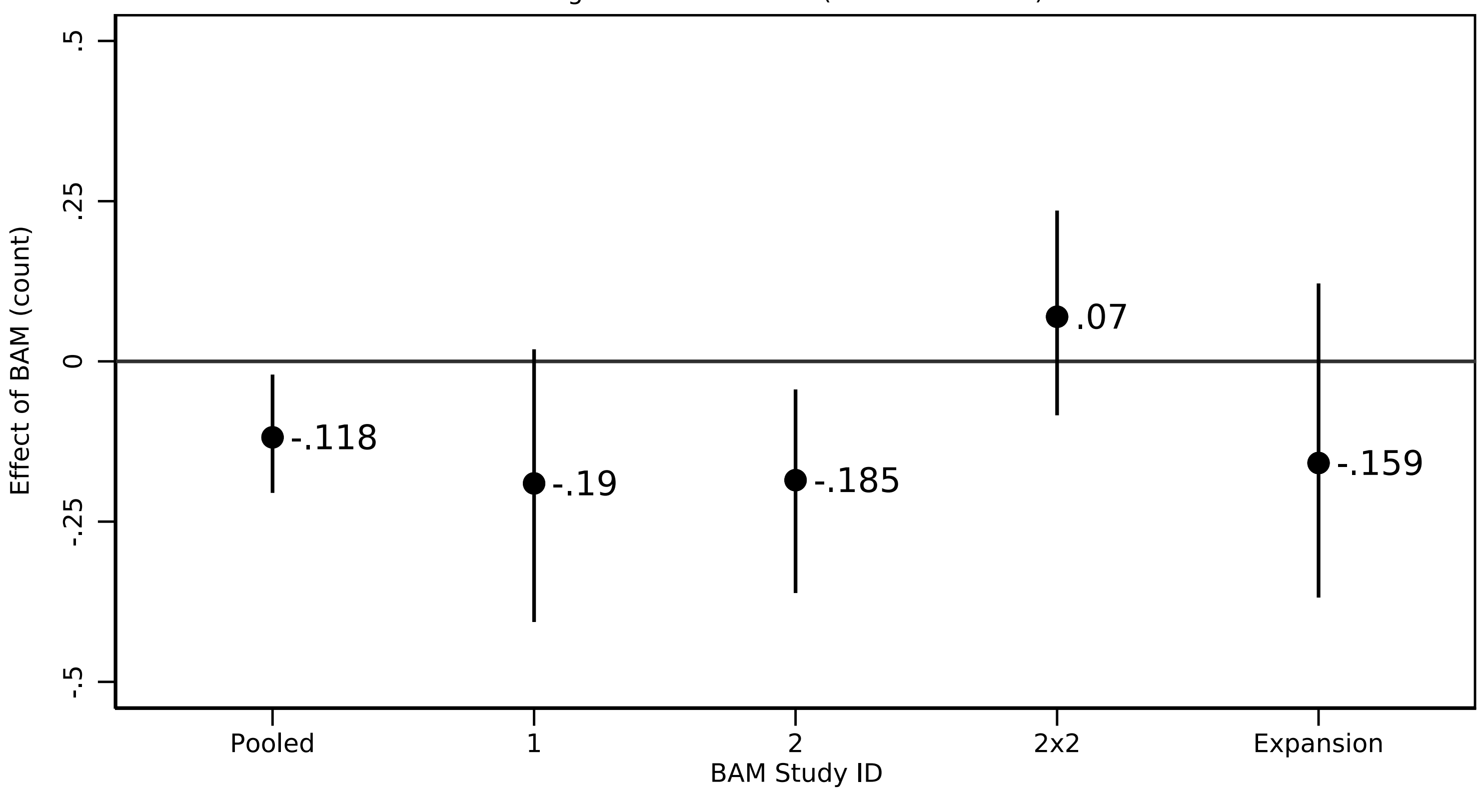
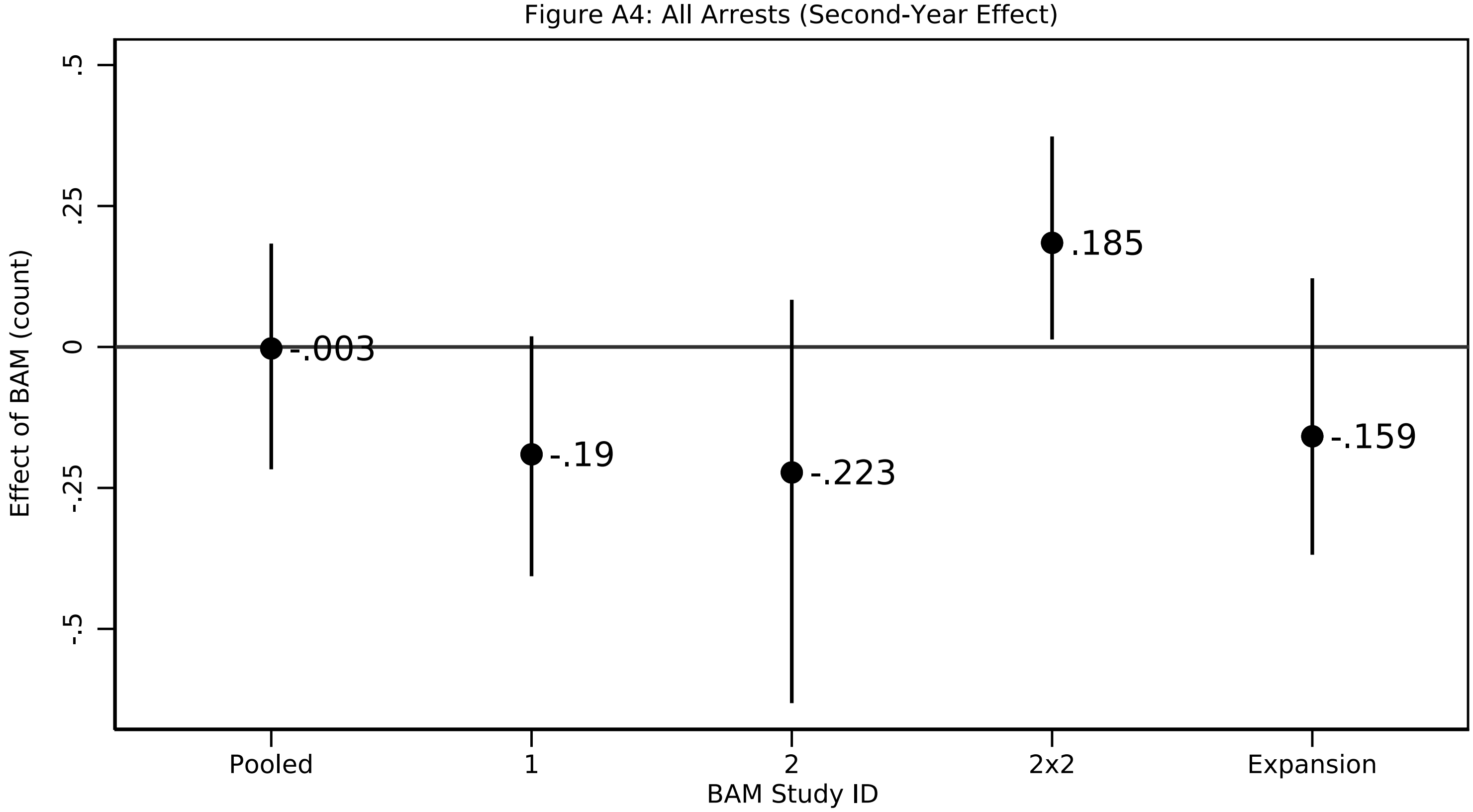



Figure A4: All Arrests (Second-Year Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).

Effect of BAM on Violent Arrests Outcomes in Each Study
First and Second Year Effects

Figure A5: Violent Arrests (First-Year Effect)

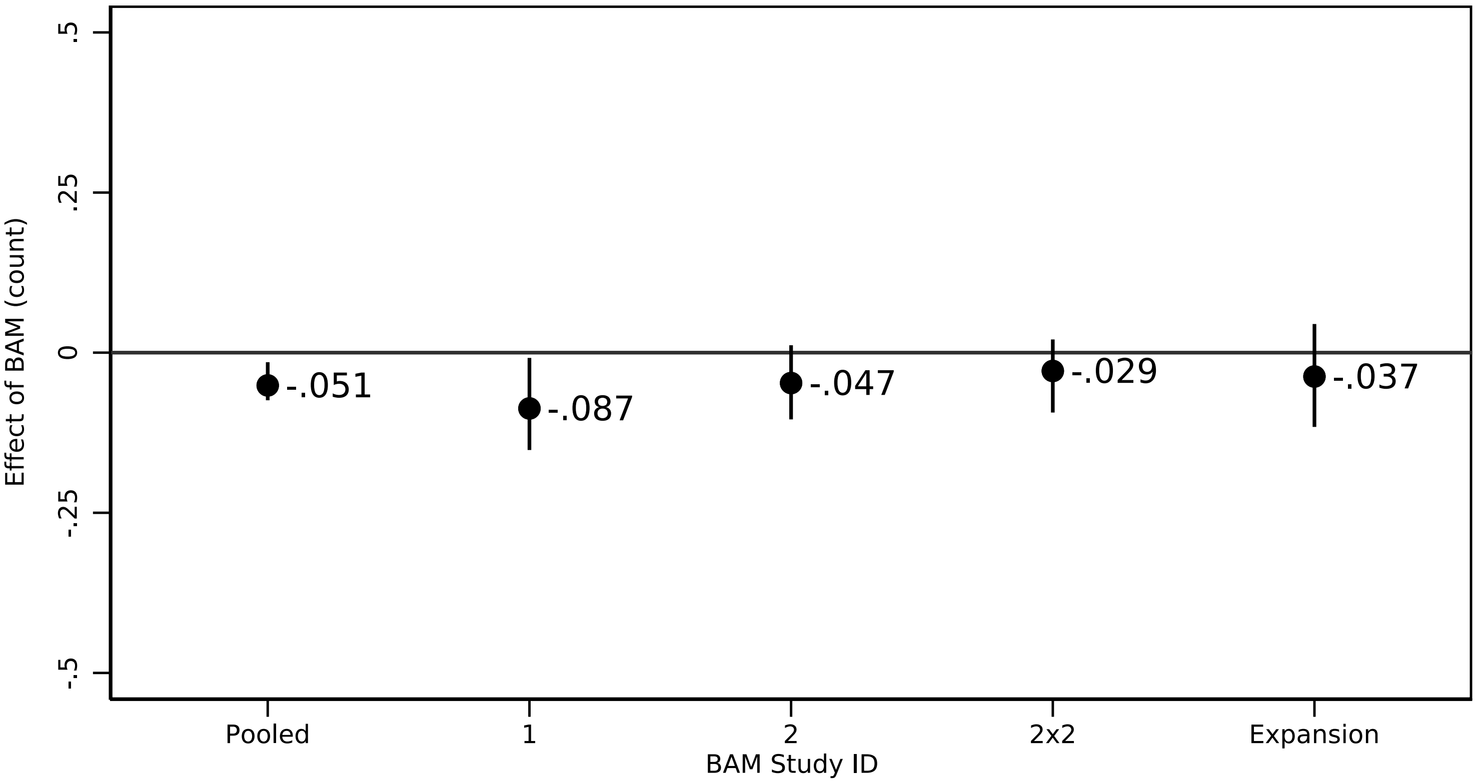
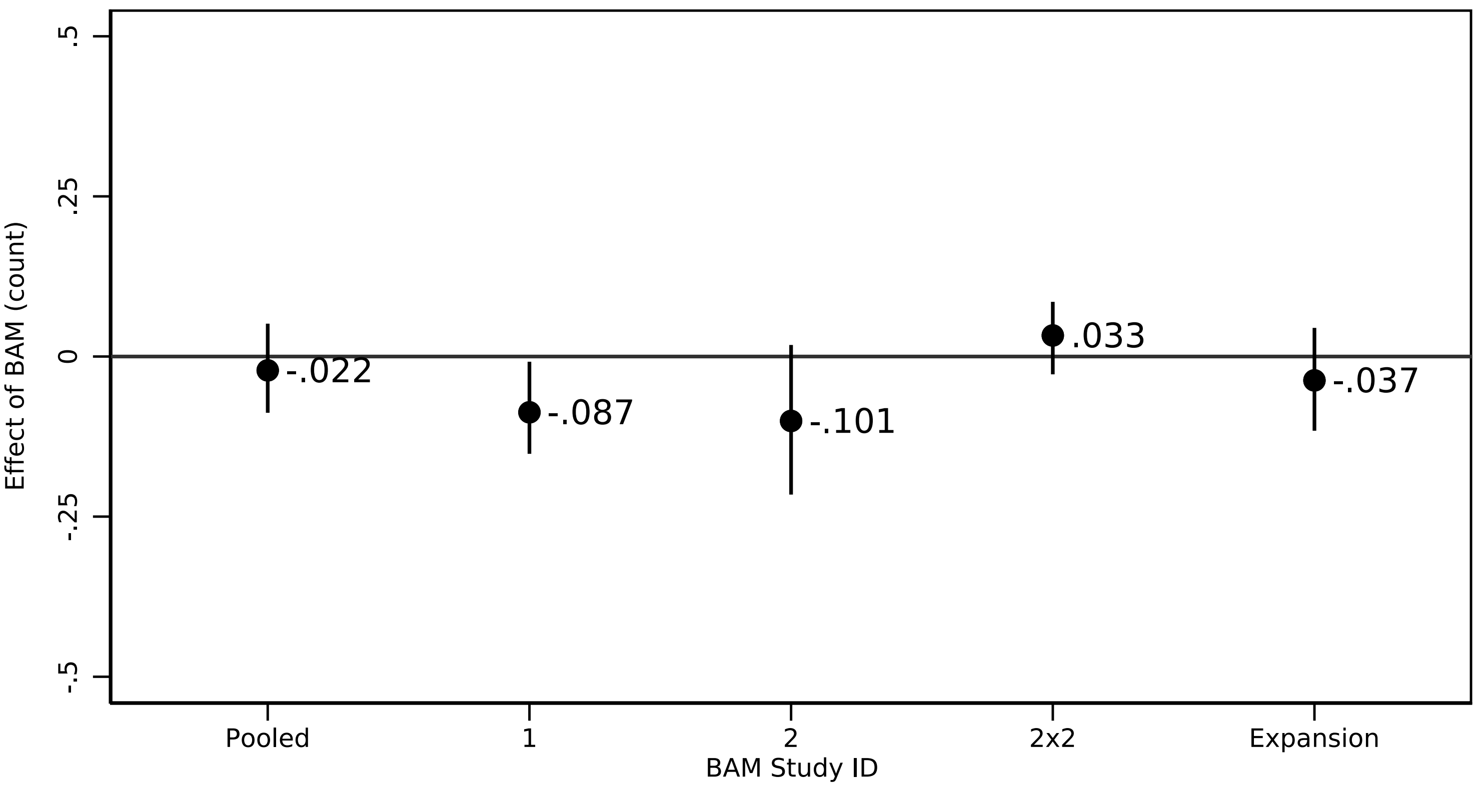



Figure A6: Violent Arrests (Second-Year Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).

Effect of BAM on Property Arrests Outcomes in Each Study
First and Second Year Effects

Figure A7: Property Arrests (First-Year Effect)

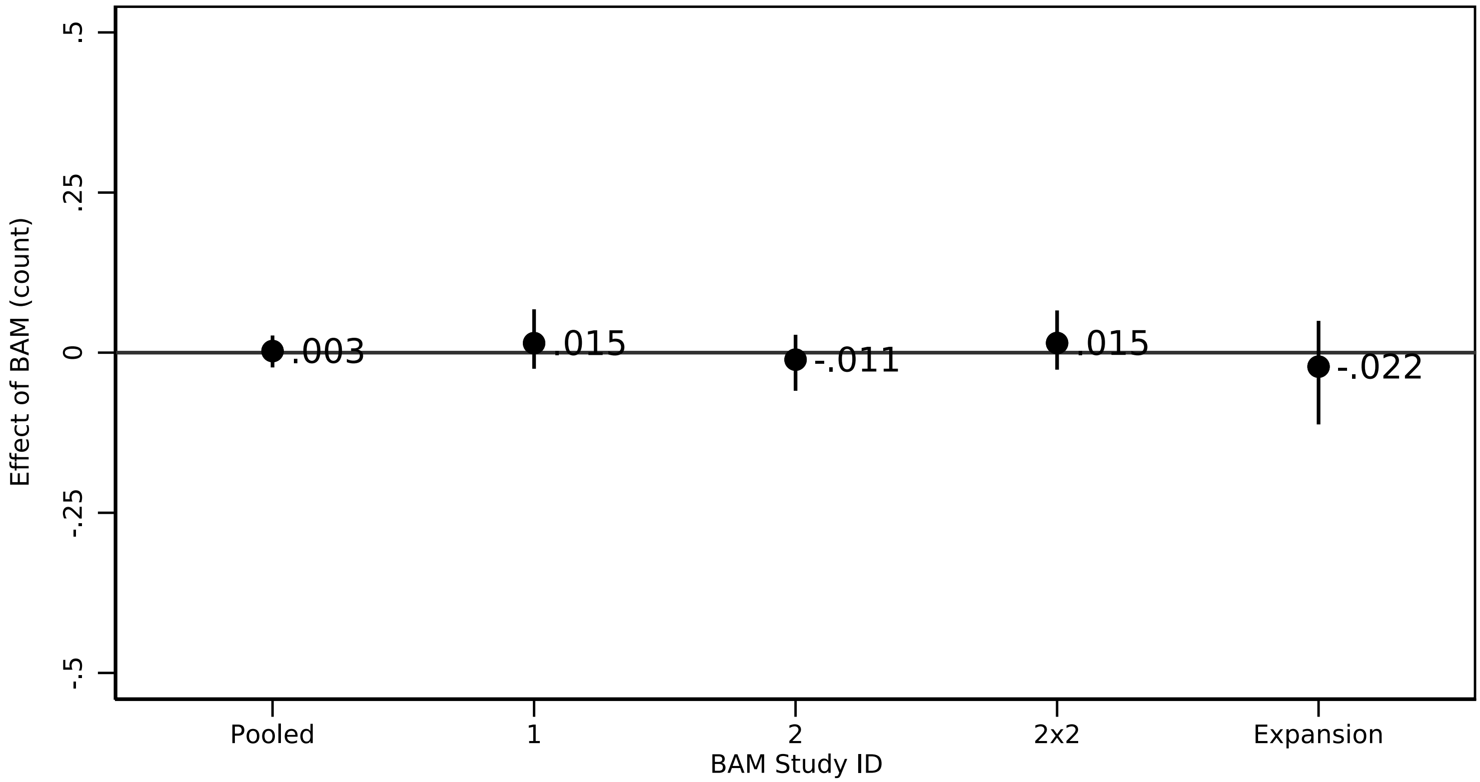
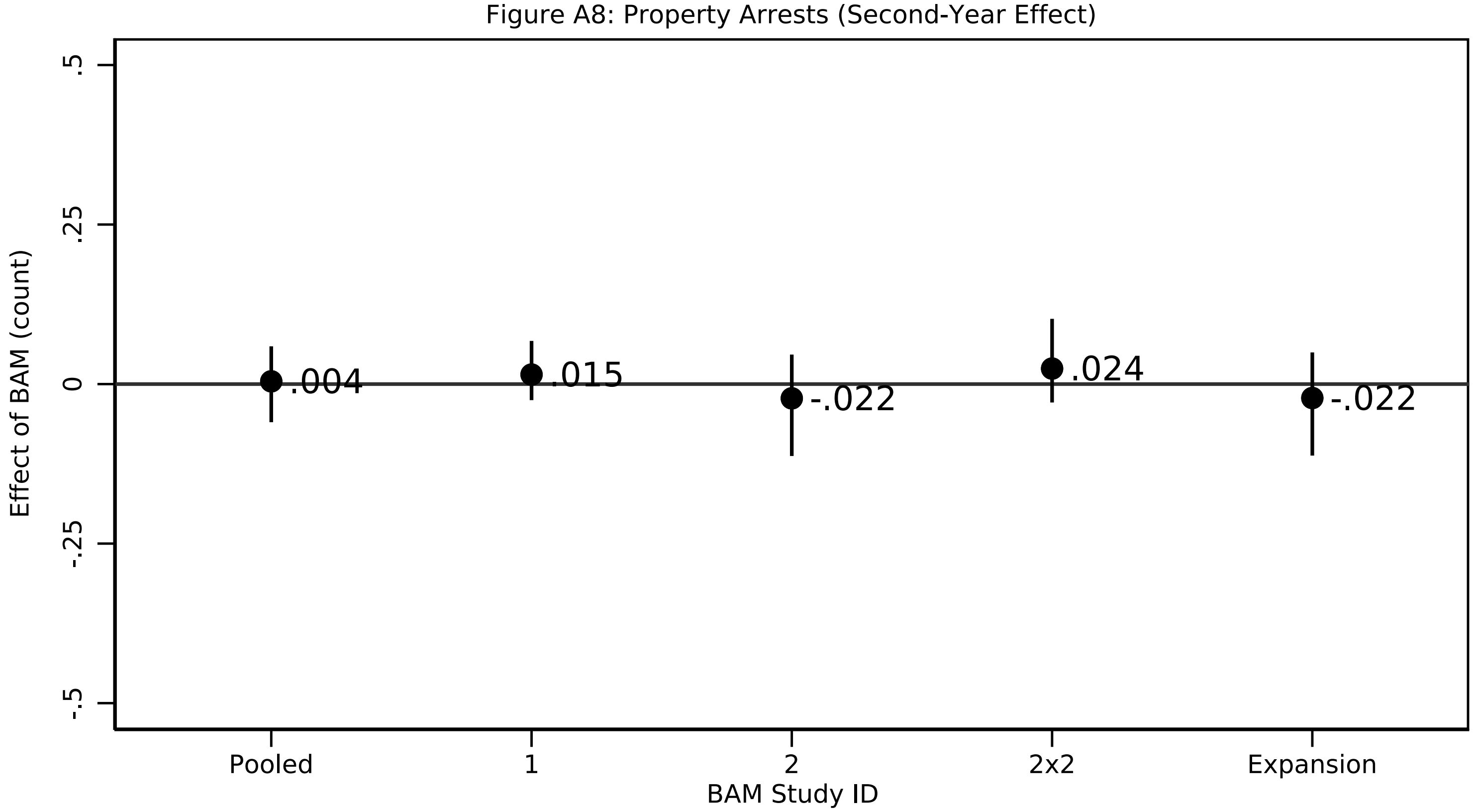
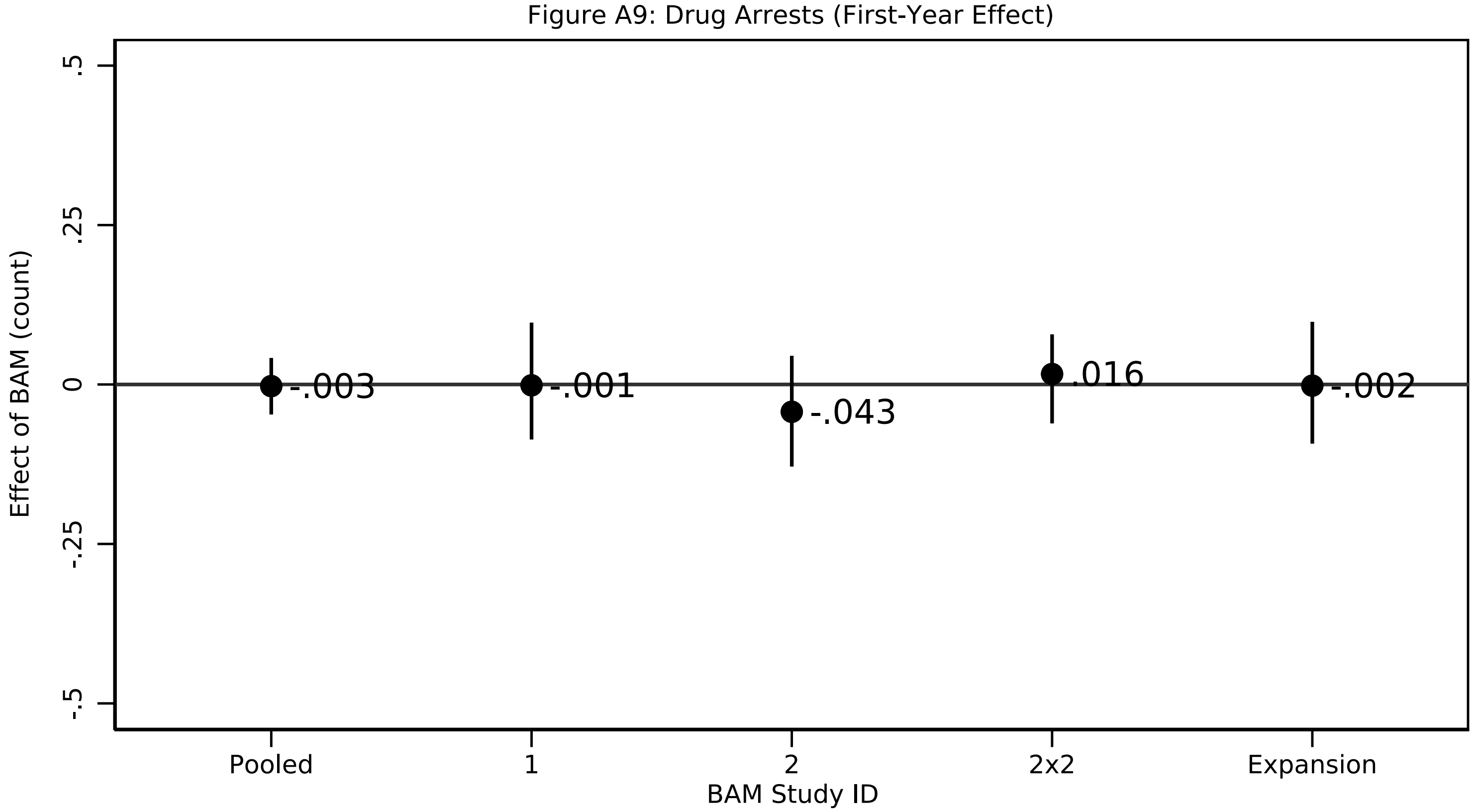


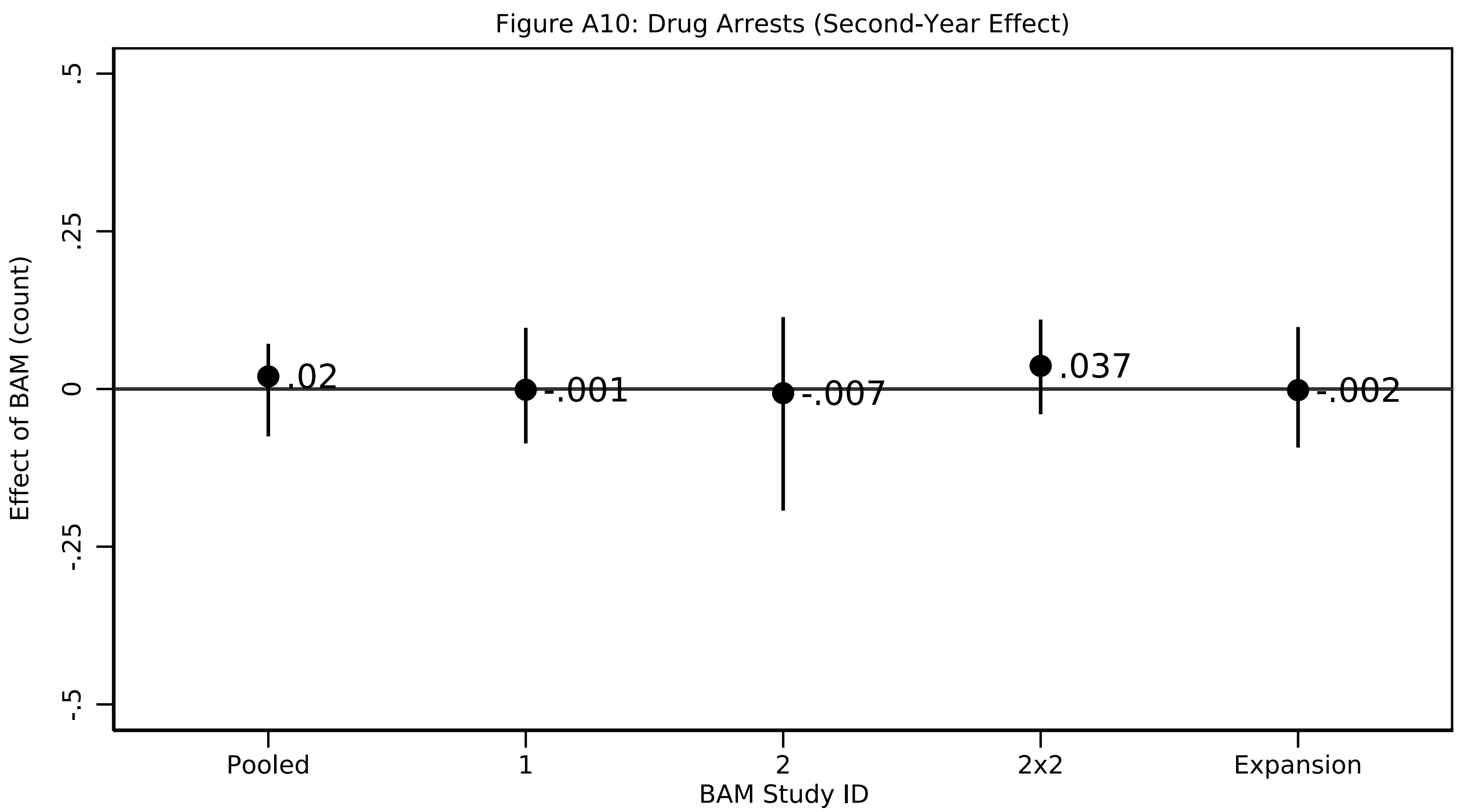

Figure A8: Property Arrests (Second-Year Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).

Effect of BAM on Drug Arrests Outcomes in Each Study
First and Second Year Effects




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).

Effect of BAM on Other Arrests Outcomes in Each Study
First and Second Year Effects

Figure A11: Other Arrests (First-Year Effect)

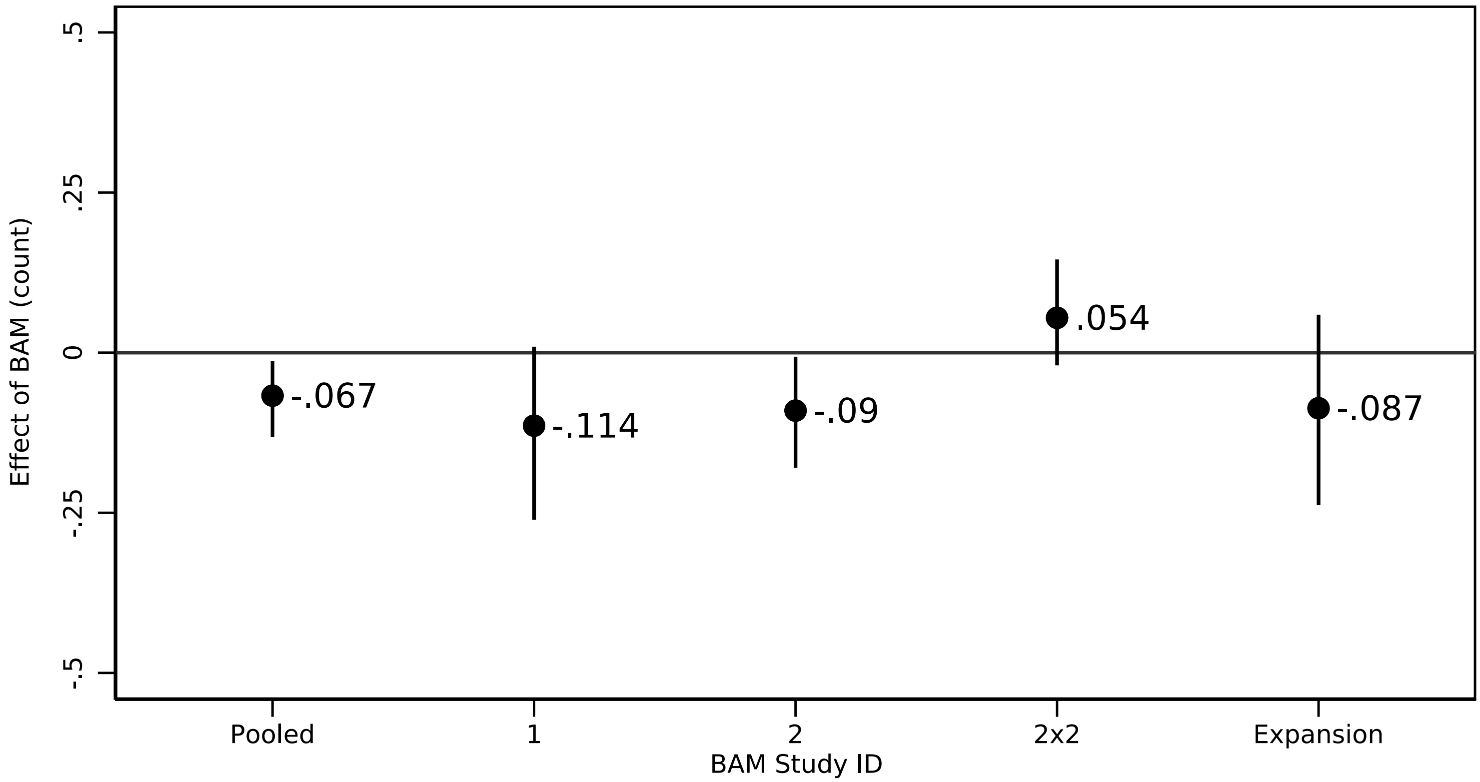
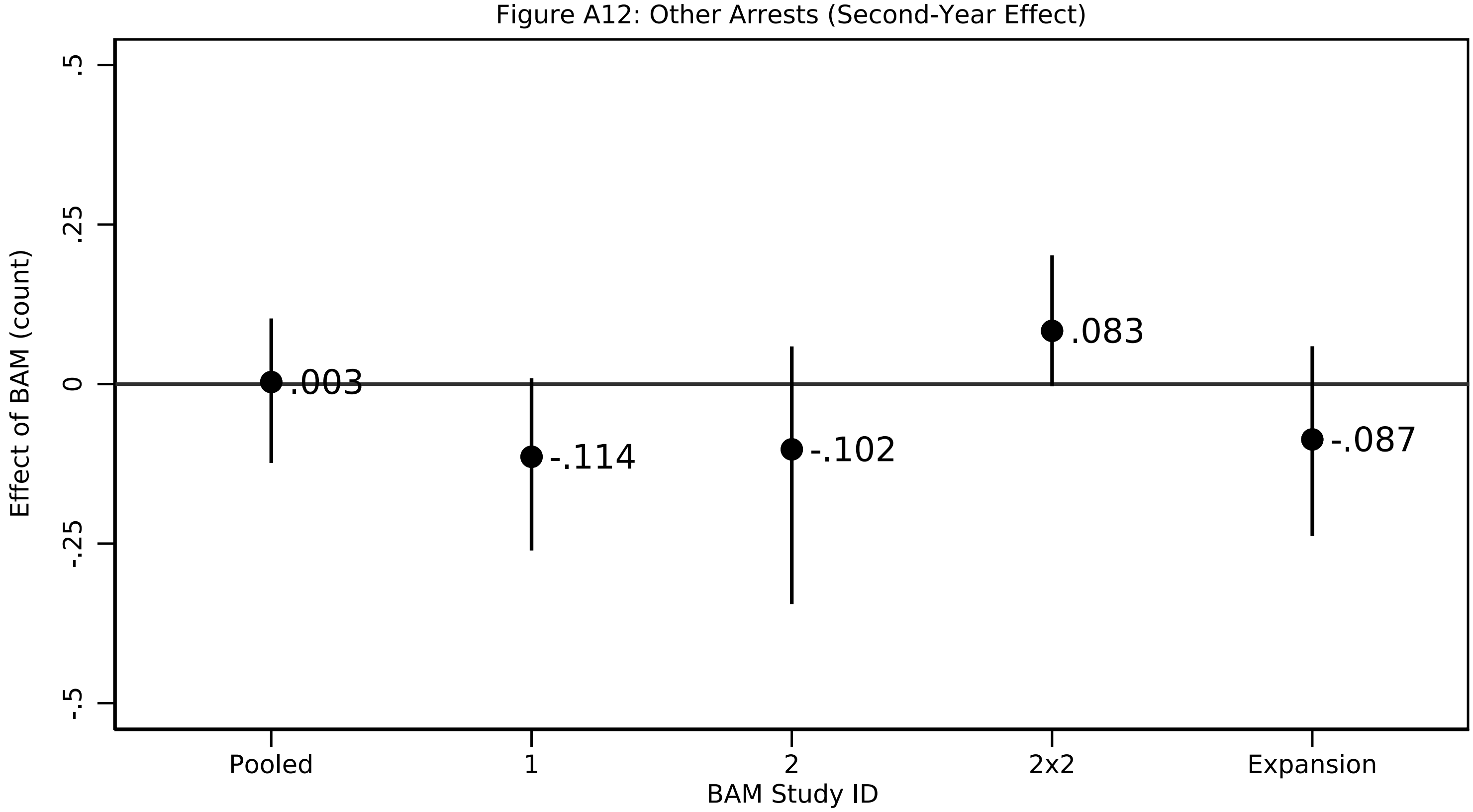



Figure A12: Other Arrests (Second-Year Effect)




Note: Point estimates are 50th percentile bootstrap estimates measured in arrest counts, from our different study samples. 95% confidence intervals are the 2.5th and 97.5th percentiles of a set of 1000 bootstrapped estimates, sampled with replacement from within randomization block groups. Standard baseline covariates and randomization block fixed effects are included in each model. The first-year effect measures outcomes from Year1 of all studies. The second-year effect measures outcomes from year 2 of Study 2 and Study 2x2 and omits Study 1 and Study Expansion (see text).