CASE 1: COGNITIVE BEHAVIORAL THERAPY IN CHICAGO SCHOOLS

Theory of Change and Measuring Outcomes

Young men participate in an activity for the Becoming a Man program in Chicago.
Photo: Rob Kozloff | University of Chicago

This case study is based on “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago” by Heller et. al. (2017), Quarterly Journal of Economics.

J-PAL thanks the authors for allowing us to use their paper as a teaching tool.
### Key Vocabulary

| **Hypothesis** | A proposed explanation of and for the effects of a given intervention. We can think of this as a claim to be tested. Hypotheses are intended to be made prior to the implementation of the intervention. e.g. *Giving textbooks to students will improve student learning.* |
| **Theory of Change** | Describes a strategy or blueprint for achieving a given long-term goal. The theory of change identifies the preconditions, pathways, and interventions necessary for success. |
| **Logical Framework** | A management tool used to facilitate the design, execution, and evaluation of an intervention. It involves identifying the inputs, outputs, outcomes, indicators and impact of an intervention. A logical framework is used to establish the causal relationships between these elements and the assumptions and risks that may influence the success and failure of the intervention. |
| **Assumption** | A precondition that underpins a theory of change or model. An assumption cannot be directly observed or verified e.g. *When students read textbooks, they learn from them.* |
| **Input** | An activity carried out as part of a program or intervention e.g. *Textbooks are given to schools.* |
| **Output** | A step in the planned implementation of a program or intervention—a.k.a. a direct result in response to the inputs e.g. *Students receive textbooks through schools.* |
| **Outcome** | A change or impact caused by the program that is being evaluated e.g. *Increase in student learning levels.* |
| **Intermediate Outcomes** | Observable changes or impacts caused by the program that are not the ultimate outcome of interest, but necessary along the way to achieving a final outcome e.g. *Increase in students who have passing test scores for the semester.* |
| **Final Outcomes** | Changes or impacts that are of ultimate interest to researchers and/or program implementers; these are often the overall goals of a program e.g. *Increase in high school graduation rates.* |
| **Indicator** | An observable metric used to measure an outcome e.g. *Student test scores.* |
| **Instrument** | The tool used to measure an indicator e.g. *A set of test questions.* |

---

1 These definitions of hypothesis, theory of change, and logical framework are based on those in Module 5.1: Theory of Change from Glennerster and Takavarasha’s *Running Randomized Evaluations*. [http://runningres.com/case-studies](http://runningres.com/case-studies)
LEARNING OBJECTIVE

To better understand measurement and the theory of change concept. Specifically, to understand the importance of theory of change in deciding what research questions to ask, what data to collect, and what outcomes to measure.

SUBJECTS COVERED

Theory of change, defining a hypothesis, selecting indicators, measuring outcomes, and measuring the impact of a program or policy.

INTRODUCTION

Violence and crime often have disproportionate harmful effects on young men in low-income settings. Violence, crime, and the root causes of violence and crime can exacerbate risk and uncertainty within communities, limiting safety, prosperity, and quality of life. Systemic factors shape the risks young men face, the choices and tradeoffs they must make, and their way of thinking, reacting, and forming decision-making strategies. While young adults in any context may struggle to develop a positive identity and strategies for responding to stress and threats, those who grow up in low-income or violent settings typically face more challenges and receive less support than their peers who grow up in higher-income or less violent settings.

Violent crime can arise from smaller, non-violent altercations that escalate. Chicago police attributed 75 percent of homicides in the city in 2015 and 2016 to altercations that escalated.2

This case study will look at a program aimed at reducing violent crime and improving educational outcomes among male youth in Chicago from low-income households through the use of cognitive behavioral therapy (CBT) in Chicago Public Schools and Juvenile Detention Centers. CBT is a short-term, goal-oriented psychotherapy treatment that takes a hands-on, practical approach to problem solving.3 Its goal is to introduce skills and change patterns of thinking or behavior, thereby changing the way individuals feel and respond.

Note: This case study focuses on specific behavioral mechanisms that contribute to violent behavior and escalation, which are the focus of the original study on which this case is based. Because the original study does not focus on societal factors such as race, socioeconomic status, or geographical location or on systemic barriers or inequities, this case similarly does not focus on these factors.

DISCUSSION TOPIC 1

1. What are some possible contributing factors that might explain why some youth in Chicago engage in criminal or violent activity?

2. Review your list. Can you suggest interventions that might reduce or mitigate these contributing factors? What are other policy options for reducing criminal activity among youth?
COGNITIVE BEHAVIORAL THERAPY

Cognitive behavioral therapy (CBT) is a comprehensive intervention to reduce self-destructive behaviors by teaching individuals to reflect upon and modify their thought processes, update their perceptions and decisions, and better understand situational factors. Rather than defining and teaching “good” behaviors, or nudging participants towards certain choices, CBT teaches concrete methods for better assessing one’s environment and practicing strategies to first recognize, and then regulate, harmful and automatic behaviors. In the terminology of Daniel Kahneman, humans make “automatic responses” when they “think fast” (System 1 thought) rather than going through reflective, slower responses (“thinking slow,” or System 2 thought.) System 1 “fast” responses can sometimes result in helpful automatic decisions (such as jumping out of the way of a swerving car). However, growing up in a violent neighborhood can shape responses and lead to situations where automatic responses are seen as misbehavior in school and/or pose risks or dangers. CBT guides participants to “think about their thinking” as a way to slow automatic responses and assess whether a response would fit the situation at hand.

Since the 1980s, CBT has been used as a mental health treatment, including as a treatment for depression. Today, policymakers around the world are increasingly interested in adapting CBT interventions to deter criminal and violent behavior. In contrast to policing approaches such as increased enforcement or broad social initiatives such as employment programs, CBT is targeted and short-term, making it a relatively inexpensive and easy-to-implement policy option.

THE INTERVENTION

In 2009, the non-profit organization Youth Guidance implemented a program called “Becoming a Man” or BAM, which included in-school and after-school programming centered upon cognitive behavioral therapy. BAM enrolled male middle and high school students with low GPAs. On average, these students had missed 6–8 weeks of the school year and approximately one-third had been arrested before.

During the intervention, young men attended CBT sessions with prosocial role models (adults demonstrating positive and helpful behaviors, decision-making processes and actions). The goal was to promote new coping mechanisms and strategies, and to update youths’ perceptions of these tools as socially acceptable alternatives to criminal or violent reactions. The program was run throughout the school year, with participating students spending an hour of class time each week in small group sessions. Additionally, the after-school programming occupied students during high-risk times after school hours.

BAM sessions were structured as weekly check-ins that include reflective, introspective activities as well as immersive, experiential learning exercises like role-playing. One example of an immersive activity is known as “the fist.” Students are paired up and one member of each pair is given a ball. The other student is instructed to get the ball from the first student. The students then switch roles. For many students, the gut reaction is to resort to violence and aggressive behavior in seeking the ball. Following the exercise, facilitators ask whether they had considered simply asking their partner for the ball. Facilitators also asked...

“How would you have reacted if asked nicely for the ball?” Often, both partners stated that they would gladly have shared the ball if they had been asked. This helps students close the gap between misperceptions held in their mind and the actual reality of the situation. This also helps them realize that different approaches can lead to greatly improved outcomes. Although this exercise doesn’t tell students the “right” way to respond, it helps them reflect on impulsive and seemingly instinctive approaches. It also guides students to recognize the automatic responses that may be necessary for safety, reputation maintenance, and survival outside of school, and prompts them to recognize that these reactions may not be appropriate in other environments such as school.

**DISCUSSION TOPIC 2**

1. We will now discuss how an impact evaluation of this program might be conducted. What is a hypothesis that an impact evaluation of this program would test?

   *(Reminder: A hypothesis is an explanation of and for the desired effects of the intervention. For example, the hypothesis of a different intervention might be “Giving textbooks to students will improve student learning.”)*

2. What are the intermediate outcomes the training program is seeking to change? (In other words, what are its short-term goals?) What are the final outcomes the program is seeking to change? (In other words, what are its long-term goals?)
THEORY OF CHANGE

A theory of change (ToC) identifies the causal link between the intervention and the final outcome.

Note on commonly confused terms:

- An **output** is a direct action or fact that is part of a program being implemented as planned. For example: “student receives textbook.”
- An **outcome** is a change (a.k.a. effect or impact) in terms of an indicator that you can measure. For example: “change in students’ reading scores.”

- **Input**: An activity carried out as part of a program or intervention e.g. Textbooks are given to schools.
- **Output**: A step in the planned implementation of a program or intervention – a.k.a. a direct result in response to the inputs e.g. Students receive textbooks through schools.
- **Outcome**: A change or impact caused by the program that is being evaluated e.g. Increase in student learning levels.

  Intermediate outcomes: Observable changes or impacts caused by the program that are not the ultimate outcome of interest, but necessary along the way to achieving a final outcome e.g. Increase in students who have passing test scores for the semester.

  Final outcomes: Changes or impacts that are of ultimate interest to researchers and/or program implementers; these are often the overall goals of a program e.g. Increase in high school graduation rates.
1. Using the format presented in Figure 1, draw out a causal chain (a theory of change) that connects the intervention to your expected intermediate and final outcomes.

For 2 minutes, let the group fill in the diagram on the last page of their handout individually or in pairs. Then regroup to brainstorm and fill out a diagram together on a board.

Below is one example of connecting the intervention to the outcomes.

If the group is hesitant, consider these guiding questions:

- “Let’s think about inputs. What are the elements that the people running BAM are “putting in” to the program?”
- “Let’s think about outputs. If BAM is running as planned, what will be happening in the moment, from a student’s perspective?”
- “The previous question asked about some “outcomes,” or short-term and long-term goals for changes that result from the program. Which outcomes are the most important to measure to tell if the program is having effects in terms of the problems youth are facing?”

<table>
<thead>
<tr>
<th>Needs Assessment</th>
<th>Youth in Chicago from low-income households die from gun violence at higher rates and graduate from high school at lower rates than peers. Youth sometimes respond automatically in ways that do not fit the situation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention / Input</td>
<td>CBT sessions are offered (during school hours in public schools, and in a juvenile temporary detention center)</td>
</tr>
<tr>
<td>Output</td>
<td>Students have access to CBT sessions and attend them</td>
</tr>
<tr>
<td>Intermediate Outcomes</td>
<td>Students slow down automatic thinking; decision-making is more responsive to situation (including school and conflict outside of school)</td>
</tr>
<tr>
<td>Outcome</td>
<td>Better graduation rates; fewer arrests for violent crime; lower homicide mortality rates</td>
</tr>
</tbody>
</table>

(Leave space on the board for an additional column on indicators, later)
- “Think through how the diagram might look if you were attempting to address a different factor that contributes to crime/low graduation rates among this population.”
- “What the ToC look like for an intervention addressing the contributing factors of “lack of jobs opportunities”?“
- “What about the contributing factor of “lack of education or training”?"

Learning objective for Discussion Topic 3, Question 2: Identify assumptions that support the ToC.

2. What are the necessary conditions/assumptions underlying this ToC? What needs to occur or be present for this chain to work?

Encourage participants to question assumptions made in earlier steps. There are many underlying assumptions/conditions that must hold in order for this ToC to be sensible.

Examples of underlying assumptions:
- Teachers and role models implement BAM’s CBT curriculum as planned.
- Students attend (and do not skip) BAM sessions.
- Automatic/impulsive thought processes lead to increased violence.
- CBT leads to better decision-making.

TA: Read for 3-5 minutes

MEASUREMENT: INDICATORS AND DATA COLLECTION

The ideal data collection plan measures indicators at every stage of the theory of change. Before deciding which data to collect, you need to:

- Clearly define the inputs, outputs, and outcomes you are targeting
- Identify the ways the intervention is thought to affect the outputs and outcomes

Defining a main hypothesis and theory of change at the beginning of an evaluation is a crucial step that will help you determine what data/information to collect.

For each step of the theory of change, you must identify indicators (what to measure) and instruments (tools for data collection, a.k.a. methods for measurement). If possible, you should also collect data to validate the assumptions underpinning your theory of change.

For every type of data and at every step of measurement, is important to consider the participants’ response process and how this and other factors may affect measurement.

TA: Work on this and discuss both questions for total of 10 minutes
DISCUSSION TOPIC 4

Learning objective for Discussion Topic 4, Question 1: Choose indicators for multiple steps in the ToC.

1. Which indicators would you measure at each step of your theory of change? In other words, what are possible indicators for the inputs, outputs, intermediate outcomes, and final outcome(s)?

To break this question down ask:

- If participants are struggling with remembering these terms:
- Zoom out and remind the group about the big picture. What do you need to track in order to understand what is happening – what’s being put in (inputs), what’s happening as a direct result, what intermediate things are happening, and where does it all end up?
- Go through the terms with BAMBAM-specific examples.

As you go, either fill out the ToC diagram on the board with indicators and data collection decisions (see next page for example of this diagram), or simply discuss verbally – this is up to you. Consider limiting discussion to one indicator per stage in the interest of time.

a. How would you find out whether the BAM intervention itself is operating as planned? (a.k.a. What are indicators for inputs and outputs of the program?)

If the group is hesitant, ask:

“What might go wrong in the operation of the CBT sessions, and how would you be able to measure that?”

Example answer: Facilitators might deviate from the curriculum. There could be spot-check observers during some of the sessions to record whether the curriculum is going as planned.

→ See diagram below for more examples of indicators

“What are indicators that could indicate: Are teachers being trained in the CBT curriculum as planned? Are students attending?”
b. How would you measure BAM’s intermediate outcomes as described in your theory of change? (a.k.a. What are indicators for intermediate outcomes?)

If the group is hesitant, consider these guiding prompts:

- If we did not measure any intermediate outcomes, we might know the sessions were happening as planned but then might still see no changes in terms of crime or educational outcomes. If that happened, you wouldn’t be able to know what part of the ToC was breaking down – maybe students were slowing down their thinking, but that in turn was not affecting crime or graduation rates.

- On the other hand, if the program did lead to changes in crime or graduation rates but you did not have information on any intermediate steps, you would know it worked but you would not know why. Perhaps the program worked but your ToC was incorrect – maybe it worked not as a result of slowing down thinking, but only because of exposure to positive role models.

Do students change their thinking as a result of BAM sessions (at least in the short term)?

Example answer: Conduct interviews or vignettes with students to see whether their thinking processes are changing as a result of being in the program. → See diagram below for more examples of indicators

c. What should you measure in order to assess whether the intervention has an impact? (a.k.a. What are indicators for the final outcomes?)

Are there changes in rates of criminal or violent activity for students who participated in the program?

Are there changes in educational outcomes for students who participated in the program? → See diagram below for more examples of indicators
Learning objective for Discussion Topic 4, Question 1: Choose indicators for multiple steps in the ToC

2. How would you collect data for these indicators?

**Needs Assessment**

Youth in Chicago from low-income households die from gun violence at higher rates and graduate from high school at lower rates than peers. Youth sometimes respond automatically in ways that do not fit the situation.

**Indicators:** rates of crime; educational measures (attendance, grades, suspensions)

**Data:** administrative records on crime (for example, violent crime arrests) and education

**Intervention / Input**

CBT sessions are offered (during school hours in public schools, and in a juvenile temporary detention center)

**Indicators:** sessions are held; curriculum delivery

**Data:** interview and spot checks, or tools like attendance sheets collected in class

**Output**

Students have access to CBT sessions and attend them

**Indicators:** attendance rates in BAM, number of sessions held

**Data:** administrative records, spot checks, interviews

**Intermediate Outcomes**

Students slow down automatic thinking; decision-making is more responsive to situation (including school and conflict outside of school)

**Indicators:** decision-making during sessions and in vignettes and surveys with researchers

**Data:** administrative records, spot checks, interviews

**Outcome**

Better graduation rates; fewer arrests for violent crime; lower homicide mortality rates

**Indicators:** rates of graduation and dropout, crime rates, mortality rates

**Data:** administrative data from schools, police departments, vital statistics
As the group answers, use the board to write down indicators and data (see diagram.)

If the group is hesitant, consider these guiding questions:

- “Attendance would be an indicator for the “output” of the intervention (with a key input being “sessions are offered”). How might you collect data whether or not students are attending the CBT sessions?”

  (Answer: attendance sheets or administrative records.)

- “An important intermediate outcome is “students slowing down their decision-making.” How might you collect data on an indicator for whether not students are slowing down decision-making?”

  (Answer: researchers used interviews and vignettes where they timed decision-making.)

- “The indicators for the final outcomes we want to know about are rates of violence, rates of crime, and graduation rates. How would you collect data on rates of crime, for example?”

  (Answer: school records and arrest records.)

Answers/ideas listed as plain text here, if you would prefer not to use the diagram.

**Needs:** Baseline indicators

**Indicators:** rates of crime; educational measures (attendance, grades, suspensions)

**Data:** administrative records on crime and education

**Intervention/Input:** Implementation of the BAM program

**Indicators:** sessions are held; curriculum delivery

**Data:** interviews and spot checks, or tools like attendance sheets collected in class

**Output:** Students participate in the BAM program

**Indicators:** attendance rates in BAM, number of sessions held

**Data:** administrative records, spot checks, interviews

**Intermediate Outcomes:** Better decision-making in sessions (and in crucial situations related to conflict/violence and schooling)

**Indicators:** Decision-making during sessions; decision-making in vignettes and surveys with researchers

**Data:** Changes in behavior measured in sessions; survey results

**Outcome/Impact:** Better graduation/dropout outcomes, fewer arrests for violent crime, lower homicide mortality rate

**Indicators:** Rates of graduation and dropout, crime rates, mortality rates

**Data:** Administrative data from schools, police departments, vital statistics
Learning objective for Discussion Topic 4, Question 3: Brainstorm challenges to measurement.

3. What challenges might arise during the data collection and measurement processes? For example: In this hypothetical plan for data collection, are survey questions, study protocols, and protocols for the intervention itself clear and easy to comprehend? Are survey questions worded so as to avoid social desirability bias? If not, what might be the effects on survey responses? How might this affect the conclusions researchers draw from the study?

Time allowing, discuss the actual data used in the study:
- Summary index of three schooling outcomes (GPA, days present, and enrollment status at the end of the year)
- “To measure effects on criminal behavior, for study 1 we use electronic arrest records ("rap sheets") from the Illinois State Police (ISP). For study 2 we use arrest data from the Chicago Police Department (CPD). Both datasets are linked to our samples using probabilistic matching on first and last name and date of birth. Arrest records avoid the problem of under-reporting of criminal involvement in survey data (Kling, Liebman, and Katz 2007) but require the assumption that the intervention does not change the chances a crime results in arrest. Because intervention impacts can vary by crime type, we present results separately for violent, property, drug, and “other” crimes. We cannot distinguish “missing data” from “no arrests,” so we cannot explore how the arrest impacts change when we change how we handle missing data.” – Heller et al.

TA: Discuss for 5 minutes

INTERPRETING THE RESULTS / DISCUSSION TOPIC 5

Keep in mind when discussing the questions below that an impact evaluation is not a “thumbs up” or a “thumbs down” about a program – whatever the results are, valuable information can be gained and critical questions should continue to be asked.

Learning objective for Discussion Topic 5, Question 1: Interpret possible meanings of null results.

1. Imagine that the study finds that there is no impact of BAM on criminal activity or educational outcomes. How do you interpret these results? As a policymaker, how would you react to these results?

If the group is hesitant, prompt with these questions:

What questions should a policymaker ask next?
[Example responses] Does the finding of no impact mean that the Theory of Change was wrong? Does it mean that the program is not effective? Does it mean the program was implemented incorrectly?
- **Implementation:** ToC could still hold up; maybe elements of the program were not implemented with fidelity. Maybe the program needs tweaks (different duration, different implementation, etc.) to be effective.
- **Program not effective:** Maybe the program doesn’t work! Getting some qualitative evidence/talking to facilitators and participants might shed some light on what happened.
- **ToC not right:** Maybe underlying issues were caused by a different mechanism

Learning objective for Discussion Topic 5, Question 2: Interpret possible meanings of a finding of impact.
2. The real study by Heller et al. found that the program reduced the number of violent arrests per youth per year by 20% and had positive effects on educational outcomes as well. How do you interpret these results? As a policymaker, how would you react? What other information might be needed to make a policy decision?

Help participants think through the meaning of results that indicate an impact.

If the group is hesitant, prompt with these questions:

- “What questions should a policymaker ask next?”
- “Does this mean that the ToC is right and there are no program tweaks or adjustments needed?”

- First, check on the intermediate outcomes (did students attend sessions? Were teachers trained well? Did students report slowing down their thinking?). This will help clarify how and why the program did or did not work.
- A policymaker might be interested in learning about the costs of the program. If the program does appear to cause desirable outcomes, information on cost may help a policymaker determine whether and how the program could be expanded.
- All results should be looked upon with a critical eye. That is not to say that we don’t believe the results, but rather that just as null results would have led to questions about program tweaks, etc. – and so should results that are in line with our hypothesis.
- Qualitative follow up can still be useful to better understand the mechanisms.

More detail on actual results of the study:
The study found that participation reduced violent-crime arrests during the program year by 8.1 per 100 youth (a 44 percent reduction), and generated sustained gains in schooling.

Learning objective for Discussion Topic 5, Question 3: Connect theory of change to the interpretation of results.

3. [Optional – time allowing] Discuss and reflect on how the program’s theory of change provides context for interpreting these results.

Without a clear theory of change and understanding of how the intervention “should” lead to the desired outcomes, it is not at all clear how to interpret the results of any study. In this case, there were a number of hypotheses about how to reduce criminal activity in this population. If the study finds null results, this might indicate that the hypothesis was not correct, but not that this particular intervention could never work.

Heller et al., 2015, Table V.
REFERENCES AND FURTHER READING


REUSE AND CITATIONS

To request permission to reuse this case study or access the accompanying teachers’ guide, please email training@povertyactionlab.org. Please do not reuse without permission. To reference this case study, please cite as:


To reference the original study by Heller and coauthors, please cite as: