Document Title:	**An Evaluation of Simulation vs. Classroom-Based Implicit Bias Training to Improve Police Decision Making and Enhance the Outcomes of Police-Citizen Encounters**
Author(s):	**Lois James, Ph.D.**
Document Number:	**306564**
Date Received:	**May 2023**
Award Number:	**2017-R2-CX-0024**

Cover Page

Federal Award Number: 2017-R2-CX-0024

Project Title: An Evaluation of Simulation vs. Classroom-Based Implicit Bias Training to Improve Police Decision Making and Enhance the Outcomes of Police-Citizen Encounters

Principal Investigator: Lois James, PhD.

Contact Information: lois_james@wsu.edu (509) 385-9386

412 E Spokane Falls Blvd, Spokane WA, 99223

Award Recipient Organization: Washington State University

Project Period: 01/01/2018 – 12/31/2020

Award Amount: $749,864

Summary of the Project

Major Goals and Objectives

The goal of this study was to conduct an experiment evaluating the impact of implicit bias training on four distinct outcomes: 1) officer behavior, measured by coding body worn camera (BWC) footage of police-community member encounters; 2) specific "discrimination" based community member complaints against police; 3) officer perceptions of training effectiveness; and 4) arrestee perceptions of treatment by police. Due to the combined effects of the Covid-19 pandemic and the civil unrest following the murder of George Floyd, the fourth outcome was dropped from the study due the participating department determining that it was overly burdensome to collect.

Two types of training were evaluated in this experiment: classroom-based implicit bias training, and simulation-based counter bias training. This resulted in four separate groups of patrol officers (N~400), who were followed over an approximate three-year period in a large municipal department: First, a group that received only classroom-based training (classroom treatment group); second, a group that received only simulation-based training (simulation treatment group); third, a group that received both training types (combined treatment group); and forth, a group that received neither training types (control group). In doing so, this is the first police experiment testing the behavioral impact of implicit bias training in the field.

Research Questions

The main research questions for the current study were as follows:

1. Can implicit bias training impact police behavior by improving fairness in their decision making during interactions with community members?
2. Can implicit bias training impact community member perceptions of police bias based on discrimination complaints?
3. Is classroom-based training, simulation-based training, or a combination of both most effective at improving fairness in police decision making and reducing perceptions of bias in community members?
4. What are officer perceptions of implicit bias training?

The specific study hypotheses tested were:

H_1: At least one of the treatment groups receiving training (classroom only, simulation only, or a combination of both) will have significantly improved fairness scores following training compared to control group officers

H_2: At least one of the treatment groups receiving training (classroom only, simulation only, or a combination of both) will have significantly reduced numbers of discrimination-based community member complaints following training compared to control group officers

H_3: The treatment group that receives both training types will have the greatest improvement in officer fairness scores and the largest reduction in discrimination-based community member complaints

Research Design

This study used an experimental research design. To evaluate the effectiveness of implicit bias training we partnered with a large municipal police department (approximately 400 patrol officers at time of testing) who were willing to give us access to data and officers to conduct an

on-site field experiment. To evaluate the intervention, we focused on three types of outcome data. First, to evaluate the impact of training on officer behavior we coded BWC footage using a validated tool for measuring police performance (i.e., how they treat people) during encounters with members of the public. Second, to evaluate the impact of training on community members' perceptions of officer bias we measured discrimination-based complaints. Third, we surveyed officers following participation in training to assess their perceptions of training effectiveness.

Following a 12-month long baseline data collection period (the calendar year of 2019), patrol officers within our partner department were assigned to one of four groups: a treatment group that received classroom-based implicit bias training (n~50), a treatment group that received simulation-based counter bias training (n~100), a treatment group that received both training modalities (n~50), and a control group that received neither intervention (n~200). Assignment to groups was randomized at the patrol district level. Following the intervention, which took place over the 2020 calendar year, data collection (coding of BWC videos and measuring discrimination-based community member complaints) re-commenced for a 10-month period (January through October of 2021).

Body Worn Camera (BWC) Coding

Footage was recorded using Axon Body 2™ cameras and uploaded to an online portal hosted by Axon Enterprise™. Over 750,000 videos were uploaded within the sampling period of present study, which included the calendar year of 2019 (baseline) and the first 10 months of 2021 (post-intervention). Over 1,000 videos were randomly sampled from the baseline period and approximately 600 from the post-intervention period. For each calendar day, three videos were sampled corresponding to the three work shifts of officers: day (06:00-16:00), evening

(14:30-00:30), and night (21:00-07:00). Videos were sampled and coded by raters until each calendar day had three videos.

For a detailed description of the tool used to code BWC footage, as well as the validation process to ensure the tool's utility for measuring police behavior please see (Brown-Elkins, James & James, forthcoming). In brief, the coding tool consisted of 25 discrete items drawn from the collected pool of items generated by the metric-development work of Vila and colleagues (see Vila et al.,2018 for a comprehensive review). These items were specifically concerned with aurally measurable officer behaviors, such as offering verbal greetings, explaining the purpose of the encounter to the community member, showing signs of empathy, trying to de-escalate, etc. Please see the coding tool used attached separately.

In addition to performance items, information was gathered from each video about community member variables such as apparent race, gender, socio-economic status and so on, as well as information about how each video ended such as with an arrest, a citation, a warning, use of force and so on. For each video, coders were instructed to score officer performance items with one of four numerical designations: 1 = "yes", 0 = "no", -1 = "not applicable", or -9 = "unknown". For example, for the item "The officer demonstrates gratitude at a community member's compliance", the coder would designate "1" if the officer thanked the community member when they complied, "0" if the officer did not thank the community member when they complied, "-1" if the officer did not have the opportunity to thank the community member for compliance (either due to the community member's active non-compliance or them not receiving instructions requiring compliance), or "-9" if it was not clear if the community member complied, thus providing the opportunity for the officer to thank them.

From these scores assigned to each item, an officer received an overall performance score for a police-community interaction by dividing the sum of the scores they received by the total score possible to achieve in an encounter. Thus, scores are expressed as a proportion of all "fair" decisions that are possible in the encounter which are measured by the coding tool. Our rational for using this tool to assess fairness in officer decision making is that police-community member encounters are by their nature probabilistic. That is, an officer could behave impeccably but the encounter could still have a negative outcome (for example a community member's arrest). Conversely, an officer could treat the community member without dignity or respect, and yet the encounter may not result in a negative outcome. By employing a direct measure of officer behavior we overcome this limitation of evaluation studies that focus exclusively on the outcome of encounters such as arrests or use of force.

Discrimination Based Community Member Complaints

To supplement our primary outcome variable of officer behavior, we chose to assess a straightforward and easily accessible variable—number of community member complaints against police officers specifically for reasons of discrimination. Number of complaints during ten of the twelve baseline data collection months were compared to number of complaints during the ten post-intervention data collection months for a comparison of community member perceptions of police officer bias.

Officer Perceptions of Training

Our final outcome variable was measured by surveying officers post training to assess their perceptions of training effectiveness. This survey was developed in survey monkey, and a

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

link was emailed to all officers who received training, the week following their participation. All survey data was collected anonymously.

Training Interventions

The classroom-based training used in the current study was provided by the American Leadership Forum and consisted of a four-hour training session including discussion on the cognitive science of implicit bias, history of race in America, othering and belonging, inclusion, racial anxiety, structural inclusion, community engagement, and adaptive leadership. The desired outcomes of the training were for participants to gain an understanding and/or application of:

- Cognitive science and implicit bias
- The history of race in America- Understand the history/background of structural racialization, and the role of adaptive leadership and social narratives, specifically regarding the police as well as people of color and other marginalized groups
- Racial anxiety and stereotype threat and how they are related and connected with other institutions, structures, and systems that reinforce the process of structural racialization
- Othering and belonging framework, including the benefits of belonging and dangers of othering
- Understanding strategies to change behavior; receive guidance, tools, and framework(s) to use
- Adaptive leadership

This classroom-based training was delivered by American Leadership Forum members, in consultation with the department. This training was delivered in the Spring and Summer of 2020.

The simulation-based training used in the current study was Counter Bias Training Simulation (CBTsim™), a four-hour scenario-based training program developed by researchers at Washington State University (James & James, 2019). Incorporating the philosophy of counter conditioning, the purpose of CBTsim™ is to expose officers to a diverse range of scenarios in a decision-making simulator so that they can practice making unbiased decisions.

The goal of CBTsim™ is for officers to treat on-screen community members with dignity and respect, while making decisions about levels of threat based on an analysis of community member actions and not characteristics such as race, socio-economic status, mental status and so on. The guiding principles of CBTsim™ are: (1) repeated exposure to scenarios in which community member characteristics are not related to scenario outcome will reduce stereotyping based on biases; and, (2) post-scenario debriefing (self-reflection, peer feedback, and instructor feedback) will enable any implicit biases to come to officers' attention, which will reduce their impact on officers' decisions.

This simulation-based training was delivered by Washington State University members, in consultation with the department. This training was delivered in the Fall of 2020.

Analytical Approach

Differences in officer performance scores between the groups and before and after the training were analyzed using multi-level mixed models. This analytical approach can account for multiple observations per participant over time, while reducing risks of Type 1 error due to potential lack of independence among data points.

Differences in number of discrimination-based community member complaints between the groups before and after the training were analyzed using a mixed ANOVA model. All data analysis occurred SPSS.

Officer perceptions of training were analyzed using descriptive statistics.

Expected Applicability of the Research

The policing profession has undergone what many have termed a "crisis of legitimacy" (James, Fridell, & Straub, 2016; White & Fradella, 2016; Terrill, 2016; Paoline, Gau, & Terrill, 2016; Nix, Campbell, Byers, & Alpert, 2017; Todak, 2017). Although racial tensions have consistently permeated the history of American policing, the 2014 shooting of Michael Brown in Ferguson brought allegations of racial bias back to the forefront of the national conversation about police legitimacy. Since 2014, a wave of police shootings and in custody deaths of Black community members has rocked the foundation of public trust in police. These deaths include but are not limited to Tamir Rice, Eric Garner, Freddie Gray, Akai Gurley, Philando Castile, Stephon Clark, Alton Sterling, Botham Jean, Duante Wright, Andre Hill, Atatiana Jefferson, Breonna Taylor, and George Floyd.

Relatedly, around the same time the law enforcement community began to pay more attention to implicit (or unconscious) bias, and in particular the notion that officers may not display deliberate discrimination or bigotry but can still be influenced by biases in their behavior, judgments, and decisions (Dasgupta, 2013; James, Klinger, & Vila, 2014). Some departments mandate implicit bias training in response to government-imposed measures such as consent decrees or oversight processes such as collaborative reforms. Other departments proactively seek out implicit bias training in an attempt to promote fairness in police decision making, enhance

public perceptions of police, and ultimately improve the outcomes of police-community member encounters. Regardless of motivation, in response to broad concerns about racial bias in police decision making, implicit bias training is becoming a staple among police departments (Worden et al., 2020).

Despite widespread adoption of implicit bias training among police departments, very little evidence exists regarding the effectiveness of this training. The most rigorous evaluation of implicit bias training in policing to date is the Worden and colleagues (2020) study evaluating the impact of an implicit bias intervention in the NYPD. This study was massive in scope—with all 36,000 sworn personnel receiving the intervention. Outcome measures included officers' knowledge about the science of bias, their attitudes about bias and intentions to use techniques learned during training in their daily activities, and disparities in enforcement activities such as arrests and use of force. The researchers found a moderate improvement to knowledge, a small impact on attitudes, and insufficient evidence to suggest disparities in police actions were reduced following the intervention. The evaluation report acknowledged a key limitation of the study—that it is extremely difficult to isolate training effects on outcomes such as use of force from other factors that have an undeniable and legitimate influence on these outcomes (such as community member actions). The researchers conclude by stating that "if disparities stem from forces other than implicit bias, then even a well- designed training that is flawlessly delivered cannot be expected to alter patterns of police enforcement behavior" (pg. vi).

Although the NYPD evaluation included outcomes such as arrests and use of force, arguably, their still exists a distinct lack of understanding about the impact of implicit bias training on how officers treat people during their interactions with them. As such, we contend that prior to this study there was still no research showing the extent to which implicit bias

training is effective for *changing behavior* in law enforcement. Ultimately, the stakes are high for understanding the impact of implicit bias training modalities on police decision making, public perceptions of police legitimacy, and the outcomes of police-citizen interactions. This is particularly true given the recent "crisis of legitimacy" in American policing, which has placed the police profession under a microscope of public scrutiny. Thus, the results of this study have widespread applicability and implication for US policing.

Participants and Other Collaborating Organizations

This project was conducted in collaboration with our partner police agency, the Sacramento Police Department (Sac PD). Approximately 400 patrol officers from Sac PD participated in this experiment.

Changes in Approach from Original Design

Three substantial changes were necessitated from the original design. The first was a change in the partner agency from the Cleveland Police Department to the Sacramento Police Department. This was due to changes in leadership between the time of proposal submission to award receipt. The securing of a replacement department, together with delays due to the Covid-19 pandemic and the civil unrest following George Floyd's murder resulted in an approximate 18-month delay to project completion.

The second change from the original design was that randomization into groups occurred at the patrol district level instead of the officer level as originally intended. This adaptation was necessary to fit with Sac PD's training schedule and needs. Given our extreme gratitude from

Sac PD's last-minute involvement, we were willing to compromise to minimize burden on the department's trainers and scheduling team.

The third change from the original design was the removal of arrestee perceptions of police treatment as an outcome. This variable would have required access to the jail and was determined to be too burdensome given the pandemic, civil unrest, and consequent understaffing.

Outcomes

Activities / Accomplishments

All major activities (excluding gathering arrestee perceptions of police treatment – see rationale above) were accomplished. The study period was extended by 18 months given various delays experienced due to a change in police agency partner, the Covid-19 pandemic, and civil unrest following the murder of George Floyd.

Results and Findings

Analysis of performance scores revealed a significant interaction between group (intervention group) and time (pre- vs. post scores), $f = 4.736$, $p = .009$. The interaction revealed a significant increase in Performance from pre- to post-intervention, exclusively for the treatment group who received both the classroom- and simulation- based training.

Performance	*Numerator df*	*Denominator df*	*F*	*Sig.*
Group	2	340.153	0.192	0.826
Time	1	1670.397	3.088	0.079
Group * Time	2	1667.398	4.736	**0.009**

Performance scores in the combined group went from an average of 50% before the training, to 61% after the training. In practice, this means that officers went from performing 50% of the items listed in table 1 that they feasibly could have done, to 61% of items after the training.

Next, we plotted the means of individual officer performance items that make up the Performance measure across groups to determine any notable items contributing to performance increase. Results suggest that individuals in the group that received both training types may remember community members' names, express concern for community members' safety, and apologize to community members for the encounter more compared to baseline and compared to other groups post-intervention. It also appears that certain kinds of behaviors were *reduced* from baseline to post-intervention for all groups except the combined training group – notably, spending time with the community member and establishing common ground with the community member. The first makes intuitive sense given the pressures of the pandemic and consequent understaffing. The second might be related to increased tension between police and community members following civil unrest of 2020.

The mixed ANOVA model testing whether discrimination-based community member complaints were impacted by the training intervention also revealed a significant interaction between group (intervention group) and time (pre- vs. post scores), ($f = 3.042$; $df = 2$; $p = .049$). As was found for officer performance scores, the interaction revealed a significant decrease in discrimination complaints from pre- to post-intervention, exclusively for the treatment group who received both the classroom- and simulation- based training.

The survey of officers' perceptions about the training revealed that 83% of officers found the training to be interesting and engaging, 83% found it to be beneficial for improving fairness

in police decision making, 75% found it to be beneficial for improving public perceptions of police legitimacy, and 75% found it to be beneficial for enhancing the outcomes of police-community member encounters. Qualitative feedback included:

"Talking over the simulations afterwards was helpful. We could discuss mindset and think a bit on WHY we acted/spoke a particular way."

"The roundtable feedback after the simulation was priceless. Within the group, there was different levels of experience so it was great to hear different perspectives."

"I didn't think it would be beneficial, but I actually really liked it"

Limitations

These study results represent a substantial contribution to the research literature. It is the first (preliminary) evidence of a potential behavioral impact of implicit bias training. Results also suggest that community member perceptions of police fairness might be positively impacted following officers' engagement in implicit bias training. However, several notable limitations need to be addressed, and results need to be interpreted and caveated with these in mind.

First, the combined effects of the Covid-19 pandemic and the civil unrest of 2020 might have influenced results in ways that are difficult to control. The presence of a control group within this study prevents the data from being meaningless in the face of extreme temporal changes between 2019 (baseline data collection) and 2021 (post-intervention data collection). However, these extreme influences may jeopardize the generalizability of study results. Sacramento Police Department's experiences with the pandemic and civil unrest might not be the

same as other departments' experiences, and this should be kept in mind when considering the national applicability of study results.

Second, and related to the first limitation, several adjustments were necessary from the original research design. We had intended to randomize officers into groups at the individual level, to approximate a true randomized control trial most closely. Given the scheduling burden faced by the department, understaffing issues, and other logistical considerations, it was determined that randomization at the patrol district level was the only feasible randomization option. In addition, the outcome measure of arrestee perceptions of police treatment was dropped from the study. It was determined that it was too big of an ask to access the jail, in light of the Covid-19 pandemic.

Artifacts

Products, Datasets, and Dissemination

Manuscripts:

"Validating a novel tool for coding body worn camera footage of police-community member interactions" (under review with Criminology and Public Policy). Elkins-Brown, N., James, S., & James, L.

"Using Body Worn Camera Footage to Investigate Situational Predictors of Officer Behavior and the Outcomes of Police-Community Interactions" (being prepared for submission to Police Quarterly). James, L., James, S., & Elkins-Brown, N.

"Can Implicit Bias Training Improve Fairness in Police Decision Making? Results from an Effectiveness Evaluation" (being prepared for submission to Nature). James, L., James, S., Elkins-Brown, N., Mitchell, R.

Presentations:

"Effectiveness Evaluation of Implicit Bias Training: Research Study Results" (April 2022, The Justice Clearing House, virtual). James, L. & Mitchell, R.

"Counter Bias Training: Evaluation Study Results" (May 2022, The American Society of Evidence Based Policing, Annual Conference, Washington DC). James, L.

Datasets:

The first data set of note contains the coded BWC footage from before and after the training intervention (2019 and 2021). This includes the following variables (coded as "yes", "no", "not applicable" or "unknown") for approximately 1,700 videos:

Community Member Items
The community member appears to be white
The community member appears to be black
The community member appears to be Hispanic
The community member appears to be Asian
The community member appears to be Native American
The community member's race appears to be "other" (not captured above)
The community member appears to be male
The community member appears to be female
The community member is armed
The community member appears to be high socio-economic-status
The community member appears to be mid socio-economic-status
The community member appears to be low socio-economic-status

The community member appears to be homeless
The community member appears to be large in physical stature
The community member appears to be medium in physical stature
The community member appears to be small in physical stature
The community member appears to be a child
The community member appears to be a teenager or young adult (appears to be 13-21)
The community member appears to be an adult (appears to be 22-50)
The community member appears to be an older adult (appears to be 51-70)
The community member appears to be elderly (appears to be over 70)
The community member appears to be impaired by alcohol or drugs
The community member has apparent mental illness or developmental disability
The community member appears to be in crisis
The community member appears to be disrespectful to the officer (e.g., contempt of cop)
The community member is verbally threatening to the officer
The community member attacks or assaults the officer
the community member attacks or assaults someone in the officers presence
The community member is wearing street type attire (e.g., hoodies, jeans, sneakers)
The community member is wearing business type attire
The community member has apparent gang indicators
The community member has had prior interactions with the officer

Officer Performance Items

The officer verbally greets the community member
The officer introduces themselves by name to the community member
The officer shows natural human emotion
The officer offers to help the community member
The officer attempts to establish common ground with the community member
The officer shows signs of empathy to the community member
The officer attempts to put the community member in their shoes
The officer explains their actions to the community member
The officer communicates with the community member while carrying out a task
The officer apologizes for the inconvenience of an encounter
The officer treats emotionally disturbed community members with dignity and respect
The officer demonstrates gratitude at a community member's compliance
The officer does not patronize or insult the community member
The officer does not point their weapon at the community member unnecessarily
The officer demonstrates concern for the community member's safety
the officer provides clear instructions to the community member

The officer makes sure they understand what the community member is communicating
The officer changes tactics when original tactics are not working
The officer recognizes when their actions are not appropriate and modifies them
The officer attempts to de-escalate a volatile situation
The officer spends some time with the community member before leaving the encounter
The officer remembers names and uses them when saying goodbye
The officer ends the encounter on a positive note
The officer leaves the community member with useful information
Use of Force Items
The officer put hands on the community member
The officer put handcuffs on the community member
The officer pointed a Taser at the community member
The officer used a Taser on the community member
The officer used pepper spray on the community member
The officer used strike techniques on the community member
The officer used hold techniques on the community member
The officer deployed a Canine on the community member
The officer pointed a weapon at the community member
The officer used lethal force (handgun)
The officer used lethal force (rifle)
The officer used lethal force (shotgun)
Outcome Items
The community member appeared to be satisfied with the interaction
The community member was arrested
The community member was issued a citation
The community member was taken to hospital
The community member was given a warning
The community member was handed off to other officer
Outcome "other" (qualitative note on outcome)

The second dataset contains community member complaints from the measurement period before and after the intervention (2019 and 2021).

References

Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. *Advances in Experimental Social Psychology*, *47*, 233–279.

James, L., Fridell, L., & Straub, F. (2016, February). Implicit bias versus the "Ferguson Effect": Psychosocial factors impacting officers' decisions to use deadly force. *The Police Chief*, pp. 44–51.

James, L., Klinger, D., & Vila, B. (2014). Racial and ethnic bias in decisions to shoot seen through a stronger lens: Experimental results from high-fidelity laboratory simulations. *Journal of Experimental Criminology*, *10*(3), 323–340.

Nix, J., Campbell, B. A., Byers, E. H., & Alpert, G. P. (2017). A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy*, *16*(1), 309–340.

Todak, N. (2017). The decision to become a police officer in a legitimacy crisis. *Women & Criminal Justice*. https://doi.org/http://dx.doi.org/10.1080/08974454.2016.1256804

Vila, B., James, S., & James, L. (2018). How police officers perform in encounters with the public: measuring what matters at the individual level. *Policing: An International Journal*, *41*(2), 215-232.

White, M. D., & Fradella, H. F. (2016). *Stop and frisk: The use and abuse of a controversial policing tactic*. New York: New York University Press.

Worden, R. E., McLean, S. J., Engel, R. S., Cochran, H., Corsaro, N., Reynolds, D., ... & Isaza, G. T. (2020). The impacts of implicit bias awareness training in the NYPD. *The John F.*

Finn Institute for Public Safety, Inc. and The Center for Police Research and Policy at the University of Cincinnati.