Final Report:
Promoting Officer Integrity Through Early Engagements and Procedural Justice in the Seattle Police Department

by
Emily G. Owens, Ph.D.
University of Pennsylvania

David Weisburd, Ph.D.
George Mason University

Geoffrey Alpert, Ph.D.
The University of South Carolina

Karen L. Amendola, Ph.D.
Police Foundation

Police Foundation
1201 Connecticut Avenue, NW
Suite 200
Washington, DC 20036
(202)833-1460 ♦ (202)296-2012
www.policefoundation.org

[DISCLAIMER]

Final Report for Project # 2012-IJ-CX-0009

This project was supported by Grant No. 2012-IJ-CX-0009, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice, the Police Foundation, or the Seattle Police Department.

Table of Contents

List of Tables and Figures

Tables

Figures

ACKNOWLEDGMENTS

We thank Maria Valdovinos and Earl Hamilton of the Police Foundation for valuable feedback and assistance at all stages of this project. We also appreciate comments provided by Dr. Travis Taniguchi, formerly of the Police Foundation and now at RTI. This research was made possible by the cooperation and enthusiasm of former Deputy Chief Clark Kimerer, Captain James Dermody, Kim Kubie, Mimi Walsh, Peggy Garcia, and Jan Hoyt, as well as all participating officers of the Seattle Police Department. We also greatly appreciate the assistance of Claudia Gross-Shader of the City of Seattle. Finally we acknowledge the ongoing support and assistance of Dr. Brett Chapman of the National Institute of Justice (NIJ), who served as our Program Manager.

ABSTRACT

Academics have long argued that the use of procedural justice is a necessary component of effective policing; yet, there is scant evidence on how the goals of procedural justice can be implemented in a practical way and on whether training officers to "listen and explain with equity and dignity" (LEED) actually translates into quantifiable improvements in field outcomes that policy makers care about. The purpose of this study was to conduct an experimental evaluation of an innovative training program aimed at promoting the use of procedural justice by officers in the Seattle Police Department (SPD).

A High Risk Circumstance (HRC) model was developed to identify officers working in "behavioral hot spots" or areas where an officer was at a higher risk of becoming involved in a potentially problematic event. Identified officers were assigned to either treatment or control. The treatment consisted of a non-disciplinary meeting with their supervisors where they were exposed to procedural justice and the principles of LEED in an interactive manner. The officers were subsequently compared to their control counterparts (N= 320), on four measures of overall activity including: a) total CAD incidents responded to, b) percentage of CAD incidents initiated by the officer, c) minutes on scene, and d) whether or not a written report was filed. The officers were furthermore, compared on three primary measures of how they responded to incidents: a) percent of incidents resolved via an arrest, b) frequency of officer involvement in incidents involving force, and c) frequency of complaints filed against the officer. As a complement to the arrest analysis, the percent of incidents resolved via a citation, a verbal warning, or by rendering assistance was also examined.

The officers who participated in supervisory meetings appeared to engage in encounters with citizens with equal frequency as their colleagues. However, those who participated in the meetings were roughly 26% less likely to resolve an incident with an arrest one week after having a meeting when compared to their colleagues who did not participate. This effect is reasonably persistent, and the results suggest that officers who participated in the LEED debriefs were 12% less likely overall to resolve incidents via an arrest over the six-week period after the supervisory meetings. The results also suggest that in the longer run, officers who participated in the meetings were over 50% less likely to be involved in a use of force incident.

Overall, we did not find evidence that officers who had additional non-disciplinary supervisory meetings were any more or less likely to respond to, initiate, or document CAD incidents relative to their peers who worked in similar situations. We also found no substantive change in the amount of time officers were officially on-scene in a given incident. Furthermore, we did not find evidence that officers who participated in the meetings were less likely to garner complaints from the public.

We conclude that non-disciplinary LEED based supervisory meetings are a promising strategy for improving police legitimacy. Officers who had at least one meeting over a six month period in which they reviewed how they approached relatively standard citizen encounters appeared to be less likely to engage in behaviors that, while central to policing, have the potential to reduce legitimacy when abused (e.g. making arrests and use of force).

Executive Summary

One of the fundamental challenges of government is to identify how to promote the interests of its constituents while maintaining a monopoly on the legitimate use of force (North, 1981). This problem is particularly acute when it comes to the administration of criminal justice; how do law enforcement agencies convince the public that they are benevolent and trustworthy agents, while at the same time making arrests, and coercing citizens? Officers must strike a delicate balance. They must actively engage the public, and issue warnings, citations, or arrest those accused of violating the law. At the same time, officers must make sure that citizens feel like they are being treated fairly, equitably, and are protected from harm.

Academics have long argued that the use of procedural justice is a necessary component of effective policing. While few practitioners would argue that treating citizens fairly, equitably and respectfully is not important, there is scant evidence on how the goals of procedural justice can be implemented in a practical way, and even less evidence that training officers to "listen and explain with equity and dignity" (LEED) actually translates into quantifiable improvements in field outcomes that policy makers care about.

To that end, we conducted an experimental evaluation of a training program aimed at promoting the use of procedural justice by officers in the Seattle Police Department (SPD). The innovation of the training program was threefold. First, we applied insights from criminology and statistics to develop a new kind of Early Intervention System, which we call a High Risk Circumstance (HRC) model. This model identifies officers working in behavioral "hot spots," small geographic areas where police officers are more likely to be involved in problematic citizen encounters such as assaults on police officers, intoxicated persons, etc. Second, sergeants trained in the concept of listening and explaining with equity and dignity (LEED) modeled this

approach to procedural justice in meetings with officers, in which the officers' were asked to discuss a recent encounter selected by the research team. This approach was designed to demonstrate a procedurally just encounter in hopes that it would translate to how officers interacted with members of the community and how they could interact with them in the future as well. Third, we provide experimental evidence on the impact of a feasible procedural justice training program that is based on two practical and quantifiable performance metrics: officer activity and incident outcomes.

The HRC Model

Our HRC model was calibrated using geographically identified, incident-level data and SPD human resource records from 2009. The model was used to assign a "risk score" to every incident recorded in the SPD's Computer Aided Dispatch (CAD) system, based on a 30 day history of events in close geographic and temporal (time of day and day of week) proximity to the incident in question.

The risk score was calculated as the frequency with which calls dispatched to that area over the past 30 days were based on 911 calls, reports of firearm involvement, officer safety notices added to calls, citizens in mental distress, domestic violence incidents, or if dispatchers frequently identified situations as substantively different from the way that officers cleared incidents. We also measured the number of officers who reported being injured in a particular block location (e.g., reported being bitten while making an arrest at a specific place) and whether or not the specific officer involved had more than the average number of sick days, secondary jobs, or overtime hours in the previous month. The contribution of each of these measures to predicted risk was based on their partial correlation with what we defined as "potentially problematic events." We operationalized "potentially problematic events" as those incidents in

which an officer used force, for which a citizen complaint was filed against the officer related to the incident, or for which an involved officer was injured.

During the field experiment, we identified officers on a bi-weekly basis who were involved in incidents in the top 12th percentile of predicted risk in their precinct based on the HRC model. SPD personnel suggested that 12 percent was the largest number of officers that could participate in the study without disrupting normal operations. Officers in the top percentile of risk were then randomly assigned to either a control group or a treatment group. In total, the experiment had 12 waves over the course of six months. Typically, officers participated in these meetings within 11 days of being notified that they were selected.

The Experimental Engagement

A total of 241 officers were assigned to treatment, and despite union and IRB rules mandating that officer participation in the engagements be voluntary, there was still a 91.7 percent compliance rate. Treatment in this experiment consisted of officers participating in a non-disciplinary LEED-based supervisory meeting, which we refer to as an "experimental engagement" throughout. Officers participating in these meetings were exposed to procedural justice techniques in an innovative and interactive manner.

Specifically, in the experimental engagement the officer was asked by his/her sergeant to discuss a recent substantive citizen encounter that had been selected by the supervisor from a list provided by the researcher. During this discussion, the sergeant actively modeled the use of procedural justice through his/her interactions with the officer, and also allowed the officer to contribute to the focus and timing of the meeting in a way that is atypical for the standard hierarchical structure of supervisory meetings. The sergeant also suggested moments in the encounter where the officer had the opportunity to implement procedural justice ideas in practice

but had failed to do so. The purpose of this was not to reprimand the officer or criticize; rather, the purpose was to prime them to recognize future similar opportunities where they could implement LEED in practice.

The aim of this innovative and interactive approach was to provide the officer with a two-fold reinforcement of how to implement procedural justice – specifically LEED concepts – in the field. During the meeting, a lieutenant oversaw the process, and if necessary, intervened with the sergeant in case he/she strayed from consistent application of the prescribed procedural justice techniques in which they had been trained.

At the conclusion of the engagement, officers were asked to fill out a confidential comment card about the experience and they were asked to mail it directly to the Police Foundation. Because completion of the comment cards was voluntary and anonymous, there was no opportunity for the researchers to follow-up with non-participating officers to request completion of the comment cards. Based on site visits conducted by the Police Foundation, it appears that encouragement by SPD personnel to complete these comment cards varied substantially, and this likely explains the low level of response we received.

Study Implementation

Over the course of 26 weeks from April 2013 through October 2013, a total of 221 procedural justice meetings were held. The research team conducted a total of five site visits during the field experiment, and observed a total of eight meetings. During those observations, it was clear that the collaborative approach to the meetings was unusual for many of the SPD officers, and while most sergeants and lieutenants engaged in good-faith efforts to conduct the meetings as they had been instructed, there was still substantial confusion about the purpose of the engagements by the officers. In particular, much of the confusion appeared to stem from the

officers' lack of understanding regarding the rationale for discussing the particular encounter selected, especially when the perception was that the chosen incident was benign or "a non-incident" as viewed by officers.

The final wave of the experiment was conducted on October 31, 2013. In order to evaluate the impact of the experimental engagements we received a dataset on all use of force incidents, filed complaints, and all CAD activity between January 1, 2013 and December 31, 2013. We then compared treated officers to their control counterparts (N= 320), on four measures of overall activity including: a) total CAD incidents responded to, b) percentage of CAD incidents initiated by the officer, c) minutes on scene, and d) whether or not a written report was filed. We also compared treated officers on three primary measures of how officers responded to incidents: a) percent of incidents resolved via an arrest, b) frequency of officer involvement in incidents involving force, and c) frequency of complaints filed against the officer. As a complement to our arrest analysis, we also examined the percent of incidents resolved via a citation, a verbal warning, or by rendering assistance.

Results

Overall, we did not find evidence that officers who participated in experimental engagements were any more or less likely to respond to, initiate, or document CAD incidents relative to their peers who worked in similar situations. We also found no substantive change in the amount of time officers were officially on-scene in a given incident. In addition to being statistically indistinguishable from control officers, we were able to rule out any impacts of the meetings on overall activity that were larger than a 7 percent change over a six-week period.

The officers who participated in supervisory meetings appeared to engage in encounters with citizens at an equal frequency as their colleagues; however, those in the treatment group

were roughly 26 percent less likely to resolve an incident with an arrest one week after participating in a non-disciplinary LEED meeting. This effect is reasonably persistent, and our results suggested that treated officers were 12 percent less likely overall to resolve incidents via an arrest over the six-week period after the experimental engagements took place. This is an important finding in the context of continuing concerns over use of force by police in a way that alienates citizens from the police such as when it appears unnecessary or excessive (Rosenbaum et al. 2005). While we did not find any evidence that officers were systematically resolving incidents in any particular alternate way, we did observe moderate short run increases in general "assistance rendered" dispositions, and longer-run imprecise increases in the number of citations given.

In addition, we also found evidence that in the longer run, treated officers were less likely to be involved in use of force incidents. While this effect was imprecisely estimated in the short run, when we compared officers over the entire sample period we estimated that engaged officers were over 30 percent less likely to be involved in a use of force incident. However, surprisingly, and in contrast to the observed reductions in arrests and use of force, we did not find evidence that treated officers were less likely to garner complaints from the public. We are unfortunately unable to draw any conclusions about this finding, as our estimates were statistically consistent with both large increases and large reductions in complaints.

We did not find compelling evidence, however, that the effect of the LEED meetings in the experimental engagements varied over time or location, or that more than one engagement over a six-month period resulted in an increasingly larger behavioral effect. We also did not identify a systematic pattern in effects across precincts but did find some evidence indicating that in the short run, treated officers in the East precinct spent more time on scene and had fewer

complaints filed against them than those in other precincts. Our site visits suggested that SPD officers were somewhat confused about the purpose of the non-disciplinary LEED-based supervisory meetings and that SPD personnel struggled with identifying a "substantive" incident to discuss in early rounds of the experiment. We think that this confusion stemmed from the fact that supervisory meetings in the SPD more typically occur when officers are involved in serious incidents in which there is a real threat of injury to citizens or themselves. Also, because of the language we used to describe qualifying incidents, specifically "high predicted risk," we believe this may have led SPD officers to conclude that the specific incident selected for discussion should have involved a substantive encounter. Although we had emphasized to the supervisors the fact that the HRC model was not activated by a specific problematic incident but rather by the officers being in situations perceived as high risk, officers frequently expressed confusion as to why they were "discussing small stuff." Despite the confusion regarding the nature of the incident selected for discussion, the feedback, albeit limited, provided by the officers in comment cards sent to researchers at the Police Foundation indicated that they were pleasantly surprised by the LEED interactions with their supervisors, reporting that they found them to be a positive change of pace from their typical meeting style. Indeed, it is likely that the culture of policing leads officers to believe that the only times in which supervisors want to talk to them is when they have done something wrong.

In sum, we identified officers based on predicted risk scores in an effort to maximize the potential plausible treatment effect by identifying officers who were most likely to disproportionately benefit from additional training in procedural justice. The logical implicit hypothesis was that the officers with the highest predicted risk scores would benefit more from having additional meetings than officers engaged in incidents with lower predicted risk scores.

In contrast, we found that the largest reduction in arrests occurred among officers with lower predicted risk scores. Over a six-week follow up period, the impact of supervisory meetings on arrests rates was roughly one-half the size among officers with one-half of a standard deviation higher predicted risk score. One potential interpretation of this result is that officers who work in the relatively lower risk areas when compared to higher risk areas are more likely to encounter situations in which the nature of the violation is less serious and less clear than in those situations encountered in higher risk areas and therefore, the feasibility of using discretion informed by procedural justice to resolve the situation over an arrest is might be higher.

Conclusion

One criticism of procedural justice training is that it may lead officers to pull back from the public in response to increased supervision or scrutiny, or what is more commonly referred to as "de-policing." This withdrawal, however, could lead to reductions in pro-actively working to solve crime and disorder problems. In this experiment, the intervention was an experimental engagement consisting of a non-disciplinary supervisory meeting in which the sergeant modeled procedural justice using LEED concepts for the selected officers. Our results demonstrated that non-disciplinary supervisory meetings that emphasized procedural justice in this manner were not associated with a reduction in officer activity ("de-policing" or withdrawal) from baseline levels. Instead, the results of our study suggested that the intervention in Seattle had an impact on the outcome measures we sought to influence; namely, we observed that officers who were assigned to these meetings were less likely to be involved in incidents that resulted in citizen arrests or in the long run, use of force.

Overall, we concluded that non-disciplinary LEED based supervisory meetings are a promising strategy for improving police legitimacy. Officers who had at least one meeting over a six month period in which they reviewed how they approached relatively standard citizen encounters appeared to be less likely to engage in behaviors that, while central to policing, have the potential to reduce legitimacy when abused (e.g. making arrests and use of force). While we found little evidence of substantial costs for adopting this training program on a limited basis, agencies that choose to experiment with LEED based supervisory meetings should, nevertheless, carefully monitor their employees for signs of de-policing.

Introduction

Police departments are charged with enforcing the law while respecting the dignity of the people in the communities in which they serve. Officers have to strike a delicate balance. They must actively issue warnings, citations, or arrests to people accused of violating the law and at the same time, officers must make sure that citizens feel like they are being protected from undue harm and are being treated fairly and equitably.

In academic circles, this balance between toughness and fairness is referred to as procedural justice (Tyler, 1988) and the research literature suggests that it is the most important factor in public assessments of police legitimacy (Hinds & Murphy, 2007). Specifically, procedural justice "describes the idea that how individuals regard the justice system is tied more to the perceived fairness of the process and how they were treated, rather than to the perceived fairness of the outcome" (Tyler, 1998). Procedural justice directly contributes to the development and establishment of legitimacy, defined as "a property of an authority or institution [such as the police] that leads people to feel that the authority or institution is entitled to be deferred to and obeyed" (Sunshine & Tyler, 2003, p. 514).

The procedural justice process discussed in this report is typically referred to as LEED, an acronym for "Listen and Explain with Equity and Dignity." Within psychology and criminology, the use of procedural justice is seen as a necessary component of police integrity in the community (Melekian, 2012; Sherman, 1998; Sunshine & Tyler, 2003; Tyler, 2005). In practice, police departments promote procedural justice by instructing recruits to use LEED concepts in their work, and then monitor their subsequent behavior on the job. This monitoring is increasingly being done with computer-based Early Intervention Systems (EIS), which typically track things like citizen complaints, sick days, commendations, and uses of force. Officers

flagged by EIS are further evaluated by at least one supervisor, who may recommend an additional departmental response (US DOJ, 2011).

Computer-based early intervention systems (EIS) are increasingly used by police agencies to identify officers who may benefit from mentoring or training in use of force techniques or cultural sensitivity. In 2003, 29% of law enforcement agencies surveyed by LEMAS reported using a computer-based monitoring system to identify at-risk officers. Four years later, 39% of surveyed agencies had such a system in place (US DOJ, BJS 2003, 2007). The diffusion of this particular technology is undoubtedly due, at least in part, to incentives put into place by the federal government; for example, the Civil Rights Division of the Department of Justice has recommended the adoption of an EIS program in practically all of its reviews of police department practices.[1]

Statement of the Problem

By definition, EIS models identify a small number of officers who have already engaged in problematic behavior. Although not intended to be disciplinary, the fact that EIS systems are based on an individual's past actions has led many officers to view these systems in an adversarial way. From this standpoint, the use of EIS may actually be perceived as violating the tenets of procedural justice–something an officer has done automatically "triggers" some sort of disciplinary review (Amendola, 2006). Further, EIS systems typically do not take into account recent advances in criminology and statistics regarding the prediction of rare events and the concentration of crime and other problems in small geographic areas.

[1] These reports are available online at:
http://www.justice.gov/crt/about/spl/findsettle.php#Law%20Enforcement%20Misc
Recommendations on EIS systems are mentioned in all save the Pittsburgh Police Department review, which was the first conducted by the DOJ.

While EIS monitoring has allowed law enforcement agencies to focus more on prevention, the profession itself is still very much hinged on deterrence strategies despite evidence that other more procedurally minded strategies, such as community oriented policing, may be more effective (Sunshine & Tyler, 2003). The end result is that officers are still very much evaluated on quantitative measures such as numbers of arrests and clearance rates, or potentially problematic behavioral indicators such as the number of complaints or use of force incidents, rather than more qualitative measures such as public perceptions of their fairness, for instance. This has resulted in a palpable tension between the core tenets of procedural justice and some policing philosophies centered around order maintenance (Gau & Brunson, 2010) and raises questions of how to best integrate procedural justice within a police department and how to best measure its effects.

In addition, there is currently a disconnect between the laudable academic goal of using procedural justice and the practical implementation of LEED concepts by police officers. In reality, the culture of most police departments is centered on the measurable performance outcomes of arrests, index crimes, and clearance rates rather than the difficult-to-quantify concept of fairness. Enacting reforms to promote the use of LEED is further complicated by the fact that most officers believe they already listen and explain themselves, affording the subject with whom they are involved a great deal of respect; indeed, the majority of citizens who interact with police think that the officers behaved appropriately (BJS, 2011).[2] While police chiefs believe that the active promotion of procedural justice is important, if officers believe they are

[2] In an experimental analysis of procedural justice training in New York, "control group" officers who were asked to rank the quality of their own use of procedural justice gave themselves an average of 3.13 out of 5. Officers who had received explicit LEED training gave themselves 3.16 out of 5 (Rosenbaum & Lawrence, 2011).

already using LEED, and an enforcement-focused, statistics-oriented, binary-decision making culture is entrenched in so many departments, how exactly do commanders make it happen?

Over the past three years, the Seattle Police Department (SPD) has been working with local, state and federal law enforcement representatives to develop a comprehensive training curriculum to promote procedural justice for citizens based on LEED principles in order to enhance the public perception of fairness and legitimacy in policing. The initiative to develop curricula and adopt LEED as an overarching law enforcement value and ethos was the product of collaboration between (then) SPD Chief John Diaz, (then) Director of the Washington State Criminal Justice Training Academy Joe Haugh, and (then) King County Sheriff Sue Rahr, who is the current Director of the Washington State Criminal Justice Training Academy.

Unfortunately, current research at the time offered little guidance as to how the SPD should implement that program. Limited and mixed evidence from the field exists regarding exactly how the actions of a supervisor can encourage officers to use LEED concepts in their daily encounters with citizens (Mazerolle et al., 2011; Melekian, 2012; Rosenbaum & Lawrence 2011). Furthermore, there is scant evidence on the effectiveness of the traditional EIS models used in the promotion of positive officer behavior (Walker, 2000; Walker, Alpert, & Kenney, 2001).

As such, the purpose of this project was to design an experiment in which we could develop and test the efficacy of a supervisory intervention approach using LEED concepts to reduce problematic/negative encounters between police and citizens and increase police legitimacy. Our goals were to contribute to the academic literature and to inform police practice by bridging the gap between academic theory and practical policing with regard to procedural justice. Our specific objectives included: a) to develop a selection process based on

geographically based risk, a process we call the High Risk Circumstance Model); b) to develop an experimental intervention involving LEED principles, which we refer to as an "experimental engagement"; c) to randomly assign officers within the HRC selections to either the "experimental engagement" or "control" condition; d) to evaluate the impact of procedurally just (LEED-based) engagements on officers performance (via several outcome measures); and e) to broadly disseminate the findings.

Literature Review

Procedural justice. Empirical assessments of the model of legitimacy and procedural justice, while primarily based on survey research, have supported the importance of legitimacy and the role of procedural justice in promoting it. Tyler and colleagues have conducted much of this research (Sunshine & Tyler, 2003; Tyler, 1990, 2001, 2003; Tyler & Fagan, 2008; Tyler & Huo, 2002; Tyler & Wakslak, 2004) finding with high consistency that (a) higher levels of police legitimacy predict higher rates of citizen compliance, cooperation, and law abidingness; (b) feelings about procedural justice most powerfully predict legitimacy when compared to feelings about favorability and fairness of outcomes; and (c) citizens' personal experiences of police processes have a strong impact on their general assessments of police legitimacy. A few survey-based studies by other researchers also lend general support to the legitimacy perspective (e.g. see Engel, 2005; Reisig, 2007; Reisig, Bratton, & Gertz, 2007).

Indeed, in order to ensure procedural justice and increase legitimacy, individuals accused of violating the law must perceive that (1) they are being treated fairly relative to others; (2) they are given the chance to explain or defend their behavior; and (3) their explanation is taken into account before any disciplinary action is taken. For example, Tyler's 1990 study of Chicago residents' satisfaction with the police found that peoples' satisfaction was not tied to the outcome

of an encounter with the police but rather to their perception of how they were treated, a result that has been confirmed more recently (see e.g., Schulhofer, Tyler & Huq, 2011), and which calls for a procedural justice alternative to the assumptions that have guided traditional policing in America.

Early intervention systems (EIS). Computer-based early intervention systems (EIS), also known as early warning systems (EWS), are data driven programs that can provide timely data on officer performance. As such, these systems are increasingly used by police agencies to identify officers who may benefit from mentoring or training in appropriate use of force techniques, cultural sensitivity, or other relevant interventions, or who should be more closely monitored (Alpert & Walker, 2000). For example, in 2003, 29% of law enforcement agencies surveyed by LEMAS reported using a computer-based monitoring system to identify at-risk officers. Four years later, 39% of surveyed agencies had such a system in place (US DOJ, BJS 2003, 2007). The diffusion of this technology into police organizations is undoubtedly due, at least in part, to incentives put into place by the federal government and in particular, the Civil Rights division of the Department of Justice has recommended the adoption of an EIS program in practically all of its reviews of police department practices.[3]

While the specifics of each system differ across agencies, an EIS has three main components. These include the identification of officers with problematic behaviors, followed by both an intervention to correct the behavior and a follow-up process to assess the success of that intervention (Alpert and Walker, 2003). Examples of problematic behaviors or common triggers often flagged by EIS systems include being named in civil lawsuits, being involved in traffic

[3] These reports are available online at: http://www.justice.gov/crt/about/spl/findsettle.php#Law%20Enforcement%20Misc. Recommendations on EIS systems are mentioned in all save the Pittsburgh Police Department review, which was the first conducted by the DOJ.

accidents, or taking an excessive amount of sick days. In some cases, positive indicators such as receiving commendations or compliments from the public are also included in EIS. Most police departments track up to 20 behaviors, which include both positive and negative actions, by using a cumulative threshold system to identify officers that cross that threshold (Walker 2003). With multiple indicators, EIS systems are thought to provide a broader and more accurate base of information (Walker & Graham, 1998; Alpert and Walker, 2003).

When any officer accumulates sufficient numbers of any particular behavior(s) during a pre-specified time period, which are often compared to other similarly situated individuals, the EIS alerts a supervisor or manager that the officer might benefit from some sort of intervention like mentoring or training.[4] What constitutes sufficient numbers of any given behavior varies by department, however. While some agencies use a department-wide model that sets a threshold at, for example, "three similar complaints in a year," relying on these simple counts of all officers exceeding that threshold, other more sophisticated systems "use more robust analysis to set the threshold utilizing the standard deviation calculation [and] in these systems, officers are compared to their peer officers who work in similarly situated work environments…" (Ortiz & Amendola, 2005). These alternative deviation models focused on quantity of numbers that cross a threshold identify officers who are observably different from officers working similar beats at similar times[5] and has become a expected standard for police. For example, in its reviews of twenty-six departments since 2004, the Department of Justice has tended to recommend

[4] It is important to note that when we refer to training in this proposal (other than when referring to the department's new supervisory training program), that it is not being used to indicate that direct provision of training is being proposed, but rather that it is for the purposes of implementing the experiment (e.g. training supervisors in how to conduct the intervention).

[5] This is an incorporation of place in EIS, but as we will describe below, it is mathematically very different from the way we will use place.

department-wide threshold models, where officers are identified based on a number of different behaviors.

Unlike most EIS systems, which are typically housed in an internal affairs unit, the High Risk Circumstance (HRC) model developed in this study was not used as tool to alert supervisors of problematic behavior; rather it was used to identify officers who run an elevated risk of encountering dangerous persons or situations. More specifically, unlike the typical EIS system, this model is designed to expand the focus from identifying officers exhibiting behaviors considered to be problematic to identifying officers working in behavioral "hot spots," or small geographic areas where they are more likely to be involved in problematic citizen encounters involving assaults on police officers, intoxicated persons, etc.

Furthermore, unlike with early warning systems where the most common type of intervention utilized for flagged officers is an informal counseling session between the officer and his or her immediate supervisor (Alpert & Walker, 2003), the design of this study utilized a similar type of intervention albeit with two key distinctions; that is, in the one-on-one engagements, officers and their supervisors often discussed an unremarkable event which would typically not be subject to discussion under standard practice. Additionally, not only were the officers prompted to reflect on their thought processes and actions during these encounters but their supervisors were trained to model LEED concepts for them in practice.

Methods

Sample

We sought to inform police management practices by drawing on cutting edge research in criminology on behavioral "hot spots," which has demonstrably helped police departments proactively reduce crime (e.g. Sherman, Gartin, & Buerger, 1989; Sherman & Weisburd, 1995;

Weisburd, 2008; Braga & Weisburd, 2010). Specifically, we used geographically identified data already collected by the SPD to identify "behavioral" hot spots which we define as street segments where officers appear to be more likely to have problematic interactions with citizens. We use the Washington State Department of Transportation's definition of a street segment, which is curb-cut to curb-cut. We defined problematic interactions as: (1) using physical force; (2) being the subject of a citizen complaint; or (3) sustaining a physical injury.

Using estimated parameters from this model, we then selected samples of officers who work in these behavioral hot spots. Note that only a small fraction of officers working in these "hot" street segments had actually engaged in any potentially problematic behavior. Indeed, the point of our model was not to identify officers who had engaged in problematic behavior but rather the point was to use geography to identify at-risk officers before an encounter occurred that could result in a problematic interaction with citizens. Officers identified by the HRC model as working in behavioral hot spots were randomly assigned to either a control group or an experimental engagement.

Identifying officers for the experimental engagement. As a first step in implementing the proposed project, the predicate trigger for randomly assigning sworn police officers to an experimental engagement was identified. We accomplished this through statistical analysis of the characteristics of police interactions, both routine and unusual, with a view towards ultimately determining the efficacy of the methods of engagement that cause or contribute to positive – or at least less problematic – encounters between the public and police. Our method of identifying officers for experimental engagement was inspired by the EIS currently being used by the SPD, but has an important conceptual difference; instead of identifying officers who have engaged in problematic encounters in the past, we sought to identify a larger pool of officers who,

statistically, were relatively more likely to engage in potentially problematic encounters in the future based on individual and contextual factors.

Instead of identifying officers based on any individual behavioral trigger, we identified officers for engagement based on their expected likelihood of being involved in a future problematic incident, which as noted above, we defined as either: (1) using force; (2) being named in a citizen complaint; or (3) sustaining physical injury. The expected likelihood of being involved in a problematic incident was calculated using a statistical model, described in detail below, that was calibrated to a small subset of an individual officer's actions, but primarily to the actions of other officers working in a similar environment.

Incorporating the experiences of similarly situated officers into an EIS is not an innovation per se, but the way in which we used the behavior of an officer's peers is. In the previously discussed deviation model, officers are identified by an EIS because their performance suffered relative to their immediate peer group. At first glance, using deviations from a group average as a way to identify low performing officers seems like a straightforward way to incorporate the external factors that affect an officer, particularly when there is imperfect information about individual officer activity (Walker, 2001).[6] This approach is potentially problematic, however.

The problem with using a deviation approach to identify poor performers is that the lowest performer in a group is not necessarily the lowest performer in the department. For example, consider a simple scenario where four high performing officers and four low performing officers patrol two beats. If each beat is patrolled by two high performers and two

[6] For the same reasons, this deviation approach is also a common technique used to test for racial profiling. Ridgeway and McDonald (2009) discuss this same critique of the deviation methodology in the racial profiling context.

low performers then a deviation based EIS system will work as intended; the two low performing officers on each beat will underperform relative to their peers. What if the first beat is patrolled by four high performing officers and the second beat is patrolled by four low performing officers, however? In this situation, no officers will be under performing relative to their peers, and the system will not identify any need for training; a group of officers will be dominated by low performers if they were all affected by an external factor, such as working the late shift in a popular bar district, or if they were all being poorly managed by the same supervisor. With regard to problematic officer behavior, we think there is good reason for concern about the deviation approach to officer identification.

In response to this concern, we proposed an EIS-style system that uses a place-based approach instead. Proactively identifying officers who work in areas where we know officers run into difficulty may be more effective at preventing problematic behavior before it starts. By positively emphasizing the characteristics of the places that officers work in, rather than discounting problematic encounters in areas where many officers run into difficulty, we explicitly incorporated advances in criminological theory, which has elevated the role of places and areas in crime, rather than simply focusing on individual characteristics (Eck & Weisburd, 1995, Sherman & Weisburd, 1995).

Behavioral "hot spots." Specifically, we intended to identify "behavioral hot spots." These behavioral hot spots had both a temporal and spatial dimension, in that we sought to empirically identify both geographic areas and times of day where officers appeared to be more likely to engage in potentially problematic behavior. For example, a city block that is home to many bars might be a behavioral hot spot during a 10 pm to 5 am shift, but not during an 8 am to 1 pm shift.

The high risk circumstances (HRC) model (overview). In order to distinguish these behavioral hot spots from criminal hot spots, and to emphasize the importance of the environment to any policing outcome, we refer to these areas as High Risk Circumstances (HRC). In lieu of targeting a few officers based on past performance, which is the focus of an EIS system, we focused on providing pro-active training for officers assigned to beats that, based on the experiences of their peers, could be more difficult to navigate. This is a critical difference from previous attempts to identify and correct problematic police behavior (Walker, 2003). The goal of the HRC model then is to identify a large number of officers who, based on the officer's own behavior, as well as characteristics of their work environment, may be involved in situations where they need to use force in the near future. Providing LEED training to officers who interact with citizens in HRC areas may help them avoid problematic performance in the first place.

The Intervention: Procedural Justice Experimental Engagement

Before we describe the engagement in more detail, it is important to carefully and purposefully describe the process that was proposed for promptly assigning officers and supervisors to either an engagement or control group. The term "engagement" was chosen with both care and purpose in mind. An "engagement," as opposed to an "intervention" is neutral in terms of consequences to the involved parties. The Seattle Police Department has an EIS that is a product of and is governed by collective bargaining agreements, and which is uniformly triggered when threshold incidents prompt the intervention. While discipline cannot be imposed as an outcome of EIS, the assignment to an early intervention is nevertheless, widely viewed as the result of an accumulation of negative, or at least problematic, incidents. At a minimum, the processes we proposed could not in any way be perceived of as an application of EIS without significant collective bargaining implications. More specifically, the SPD is precluded from

undertaking discipline in any manner other than that prescribed in the due process provisions of collective bargaining agreements between the City and SPD sworn employees. Furthermore, the quality and value of sworn employee participation in this experimental program would have likely been tainted if the outcome of the assignment had been viewed as a negative referendum on the employee's actions or decision-making.

Pre-engagement procedural justice training for lieutenants. Prior to the start of our experiment, all SPD lieutenants were provided with classroom training and then asked to model LEED techniques for their sergeants. This entire procedure is detailed in **Appendix A** in which we describe the training lieutenants received in order to facilitate training of sergeants to conduct the experimental engagement.

Pre-engagement interventions. Prior to the start of our experiment, all SPD employees received a memo announcing the initiation of a new supervisory training program. This memo stated that over a six-month period, some officers would be asked to check-in with their supervisors to discuss their recent interactions with citizens. The memo also emphasized that the discussions would not be related to the officer's performance, but were intended to help the Police Foundation evaluate the new training method. Officers who were identified by our HRC system, and were assigned to the treatment (engagement) condition, were then asked to participate in a LEED incident walkthrough (the experimental intervention).

Assignment to treatment and control conditions. During the field experiment, we received an extract of CAD events from the Seattle Police Department Data Analysis Unit every two weeks. On the 9th of each month, we also received an extract of injuries, sick days, and secondary job permits from the Seattle Police Department's Human Resources Unit. The next

step was to construct event histories for all CAD calls, and estimated predicted risk scores for all priority 1 and priority 2 incidents, using the estimated logit parameters.

After predicting risk scores for all CAD incidents, we then excluded all officers identified by the agency as unlikely to comply. After excluding these officers, we then ranked the remaining officers in each precinct based on their highest predicted risk score over the previous two weeks. Officers in the 12th percentile of their precinct's distribution of risk were then randomly assigned to the engagement (treatment) or control group. A list of treated officers, along with the incident numbers of anywhere from one up to seven CAD incidents the officer was involved in (not necessarily high risk incidents) was then transmitted via secure FTP to the department who then notified the selected officers and their supervisors.

Officers selected by the HRC model and assigned to the engagement group were called in to meet with their supervisors in the presence of a lieutenant. The engagement can be thought of and described as a style of cognitive interview. "The cognitive interview technique is based on laboratory-tested principles of memory retrieval, knowledge representation, and communication" (Fisher, Geiselman, & Amador, 1989, p. 722) and was used by Fisher and Geiselman (2010) to increase eyewitness memory of crime victims and witnesses. In essence, the cognitive interview promotes a means of questioning that is more open. According to Fisher, et al. (1989):

> The Cognitive Interview is a set of instructions given by the interviewer to the witness at the beginning of the interview. The goals of these instructions are (a) to encourage the witness to reinstate the context of the original event and (b) to search through memory by using a variety of retrieval routes (see Geiselman, et al., 1985, for specific details). (p. 722).

As such, the sergeants and officers interacted in a dynamic way, with each person's behavior influencing the other. For the interview to be considered successful, the two members needed to have coordinated their roles effectively while remaining sensitive one another.

The LEED engagement (experimental intervention). The experimental engagement was designed to be a short, 20 to 30-minute demonstration session by the sergeant in procedural justice for the officer, based on the officer's own experiences. While not an intensive treatment per se, the engagement was expected to "double-down" on LEED training. Not only would the sergeants point out where LEED could have been used but they would also model procedural justice for the officer during the engagement. This design not only provided the pedagogical benefit of modeling the desired behavior but in adhering to the tenets of procedural justice (such as making the officer feel validated and in control of the outcome of the encounter), the potential effectiveness of the engagement was expected to be maximized (Tyler & Fagan, 2008).

We hypothesized that LEED engagements would provide an important experience to officers who were identified by our HRC system, even if they had personally had only positive encounters with citizens. One of the benefits of cognitive interview techniques outlined in Fisher and Geiselman (1992) is to enhance the interviewee's ability to retrieve memories. Encouraging an officer to remember and reflect on an uneventful encounter with a citizen reminds him/her of both the successful and unsuccessful actions he/she took that contributed to the outcome of that interaction – what was it about this encounter that kept it from escalating into a more serious situation? In retrospect, when did LEED techniques help an officer? Were there any missed opportunities to use LEED? Instead of being mentally discarded, the particulars of a mundane

citizen interaction in a high-risk circumstance can be utilized to remind officers how to behave in the future.[7]

At the beginning of the engagement, the sergeant followed standard introductory instructions adapted from Rosenbaum and Lawrence (2011) and Fisher and Geiselman (1992). In essence, the officers were reminded that this interview was intended to help the supervisor sharpen his/her training skills, rather than discipline the officer. Officers were then asked to discuss an event that happened during their shift on a particular day, where the day in question contained an event with a high- predicted risk value. The entire script provided to sergeants for use during the experimental engagements is provided as **Appendix B**.

In order to comply with IRB protocol and union regulations, once the sergeant transferred control to the officer, the engagement was deemed "complete," and the officer could choose to terminate the meeting without any further discussion. After the supervisory meeting occurred, the supervisory lieutenant reported the date of the engagement back to the department.

If the officer chose to continue, the rest of the engagement was designed to follow a structure that included either a physical or mental walk through, and focused on the methodology of the interviewer – a first-line sergeant who, as we will discuss below, was specially trained in conformity with the LEED principles. Specifically, as the sergeant walked through the incident with the officer, he or she was to explore whether there were opportunities to undertake – or impediments to introducing – the actions of listening and explaining, and whether, in the officer's view, the encounter possessed the qualities of equity and dignity. While the sergeant

[7] In situations where the officer does recount a stressful or negative citizen encounter, participating in the LEED walkthrough may provide the officer with an additional direct benefit. A long literature in psychology and management, dating back to the 1950s argues that people who participate in LEED interviews after stressful events are better able to handle demanding encounters, and feel more self-control in extreme situations (Kaplan, Iancu, & Bodner. 2001; Fisher & Geiselman, 2010).

was conducting the interview, lieutenants were present as supervisors for the sergeants. The presence of the lieutenant served two purposes. First, it reinforced the institutionalization of the procedural justice training already provided by the Seattle Police Department. Second, the trained lieutenants were available and able to intervene if a sergeant strayed from the procedural justice model during the interview, protecting experimental integrity.

The debriefing and walk through focused specifically on what happened and what the officer observed or believed, and how he/she responded. The sergeants were instructed not to pressure the officers to speculate about details, or observations about which they were unsure of or confused about. The sergeants were also instructed to expect gaps in conversation and silence while participants processed and searched through their memory and attempted to recall details and emotions. Indeed, as Koriat and Goldsmith (1996) have noted, it is important for participants to communicate only from recollections and not guess about details. Since recall accuracy is influenced by the format of the question, open-ended questions such as "describe the person," rather than "what was the person wearing or how tall was she" aimed to help officers recall the information. While sergeants were allowed to ask follow-up questions, the idea of the engagement was to have the officer recall information first on his/her own terms.

This engagement put the instructional focus on the sergeant, and the didactic model involved a kind of role reversal, where the involved officer was allowed to examine the success of the interviewing supervisor or commander in adhering to the LEED methodology. It is in the realm of this engagement that the officer interviewed was in turn, allowed to interview the sergeant about his/her own application of LEED principles.

By construction, the officers in our engagements had a higher probability of being involved in a potentially problematic event, but it is important to note that they may not have

actually been involved in a negative citizen encounter. The purpose of the engagement was to remind officers how they could use LEED concepts on the job, while the sergeant simultaneously modeled procedural justice for them in practice. The fact that sergeants used the LEED concepts themselves is critical. We know that people are more likely to obey laws when police officers use LEED (Sunshine & Tyler, 2003; Tyler & Fagan, 2008); therefore, we proposed that officers would be more receptive to the engagements if their supervisors used LEED concepts with them.

Outcome Measures

Using administrative data already collected by the SPD, we compared the post-engagement behavior of officers who worked in behavioral hot spots to a number of dimensions during one and six week periods (short run and long run). In addition to a follow-up survey of officers immediately after they participated in the LEED engagement, our primary analytic focus was on tractable and policy relevant measures of police performance and safety in the field. Officers in the treatment and control groups were compared based on the number and type of incidents that they responded to, their probability of using force or having a complaint filed against them, and the frequency with which they resolved incidents with a formal arrest, rather than a less disciplinary measure. Note that some of these outcomes have low base rates, but others measured more routine aspects of an officer's job. These were also outcomes already collected by the SPD.

ANALYSIS

Given the large number of officers involved, the experimental intervention relying on the use of LEED concepts was less disciplinary or corrective than most supervisory meetings, and survey results from officers confirmed that they did not feel they were being penalized or

disciplined during these engagements. Instead of simply telling officers they should treat citizens with respect, the engagement was designed to "show" the officers what they should be doing to promote the ideas of procedural justice in practice. In the presence of a specially trained lieutenant, the sergeant and officer discussed a recent interaction that the officer had with a member of the public. During this talk, the sergeant not only used LEED principles in discussing the interaction, but also pointed out how the officer either successfully or unsuccessfully used procedural justice at the time; in particular, by adapting their actions to their observations of the situation at hand. Simultaneously, the sergeant modeled desired behavior by treating the officers with respect, equity, and dignity, offering the officer the chance to dictate the pace of the meeting, and thereby allowing them to have more control over what is discussed than was typical for the department. Our procedural justice intervention was therefore, two-fold; not only were officers being reminded about what procedural justice actually means in practice, but they were also experiencing LEED, rather than being lectured by their superior on it.

The High Risk Circumstance (HRC) Statistical Model

The first step in developing an HRC system was to define exactly what types of behaviors we wanted officers to avoid. These behaviors were the dependent variables used to calibrate our statistical model. In the case of the SPD, our primary definition of such behaviors consisted of incidents in which an officer filed a use of force report, incidents in which an officer was named in a citizen complaint, and incidents in which the officer was injured.

We stress that use of force and citizen complaints are not always indicators of poor officer performance. Indeed, more often than not there is a legitimate and justifiable reason for an officer using force, and a high frequency of citizen complaints may reflect a retaliatory group of arrestees, or a department that is viewed as open and responsive by the public. We defined

these events as high risk, however, because the Department of Justice recommends that *any* use of force, and *any* citizen complaint, regardless of its legitimacy, be used as a precursor to an intervention in an EIS system (US DOJ, 2011). As this is how the DOJ defines a problematic encounter, we formally defined problematic behavior in the same way.

In any given week, there was a very low probability that an officer would engage in one of these risky events. In 2010, SPD officers reported using physical force of any sort in only 0.12% of all of their interactions with citizens, and only in roughly 2.6% of all arrests (Walsh, 2011). Out of approximately 1,261 SPD officers patrolling 51 beats that year, just about a third (n=461) used force at all, and only half of that group used force more than once. Given these statistics, we designed the HRC model in a way that identified officers who responded to any incident that had a high level of predicted risk, rather than an incident that actually involved a potentially problematic event.

During the experimental period, a predicted risk score was assigned to all CAD incidents on a biweekly basis. The predicted risk of any given incident involving officer 'j' was based on characteristics of all incidents occurring in or near the incident in the past 30 days, along with a rough measure of secondary jobs, sick days, and overtime worked by the individual officer over the previous calendar month. We defined events "near" a particular CAD incident as all events that occurred in the same census block and all events occurring on the same street segment as the CAD incident.

In order to construct the predicted risk score based on these location and officer-specific elements, we used data on all CAD incidents from 2010 to predict the likelihood that a potentially problematic event occurred. A potentially problematic event was defined as an incident in which an officer used force, an officer was injured, or a complaint was in the process

of being filed or was filed against an SPD officer who was working in that area on that day. In collaboration with SPD employees, we identified a number of data fields such as the number of 911 calls initiated from that area or the number of officers injured on that street segment during the previous month (see **Appendix C** for the complete list), which were used to predict risk and uncertainty associated with an encounter. All of these data fields are collected by the SPD as part of their normal operations.

We modeled predicted risk in four ways: a simple linear probability model, a logit model, a probit model, and a skewed logit model. We then compared the spatial distribution of predicted risk scores to the spatial distribution of actual risk scores across precincts and reporting areas using CAD and HR data from February and March of 2012.

In terms of capturing the relative frequency of complaints, force, and injury across districts, a standard logit model appeared to actually fit the data better than the other models. For example, there were 104,821 unexceptional CAD incidents in the SPD in February and March of 2012. The mean predicted risk score of these events was 0.0029, with a standard error of 0.000049. On the other hand, there were 309 incidents that were potentially problematic, and the mean predicted risk score of these events was 0.0442, with a standard error of 0.0126. Based on a t-test of the equality of these means, we are confident that there is no chance that these predicted risk scores are the same given that the risk score for problematic events is more than 71 standard deviations away from the mean of the unexceptional events.

We furthermore estimated the average predicted risk score, and potentially problematic event rate, in the different precincts and sectors of the SPD in those same two months, February and March 2012. These results are displayed in Table 1. The overall correlation in actual and predicted risk scores is 0.27 indicating that for the most part, places with higher risk scores do

have a higher rate of events within a precinct (for example, sector N in the North Precinct), but this is not always true. For instance, when the predicted risk score was different from the actual rate, this indicated that something excluded from the HRC model was important in predicting the rate at which officers were involved in potentially problematic events. For example, it could be the case that there were a large number of reports involving excessive alcohol consumption in sector U, which includes the University of Washington, but that interactions with college students rarely escalated to the point at which officers used force or were injured.

< Insert Table 1 About Here >

We then saved the estimated parameters of the logit model of potentially problematic circumstances. These saved parameters are reported in Table 2. Since the logit parameters themselves are not directly interpretable, in Table 3 we also present the estimated coefficients from a linear probability model along with the total impact the change in that particular data field would have on the total risk score for that incident.

< Insert Table 2 About Here >

< Insert Table 3 About Here >

For the most part, the signs and magnitudes of the HRC model are intuitive. At the census block level, there is a positive relationship between potentially problematic events and the number of 911 calls, incidents involving someone in mental distress (including alcohol or drug use), domestic violence, incidents identified by dispatchers that are higher priority than usual, and the frequency with which dispatchers identify an incident as less serious than the officer ultimately decides. Typically, conditional on the characteristics of the census block, the correlation between the incident's history at the street segment is actually negative; and, counter to the observed finding that crime "hot spots" are extremely small, this is more consistent with a

model in which reports of officer injuries or use of force incidents are general such as officers learning that someone was injured on the 900 block of Cherry Street or near a city park, rather than on 910 Cherry Street.

That said, while the net conditional effect of street segment level activity tended to reduce predicted risk when we focused on marginal effects that are statistically different from zero on their own (as reported in the 4th column of Table 3), street segment level activity was typically positively correlated with risk. For example, the net conditional correlation between domestic violence reports at the street segment level and risk was negative, but the correlation between street segment domestic violence incidents at the same time of day and the same day of the week was positive and highly unlikely to be zero. There was also a positive correlation between predicted risk and the fraction of incidents cleared as more serious than dispatched on the same street segment and the same day.

Officer level human resources data, which was collected with a lag, was less associated with risk. Only the number of injuries reported on a street segment in the previous month was found to be a statistically significant predictor of risk, suggesting strong spatial correlation in violence against SPD officers.

Two practical distinctions between HRC and EIS are worth highlighting. By construction, the roughly 100 officers a month that were selected had predicted risk probabilities greater than the 12th percentile of their precinct. This is much greater than the number of officers who are actually involved in a problematic event. Indeed, the point of the predicted risk measure was to identify and engage with more officers than an EIS system would. Because the HRC identifies more people than EIS, the engagement spurred by HRC identification must be less intensive than an engagement spurred by EIS identification. The second key distinction between

the process of experimental engagements undertaken and EIS is that the random assignment to an experimental engagement was essentially immediate, while review spurred by EIS may occur as many as six months following the first triggering event.

Post-engagement check for internal validity. Immediately following the engagement, officers were asked to fill out a short, confidential, comment card. This card consisted of nine closed ended questions regarding the officer's perceptions of the engagement, as well as an open-ended question soliciting comments about the engagement experience. This served as a post engagement check for internal validity for the researchers. The comment card is attached as **Appendix D**.

While the self-reported experiences of officers with the engagements were not the primary focus of our research, a qualitative and quantitative evaluation of these responses provided us with some evidence on the practicality of LEED style training, although the response rate was particularly low. This will be detailed later on in this report.

Site monitoring. During the course of the experiment, the Police Foundation research team made regular site visits to ensure that the experiment was being properly implemented, and to check-in on the processes for scheduling engagements, get feedback from the sergeants and lieutenants on the process, and to identify any issues or concerns that needed to be addressed by the agency or researchers. The research team including the Co-PIs, were expected to make two to four site visits each. While inevitably, there were some unforeseen issues that arose over the course of the experiment, we did not anticipate nor did we experience a large amount of attrition or difficulty with data collection. We were confident of this because the SPD had already been planning to implement LEED training, and also because the experiment was conducted within the framework of ordinary officer supervision and performance monitoring. In addition, the

research team had numerous preparatory discussions and meetings with SPD command staff during the planning of the experiment. As such, the purpose of these site-monitoring visits was more oriented towards documenting any unforeseen issues that may have developed over the course of the experiment.

Police Foundation Project Manager, Dr. Karen L. Amendola, and Research Associate, Maria D. Valdovinos conducted two multi-day site visits during the data collection phase, a number of weeks after the program had begun. Two additional multi-day site visits were conducted by co-PI David Weisburd and PI Emily Owens subsequently. Between all four site-monitoring visits, a total of eight engagements were observed, in addition to meetings and follow up with the deputy chief, and other lieutenants and sergeants, as well as a few officers who had previously participated in an engagement. A procedural justice engagement monitoring checklist was created using the training model designed for SPD sergeants to conduct the experimental engagements and as such, served as a check of how "fresh" the training was for the sergeants. The engagement monitoring process and checklist is provided as **Appendix E**.

Observations. During the monitoring sessions, it became clear that the incident selection process was not clearly explained to the officers, sergeants, or lieutenants. For the sergeants and the officers selected for the LEED debrief, there was a strong assumption that any incident with a "high risk circumstance" would be inherently substantive, or involve an arrest or citation. This was specifically counter to the intent of the model, which explicitly excluded any characteristics of the particular incident in question, and was intended to identify officers who regularly worked in areas where there was a higher risk of a potentially problematic event occurring. At times, this resulted in the selection of incidents in which the officer selected for the LEED debrief was not primarily involved (arrived after the incident was over, or arrived as backup) or which didn't last

long enough to warrant much discussion. One particular officer had been in an engagement twice before the engagement observed by the research team, and all three had seemed highly irrelevant to her. She expressed a desire to have discussed another incident she was involved in, which in her opinion was more relevant.

The underlying cause of this confusion in implementation was the unintended consequence of trying to "jump start" the supervisory meetings by suggesting an incident to discuss that was based on CAD data. This was most problematic during the first experimental round, where in one event, the incident with the highest predicted risk score was suggested, and the wording of the notification implied that the sole purpose of the meeting was to discuss that particular incident. In all other rounds, multiple incidents were proposed, first 4, and eventually up to 7. An assistant chief communicated to the lieutenants that it was acceptable for officers and sergeants to select any incident to discuss. PI Dr. Owens reinforced this point in individual meetings during her second site visit. While this collaboration between supervisor and officer is in fact one of the intended goals of the LEED model, it initially appeared to be the case that SPD employees viewed this selection of incidents, many of which were relatively mundane, or something the officer responded to only as backup, as a failure, or problem, with the underlying experimental methodology.

The researchers also observed a certain level of discomfort among the sergeants with the script. They observed them read straight from the script, although it was really meant to be used as a guide. As such, in several of the engagements observed, the interaction between sergeant and officer seemed a bit awkward. Indeed, there was a certain level of discomfort from both the sergeant and the officer. The sergeants seemed uncomfortable with how to use the script as a

guide and the officers participating in the experimental engagement seemed uncomfortable with critiquing the sergeant on his or her execution of the LEED debrief.

In addition to the discomfort with the script, the research team also observed minor oversights in how the study protocol was followed. For example, in one particular debrief observed, the officer was not provided with a comment card. In another debrief, the assigned lieutenant failed to be present. An in another observation still, the officer debriefed was not aware of the LEED concept and what it stood for. The researchers agreed that refresher training was necessary to remedy these minor problems, and this sentiment was echoed and reinforced by department management personnel. Refresher training was conducted in week 6 of the experimental engagements.

EXPERIMENTAL EVALUATION RESULTS

Quantitative Analysis of Actual Field Outcomes

We now turn to the quantitative evaluation of the impact of LEED-based experimental engagements on officer outcomes, using administrative data regularly collected by the SPD. As described, randomization occurred at the precinct-wave level, creating 60 randomization blocks (five precincts, 12 waves). In order to construct a counterfactual outcome for each officer in the experimental condition that reflected this block randomization, we constructed performance measures from each of the eligible officers in that officer's precinct, using treatment dates for the officer. For example, in the Southwest precinct, the six officers who had responded to CAD incidents with the highest risk scores in the previous 14 days were identified as "eligible" in each experimental wave. In each wave, two were randomly chosen to be in the experimental group, and four were randomly assigned to the control group.

Summary statistics. As would be expected given the random allocation of officers to treatment and control conditions, across all precincts and all waves, the mean risk score for treated officers was very similar. For the treatment group, the mean risk was 0.116 (sd = 0.186), and for the control group 0.120 (sd = 0.195). Assignment was "blind" to whether or not the officers had been previously selected for either treatment or control. We should also note that the risk score of an incident is not related to the behavior of the responding officer or the outcome of the particular incident; rather, it is based on the history of events in the area of the incident, and the officer's previous sick days, overtime, and secondary jobs. This means that any plausible effect of previous treatments should not affect whether or not an officer was selected. Of all treatment notifications sent, 75% were sent to officers who had never been selected for an engagement (compared to 79.8% of control officers), 21.7% had previously been notified once (compared to 13.9% of control officers), and 3.2 of treated officers (compared to 5.4% of control officers) had been selected twice before. Consistent with randomization, treated officers had been previously notified of being selected for a supervisory meeting 0.008 more times than control officers in the same experimental block, which is statistically indistinguishable from zero (p=0.86).

In each experimental wave, all treated officers were simultaneously notified of their selection, but engagements frequently occurred at different times depending on the work schedules of the officers, sergeants, and lieutenants. Almost all notifications were sent out on Monday morning (some were distributed on the Friday before). Figure 1 displays the distribution of the amount of time between notification and engagement. On average, engagements happened 14 days after notification, and 43% occurred within one week after notification emails were sent. Not surprisingly, there were also a relatively large number of engagements that occurred 11 days

after notification, which corresponds to the second Friday after a Monday notification was sent out and the last Friday before the next experimental wave would occur.

< *Figure 1 about here* >

The fact that officers participated in the engagements at different points in time introduces some complexity into our analysis, in particular, how we measured "post" outcomes for our control group. Figure 2 is a graphical description of our approach, based on the Southwest Precinct example of two treated officers and four control officers.

< *Figure 2 about here* >

Panel A of figure 2 identifies the pre and post periods for each of the two experimental groups created in each wave in the Southwest Precinct, based on the date of notification. Panel B identifies the pre and post periods based on the date of notification and the dates of engagement. We evaluated the impact of early engagement on policing outcomes by comparing on-the-job activity measures for officers in the experimental condition before they had been notified that they had been selected for engagement and the behavior of the same officers during two subsequent time periods, after they had been notified and after they had the engagement. Our experimental counterfactual is the change in behavior of control officers in the same experimental group over the same time periods.

For expository convenience, we refer to the set of one treated officer and four control officers as one "experimental group." In the Southwest precinct, there were two experimental groups created in each wave, with a total of 10 officer level observations (two treated observations and eight control observations).

The comparison in Panel A (before and after notification) is similar to what is commonly referred to in the experimental literature as an "Intent to Treat" effect; in other words, some officers had engagements on the same day that they were notified of their selection, meaning that their next day at work was "treated," and other officers were not yet treated, but were exposed to treatment eventually. The comparison in Panel B (before notification and after engagement) is similar to the "Treatment on the Treated." This analogy is not exact, however, because in some ways, there are two different mechanisms at work; that is, officers who are notified that they have been selected for engagement may change their behavior in anticipation of a supervisory meeting, and the meeting itself may have a further change on their behavior. Differences in the two effects (after notification and after engagement) will reflect a combination of the differences between the intent to treat effect and treatment on the treated effect, plus a potential change in behavior due to anticipation of treatment. In addition to identifying two "post" periods, we also examined the change in officer behavior during two follow up windows: one week and six weeks after treatment.[8] Each time period is bounded by the last day the officer was recorded as being active in CAD before the day of notification and by the first day after notification (or engagement) that the officer was active in CAD.

< *Table 4 about here* >

Table 4 presents some descriptive summary statistics for our measures of officer performance, with one observation per officer all measured prior to the initial treatment notification date. Overall there were a total of 240 experimental groups in our sample, one for each officer who was treated, and a total of 1,562 officer observations across treatment and control groups. For the majority of our analysis, we restrict our sample to 1,434 observations, as

[8] More specifically, our follow up periods are 7 days and 44 days (one month and two weeks).

10 control officers were unlikely to be treated due to military, vacation, and sick leave. It is clear that there is very little difference across control and treatment groups on any of these dimensions, and in fact, the control and treatment groups are statistically indistinguishable with 95% confidence on all dimensions.[9]

Importantly, all of our outcome measures are recorded in existing data systems (CAD, AIM, and RMS) maintained by the SPD. SPD employees already collect these data in real time, and we had a plan in place to extract the relevant, officer-identified data with the SPD. We believe that the actual, measurable field performance of officers is ultimately the goal of any police training program.

Statistical power. We also include a power calculation in Table 4, where we estimate the probability of rejecting the null hypothesis (of no treatment effect) with 90% certainty, if the non-disciplinary supervisory meetings actually cause a 10% change in officer behavior. Our tests of officer activity have high levels or power, almost all over 80%. Our tests for incident response have lower levels of power, meaning that a failure to reject a null hypothesis should, for the most part, be interpreted as a failure to draw any conclusions, rather than finding a null effect of the engagements.

On average, selected officers responded to 40 CAD events in the one-week period before they were notified of their engagement. Of those incidents, the officer initiated 32% of them, and officers spent an average of 40 minutes on them from initial dispatch to their return to service. On average, 30% of incidents were serious enough for the officer to file a report, and 6.5%

[9] This was statistically verified by re-estimating 20 modified versions of our central outcome equation, where the dependent variable was $PreOutcome_{ijw}$ for each of our treatment windows and all outcomes. None of our pre-notification outcomes are statistically distinguishable with 95% confidence, and only one (CAD events resolved by issuing a non-criminal citation over one week period) was statistically precise at the 90% level of confidence (p=0.083).

resulted in an arrest being made; however, officers resolved most incidents simply by rendering assistance. In a given week, 0.003 complaints were filed against each officer, and each officer filed 0.2 reports about being involved in a use of force incident.

When we examined the six-week period before the treated officers were notified of their selection, we found that officers were involved in more CAD events, had more complaints filed against them, and were involved in more incidents where force is used, which is what we should see, since we are looking at a longer time period. Turning to the measures of typical incident outcomes for each officer, we find less of a difference from the shorter time window.

Quantitative Evaluation

Since we used a block randomization design with multiple experimental waves, a simple pre and post comparison of the mean outcomes would not take the experimental design into account. Instead, the proper way to identify the impact of treatment is to adjust the comparison of control and treatment outcomes for the experimental block level randomization, and for the fact that each individual officer appears in the sample multiple times. We have repeated observations per officer for two reasons: 1) because of the repeated use of control officers across the experimental blocks, and 2) because the same officers could be selected in multiple waves of the experiment.

The mathematical expression that describes both of these adjustments is below:

$$PostOutcome_{ijw} - PreOutcome_{ijw} = \alpha_{jw} + \beta Engaged_{ijw} + \varepsilon_{ijw}$$

Where $PostOutcome_{ijw} - PreOutcome_{ijw}$ is the pre-post difference in outcomes for officer i, in experimental group j, during treatment wave w. The experimental group specific intercept

α_{jw} essentially subtracts out the mean difference in outcomes for each officer in the experimental group. Note that this eliminates any variation in outcomes over time or across precincts that affects all of the officers in each experimental group, meaning that department-wide changes cannot be driving our results. The dummy variable $Engaged_{ijw}$ is equal to one for treated officers only, meaning that β is the average difference in the treated officers from the untreated officers in their experimental group across all experimental groups.

The remaining unexplained component of the difference in outcome, ε_{ijw}, captures all other possible confounds of officer behavior and, by experimental design, is assumed to be uncorrelated with $Engaged_{ijw}$. This unexplained component will be correlated across pre-post outcomes for the same individual officer, and failure to take this into account will overstate the statistical precision of our estimates. We statistically corrected for this by clustering our standard errors at the officer level, explicitly allowing for this within-officer correlation. As a result, the estimated standard errors of our experimental effects are based on the number of unique officers in the experiment, rather than the number of observations.

The outcomes that we analyzed can be divided into two conceptual groups. "Officer Activity" measures reflect an officer's engagement with the community. These include the number of CAD incidents the officer was involved in, the fraction of CAD incidents initiated by the officer ("on-views"), the average number of minutes an officer spent on-scene per incident, and the fraction of incidents for which the officer filed a report. A reduction in any of these measures would indicate that additional supervision is associated with lower levels of general officer activity and engagement with the community at large, which we define as de-policing. We also examined "Incident Outcome" measures, which included the fraction of incidents that resulted in an arrest, use of force, and citizen complaints. While a reduction in any of these

measures is not a clear policy goal, an excessive number of arrests, force, or complaints relative to the rate at which officers interact with the community in a more positive manner, is likely to reduce perceptions of police officer integrity. We also broke down the non-arrest outcome into citations issued, verbal warnings given, and assistance rendered.

Relative to arrests, use of force and complaints are infrequent events, and the low rate at which these occur makes it difficult to interpret the results of traditional statistical analysis techniques. Because of the low frequency of these particular incident outcomes, we also calculated an even longer run analysis period for these events, which consisted of the total number of force incidents and complaints recorded between May 1st and December 31st 2013 for each officer before and after treatment and notification.

Results. The following are the findings related to officer activity, incident outcomes, alternative means of case resolution, and effect heterogeneity.

Finding #1 – Officer activity. We do not find evidence that additional non-disciplinary supervisory meetings result in major de-policing. When officers learn they will be having a supervisory meeting, they may interact with people slightly more**.** Table 5 displays our estimates of the impact of LEED meetings on officer activity measures at one and six weeks before and after treatment. While officer activity is not necessarily a direct outcome of a LEED based supervisory meeting, one potential adverse impact of supervisory meetings is that officers respond by "de-policing." While we are not aware of any formal evaluation regarding this type of change where officers stop initiating encounters with citizens and attempt to limit their time interacting with citizens when they do respond to calls, we are aware that it has been reported as a response to Early Intervention Systems or Early Warning Systems (Amendola, 2003).

< Table 5 about here >

In contrast to anecdotal reports on Early Intervention Systems, we find no evidence that treated officers reduced their activity in response to being notified about, or undergoing a LEED based supervisory meeting. In fact, our estimates suggest that, relative to otherwise similar officers, treated officers may have checked in on about two more incidents in the week after being notified or treated, although there is a 12 to 20% chance this difference could have been observed at random. When we extend our time period of analysis to six weeks before and after LEED meetings, we find even less evidence of a change in overall activity.

We also do not find evidence that SPD officers who have LEED based supervisory meetings are any more or less likely to initiate citizen encounters (see Table 6). When we look at the fraction of CAD incidents that are "on-views," we observe a slight increase in officer activity in the week after they are notified that they will have a supervisory meeting, which is statistically significant at the 10% level. However, in the longer term, we do not observe any substantively meaningful increase or decrease in the propensity of treated officers to engage the public, relative to their peers.

< Table 6 about here >

Consistent with the absence of evidence that adding these additional supervisory meetings reduced the amount of contact SPD officers had with the public, we also do not observe a statistically significant change in the amount of time that SPD officers spend on call after being notified of or participating in an experimental engagement (see Table 7). Of course, our point estimates do consistently imply that engaged officers increased the amount of time they spent on scene by between 1 to 3 minutes, and there is only a 7% chance that the 1.6 minute increase in the average time spent on scene after participating in an experimental engagement would occur

simply by chance; however, this increase is very small relative to the pre-notification mean of 39 minutes from time of dispatch to when the officers returned to service.

< *Table 7 about here* >

Finally, we examined whether or not treated officers became less likely to involve themselves in serious incidents, where the "seriousness" of a given incident is determined by whether or not the officer filed a written report. Officers are required to file written reports every time they make an arrest, but also if they issue a citation or a ticket. Officers are not required to file a written report if they simply assist a person on scene, but filing a report does indicate that the officer considered the incident to be subjectively important enough to thoroughly document.

We found no evidence that LEED based supervisory meetings affected the probability that officers filed written reports about their on-the-job activities (see Table 8). Not only are the point estimates very close to zero at one week and six weeks after treatment, but the differences that we do observe would be expected to occur over 50% of the time in the long run if the intervention truly had no impact on the propensity of an officer to file a written report.

< *Table 8 about here* >

Overall, we conclude that concerns about LEED supervisory meetings resulting in de-policing are not supported by our experimental evidence. Comparing the activity of officers who were randomly assigned to have additional supervisory meetings with similar officers who were not selected, treated officers were involved in roughly the same number of incidents, initiated those incidents at roughly the same rate, spent about the same amount of time on scene, and appeared to be equally likely to file written reports after the encounters. Importantly, the estimated standard errors on these null estimates are also small relative to the sample means.

Failing to reject the null hypothesis of no treatment effect is not the same thing as concluding that there is no effect of treatment. If our standard errors are very large, it is possible that our "null" results would be consistent with substantial reductions in police activity. A holistic way to think about statistically insignificant results is to use the estimated standard errors to construct 95% confidence intervals around our point estimates. The upper and lower bounds of these confidence intervals reflect the largest possible treatment effects that could plausibly exist, based on our observed data. In other words, we now ask the question: how much de-policing could have actually occurred such that we would still generate our experimental results?

Based on the estimates and standard errors in Table 5, we can conclude with 95% certainty that the non-disciplinary LEED based supervisory meetings could have caused no more than a 2% reduction in the total incidents that an officer was involved with. Over a six-week period, we can rule out any more than a 6% reduction in total CAD activity as inconsistent with our data. With respect to officer-initiated incidents, the results in Table 6 suggest that there was at most an 11 % reduction in "on views" one week after participating in an experimental engagement. Over a six week period, we can reject any adverse treatment effect size larger than an 8% decrease.

Our results in Table 7 also allow us to rule out more than a 6% reduction in time on scene in the short term, and any effect size greater than a 2% reduction over the longer term is also statistically improbable. Finally, with regards to the probability that officers deem an incident significant enough to file a written report, in Table 8, we find no statistically significant change and can also rule out anything more than an 11% reduction in the probability that treated officers filed a report after one week. Over a six month period there was, at most, a 5% reduction in report writing.

Overall, we conclude that there is no evidence that non-disciplinary LEED based supervisory meetings affect the amount of engagement that officers have with the community. Of course, there are other ways in which police officers could engage in "de-policing" that could be elicited through other methods, such as detailed surveys. However, in terms of actual data that the SPD collects, and official metrics used to publicly characterize community engagement, we find no evidence of an adverse effect of these meetings, and are able to rule out very moderate effect sizes. We now turn to what the officer actually did while on scene by examining the final disposition of these incidents, as well as reports of force and citizen complaints.

Finding #2 – Incident outcomes. We find evidence that non-disciplinary supervisory meetings (the engagements) result in a reduction in the frequency with which officers resolve incidents by making an arrest. We also find evidence that, in the longer run, officers who have non-disciplinary supervisory meetings may be less likely to be involved in incidents where an officer uses physical force.

Arrests. The first outcome that we used to measure whether or not LEED supervisory meetings changed the way in which officers interacted with the public was the probability that a given CAD incident ended in an arrest (see Table 9). While reducing the number of arrests that officers make was not a goal of the experiment, if officers who have additional training in procedural justice techniques are able to resolve conflicts or suspicious scenarios without using their arrest powers, this may promote public perceptions of police integrity and fairness in the long run. Moreover, from the perspective of criminal justice costs, if incidents can be resolved without an arrest, the state is saved the additional costs of criminal justice processing, and the individual and their family does not incur any of the social costs (e.g. eligibility for subsidized housing, financial aid, employment restrictions) associated with potentially having a criminal

record. While we failed to find evidence of de-policing on the part of officers, the outcome of interest is the fraction of CAD incidents that end in arrest rather than total number of arrests, which officers could reduce by not engaging with the public.

< *Table 9 about here* >

We do find evidence that officers who participated in non-disciplinary LEED based supervisory meetings are less likely to resolve citizen encounters in non-disciplinary ways. While we do not observe a systematic change in the number of arrests per CAD incident made by officers one week after they are notified that they will be selected for participation in the experimental engagement, when we compare their CAD incidents before they were notified to after they have the supervisory meeting, we do observe that incidents are about 1 percentage point less likely to be resolved in a disciplinary way. When we compare how treated and control officers resolved CAD incidents over the 6 weeks before and after the engagements, we find that this 1% reduction is quite constant, and there is less than a 6% chance that this reduction could simply be due to chance.

Use of force. We now turn to the likelihood that treated officers are involved in incidents where force is used. Whenever an officer uses force to gain control of a citizen, that officer is required to fill out a specific use-of-force incident report. In addition, any officer involved in that incident is required to file a report. All of these filed reports are then reviewed by a series of supervisors in the SPD, and a committee determines whether or not the force was justified and is consistent with SPD policy. Our outcome is simply whether or not an officer filed a use of force report- not whether or not the officer actually used force, or whether or not the force was justified or deemed consistent with SPD policy.

We found little change in the likelihood that engaged officers, or those in the experimental condition, are involved in force incidents one week before or after engagements (see Table 10). This should not be surprising, as the likelihood that there would be any incident involving force over a one-week period is very low (fewer than 0.02 events on average, or less than 0.05% in any CAD incident). However, once we expand our follow up period to cover the six-week period after notification and engagement, we estimate that engaged officers are roughly 60% less likely to be involved in force incidents than control officers, and that there is only a 13% chance that a reduction of this size would be observed by chance. While not considered statistically significant by conventional standards, given the importance of force in public perceptions of legitimacy, the relative low power of the test, and the relatively low cost of this intervention, we consider this result highly promising, and suggestive of further experimental analysis.

< Table 10 about here >

Finally, we expanded our time frame to include the entire sample period, reducing the number of zero incidents that we observe in the data. Essentially, we are now comparing the temporal distribution of force events across control and treatment groups, and determining whether or not treated officers are less likely to be involved in force incidents at any time after their LEED based, non-disciplinary supervisory meeting when compared to the likelihood of the same prior to the meeting. Roughly one out of every two officers working in the highest risk circumstances are involved in events with force prior to engagement, and we estimate that there is essentially a 50% reduction in the likelihood that treated officers are involved in these potentially problematic situations after engagement. The precision of our estimates suggests that it is highly unlikely that we would observe this pattern of behavior simply as a matter of chance.

Our previous measures of officer activity and incident outcomes were all continuous variables, which primarily ranged from zero to one. Use of force, in contrast, is a much more restricted variable that takes on one of six values. If the observed average reduction in force was driven by one officer who was involved in six incidents prior to engagement and none afterwards, this might limit the interpretability of our results.

< Figure 3 about here >

In figure 3, we plot the distribution of the number of force incidents that control and treated officers are involved in prior to the treated officers being notified of selection. These distributions are quite similar, with most officers not being involved in force incidents with the exception of a small number of officers in both groups using force multiple times. In figure 4, we plot the pre-post difference in involvement in force incidents. This figure suggests that in fact, our identification of a reduction in force incidents appears to be driven by multiple treated officers engaging in roughly one less use of force incident after engagement, rather than one treated officer making a large change. This type of response seems more plausibly a result of additional supervisory meetings.

We confirm that outliers are not driving our results by replicating Table 10 using a "trimmed" sample of force incidents (See Table 11). In this sample, any officer who was involved in three or more incidents where force was used was recorded as being involved in exactly three force incidents. Not only did this help mitigate the influence of outliers but we also essentially ignored any change in the frequency with which officers who regularly use force do so (e.g. a change from five incidents to four incidents is treated as no change at all).

< Table 11 about here >

Even in this trimmed sample, we still observe a statistically unlikely reduction in the long run use of force by officers who had non-disciplinary LEED based supervisory meetings. In fact, the point estimate and precision is only slightly different from the full sample. This confirms the graphical analysis in figures 3 and 4, which suggested that LEED based supervisory meetings reduced the rate at which officers who were relatively less likely to use force became involved in it at all.

Because not every officer who filed a use of force report necessarily used force, our results must be interpreted carefully. Our data do not allow us to determine exactly which officer who filed a force report actually did, in fact, use force against a citizen; rather, what we observe is that the officer was involved in an incident where one or more officers used force to regain control of a situation. Therefore, what our results tell us is that officers who have had non-disciplinary LEED based supervisory meetings are less likely to be involved in an incident where any of the involved officers feel that physical force must be used to regain control of the situation. Only one officer using force may cause multiple officers to file force reports, but if any one officer is able to maintain control of a situation without resorting to physical force, it is plausible that no force will be used at all.

Our final incident outcome is the rate at which citizens file complaints about particular SPD officers. Like our measure of force, we only observe that a complaint was filed against an officer at a particular date; we do not observe whether or not the complaint was substantiated or not, and we are very limited in our ability to link a complaint to a particular CAD incident given that the people filing the complaint frequently do not report and likely do not know the specific incident number. As such, we merely observe that a complaint was filed in a particular time period. It is possible that complaints that happened prior to notification are complained about in

the "post" period. This sort of measurement error should affect both our control and treatment groups equally, and therefore, is not anticipated to affect our experimental estimates.

An additional source of measurement error is the possibility that citizens will file complaints about incidents that occurred in the "post" period after our sample window. If LEED based, non-disciplinary supervisory meetings affected officer behavior in a way that changed the amount of time it took for citizens to file complaints, this type of measurement error would affect our estimates. As our sample window gets larger, this type of measurement error is less likely to be a problem because we will be allowing for a longer period of time between incident and complaint.

While we did find evidence that officers were less likely to make arrests, and less likely to be involved in incidents where force was used, we find no evidence that officers who participated in experimental engagements have complaints filed against them at different rates than officers in the control groups (See Table 12). All of our point estimates are positive, and represent a reduction of roughly 15% of the pre-notification control mean on average. However, these estimates are also statistically imprecise, and we could not reject the null hypothesis that LEED based, non-disciplinary supervisory meetings increased complaints by more than 90%, or fell by over 70%, with 95% certainty. Therefore, we must conclude that the available evidence does not allow us to say anything about whether or not these additional meetings had any effect on the frequency with which citizens file complaints against SPD officers.

< Table 12 about here >

Finding #3 – Are officers systematically resolving incidents in other ways? Our results suggest that officers who have LEED based, non-disciplinary supervisory meetings are less likely to resolve CAD incidents by making an arrest, and are also less likely to be involved in use

of force incidents. If officers are systematically resolving incidents that previously would have resulted in an arrest in a less punitive way, then we might observe an increase in the fraction of CAD events resolved by issuing a non-criminal citation, a warning, or rendering assistance. There are also many other ways that police can resolve an interaction, although these four (arrest, non-criminal citation or infraction, warning, and assistance rendered) are the most common, and almost 90% of all incidents are resolved in this way. However, given that arrests are so infrequent (roughly 7% of all incidents), a large reduction in arrests may not produce a statistically detectable increase in any other non-arrest outcome.

Non-criminal citations. We begin by looking at the probability that an officer issues a ticket or citations (See Table 13). After making an arrest, this is plausibly the most serious way that an officer could resolve in an incident; a formal record of the encounter with the individual is made, but there is no sense in which the individual is taken into police custody. We find no strong evidence that officers are systematically more or less likely to resolve incidents in this way after engagements, and can rule out with 95% certainty a 10% increase or 10% decrease in the short run, and anything larger than a 3% increase or 5% decrease in the longer run.

< *Table 13 about here* >

Verbal warnings. Turning to verbal warnings, we also fail to find statistical evidence that officers are systematically more likely to let citizens go with a verbal warning (see Table 14). In fact, the point estimates are consistently negative, so although no difference is statistically precise, there is a suggestive pattern that engaged officers may be more likely to document any citizen encounter in a formal way, rather than resolving incidents informally. Therefore, we cannot rule out potentially large short run changes in warnings issued (from a 40% increase to a 27% reduction), which is consistent with the power analysis in Table 4 In the longer run, there

may have also been an 11% increase in the probability that an officer clears an incident by issuing a verbal warning, but there could have also been a 17% reduction; thus, we can rule out any change larger than these two extremes.

< *Table 14 about here* >

Rendering assistance. The final outcome that we examined separately is entirely non-punitive; officers resolving incidents by reporting that they helped the citizen (See Table 15). This is how officers resolve the plurality of CAD calls (48% of all CAD calls in our sample are resolved this way). Our experimental results suggest that officer are potentially are between 3 to 13% more likely to resolve outcomes in this way after having LEED based non-disciplinary supervisory meetings, but none of the increases are statistically different from zero.

< *Table 15 about here* >

Finding #4 – Do we observe substantial effect heterogeneity? Finally, we examined our data for potential sources of heterogeneity. Specifically, we allowed for the treatment effects to vary by the number of times an individual officer was notified of their selection for treatment (up to three times), the value of the predicted risk score of the qualifying incident, precinct, and the experimental wave of assignment.

We find little evidence of systematic variation in the impact of additional non-disciplinary supervisory meetings across precincts or over time. We also find little evidence that more than one additional meeting over a six-month period has larger effects on behavior. There is some evidence that officers working in areas with the highest levels of predicted risk are less likely to change their behavior after additional non-disciplinary supervisory meetings.

There are four dimensions along which we expected to feasibly have heterogeneity in the impact of these supervisory meetings. Two are related to potential variation in implementation of the experimental supervisory meeting, and two are related to potential variation in the benefit of the meetings. We tested our central results (officer activity and incident outcomes) for each source of variation, and so it is important to keep in mind that, conservatively, the standard "critical thresholds" of statistical significance of 5% and 10% type-1 error probabilities are too high.[10]

Multiple supervisory meetings and officer activity. The first dimension we explored was whether or not officers who had multiple meetings behaved differently than those who had just one (see Table 16a). In practice, we tested this by allowing the impact of treatment to vary based on how many times, at the date of notification, the officer had previously been notified that they were selected for a supervisory meeting.

< *Table 16a about here* >

In panel A of Table 16, we present our results for officer activity. We find no compelling evidence that officers who previously had supervisory meetings responded any differently than officers having the first meeting. Even ignoring the potential for spurious statistically significant results, we find no statistically meaningful variation in the impact of additional meetings on the

[10] Based on a simple Bonferroni correction for multiple hypotheses testing, the appropriate critical values could be as much as eight times lower (as there are two primary hypotheses times four sources of heterogeneity). However, to the extent that many of these tests represent variation in integrity of implementation, this simple correction will likely lead to under-rejection of the null hypothesis. In addition, and also importantly, as of these tests was anticipated at the outset of the study, it is arguable that the tests should be considered independent trials. In some sense, these comparisons may be viewed as additional sensitivity tests of the data.

number of incidents, the fraction of incidents that are initiated by the officer, the time spent on scene, or the fraction of incidents for which a report is written.

Multiple supervisory meetings and incident outcomes. In panel B of Table 16a, we test whether or not the resolution of the incident varies with the number of meetings the officer was involved in. While there is some limited evidence that there is a diminishing impact of meetings on force reports during the week after engagement, these effects are less evident over the longer run six-week follow-up period (See Table 16b).

< Table 16b about here >

Higher predicted risk scores and officer activity. All officers that were eligible for treatment responded to CAD incidents in areas where other SPD officers were involved in potentially problematic events at higher rates than other officers in their precinct. In Table 17a, we include the first order impact of predicted risk score, and an interaction of risk score and selection for treatment. In order to ease interpretation, risk scores are standardized to have a mean of zero and a standard deviation of one.

A priori, it is not clear whether or not, within that group, officers with higher predicted risk scores were more or less affected by the engagements; what matters instead is the frequency with which officers were involved in events that are marginally problematic. Intuitively, there are some circumstances where officers will always make arrests or always use force, and some circumstances where neither of those two things would ever happen. Additional training in procedural justice techniques will have the largest impact on incident outcomes in cases where interactions may or may not become problematic depending on the officer's skill in handling the situation.

< Table 17a about here >

In the short run (panel A), there is no evidence that officers who work in more or less "riskier" areas are disproportionately affected by engagements. However, once we turn to longer run, six-week outcomes, we find that officers who work in areas with higher predicted risk scores have lower reductions in arrest rates. Specifically, on average, engaged officers are 0.7 percentage points (roughly 10 percent) less likely to resolve an incident via arrest; however, treated officers who work in areas that are one standard deviation higher in predicted risk are equally as likely to resolve an incident via arrest as similar officers who did not have an additional supervisory meeting. Similarly, officers who work in areas that are one standard deviation less risky than average are 20% less likely to resolve incidents via arrest than their otherwise similar colleagues.

< Table 17b about here >

Heterogeneity in supervisory meetings by experimental wave. We are including all 12 experimental waves in our results, removing any variation in outcomes over time with our randomization group fixed effects. While these fixed effects will absorb the impact of many potential threats to treatment validity that affected both control and treated officers, it is possible that there could be heterogeneity in the impact of supervisory meetings over time.

In Table 18a we allow for the average impact of supervisory meetings to be different in every treatment wave. In the short term, we find some evidence that officers who had supervisory meetings became more active in the first wave of the experiment and that relative to the average change, the treatment effects in later waves are consistently lower, and in a few treatment waves, the impact on overall activity is statistically lower than average. We also find that, in the short run, reduction in the use of force was concentrated in the early treatment waves.

< *Table 18a about here* >

In the longer run (Table 18b), we find less evidence of variation in treatment effect over time, although it appears to be the case that the reductions in arrests and force incidents were smallest in the earlier waves of the experiment (waves two through four). This finding is consistent with the qualitative analysis, which suggested that some SPD officers were, at least initially, skeptical as to the purpose of the supervisory meetings and in particular, were skeptical about why "small stuff" was being discussed. During site visits by project staff, sergeants and lieutenants had the chance to discuss the engagements, and this may have improved the integrity of the experimental implementation over time. In addition, the deputy chief and project staff refined the type and number of incidents that could be suggested as possible events to discuss in the engagements during the first weeks of the experiment.

< *Table 18b about here* >

Heterogeneity in supervisory meetings by precinct. Randomization occurred at the precinct level, and our main effects can be interpreted as averages across all precincts. In Table 19a, we allow for the impact of supervisory meetings to vary across precincts. These results suggest that, in the short run, officers in the East precinct (the omitted precinct) responded to treatment by spending more time "out of service" after CAD incidents on average. Of course, the observed effect, roughly 20 additional minutes on scene for treated officers, is so large that it suggests a statistical anomaly. However, the change in the time that East precinct officers are spending on scene is still evident six weeks after engagement, although the effect size is smaller and more in line with plausible estimates of an increase in time spent talking to citizens (an additional 3 minutes on scene).

< *Table 19a about here* >

There is no strong evidence that officers in different precincts responded to incidents in a differential way after having supervisory meetings. One exception to this is the observation that treated officers in the East precinct were also less likely to have complaints filed against them relative to the average. However, given the low rate of any complaints being filed, we conclude that it is simply not clear that LEED based supervisory meetings are related to the likelihood that Seattle residents file complaints against police officers.

< Table 19b about here >

Qualitative Analysis of Officer Responses via Comment Cards

The proposed experimental engagements involved increasing the amount of monitoring that officers were subjected to. While there may be positive outcomes from this additional supervision, it is important to weigh evidence of any desired changes in officer activity against the costly burden that these meetings place on the SPD. Therefore, we conducted a qualitative evaluation of these engagements in order to both evaluate the integrity of the experimental implementation and also to provide context for interpreting the quantitative results. It is quite common for social scientists to rely on a 5% level of type one error as a universal threshold for "significance." This broad-brush strategy ignores the fact that the actual cost of making a type I error varies dramatically across contexts. If engagements were viewed positively, or relatively innocuous by the officers, and there are signs of desirable effect on officer activity, then a higher level of type 1 error may be acceptable. In contrast, if engagements were viewed as particularly distracting or otherwise costly to officers, we would only consider improvements in officer activity to be meaningful for policy purposes if there was an incredibly low probability that they were generated by chance.

Response rate. There were a total of 67 comment cards received from participating

officers. While there was a box on the bottom left of the front of the comment card for the precinct to be filled in by the supervisors (not the officers themselves), it is clear that the supervisors did not get this message, as in any one precinct, the writing was in the same handwriting as the rest of the responses on the comment card. Of these 67 completed comment cards, only 64% (n = 43) included the precinct. For almost 7.5% (n = 5) responses, the front page of the comment card was left blank altogether, perhaps further demonstrating that the supervisors did not carefully instruct the officers in the completion of this task. It is not possible to estimate how the remaining 24 comment cards were distributed across precincts, nor whether the lack of completion had to do with concerns over confidentiality.

Assuming that there were equal numbers of interventions across the five precincts, we might expect a fairly equal distribution of responses/comment cards. However, the extent to which individuals respond to requests for feedback is motivated by a number of factors such as having strong views, feeling confident that one's individual responses cannot be identified, being encouraged by a supervisor, etc. In this study, the greatest number of precinct-identifiable responses came from the West precinct (n=15), with the remaining precincts ranging from 5 (South) to 9 respondents (East). The fact that over a third of responses that included a precinct identification came from the West precinct may reflect a higher level of expectations regarding confidentiality or some other factor. As such, we will examine the responses from that precinct compared to all others later in this section.

Item Analysis. The item analysis utilized the following response scale:

Response scale:

1 = not at all;

2 = to a very small degree;

3 = to some degree;

4 = to a pretty high degree;

5 = to a great degree; OR

n/a = not applicable

***Item 1*:** To what extent did this meeting with this supervisor make you feel like what you had to say actually mattered to this supervisor?

< Table 20A about here >

These responses indicate that for the most part, officers felt to a pretty high degree that the meetings made them feel like their responses mattered to their supervisors.

Item 2: To what extent did this meeting with this supervisor make you wish more supervisors in the SPD used this type of feedback approach?

< Table 20B about here >

These responses indicate that officers were not that enthusiastic about this approach, or that perhaps, they already feel that supervisors in SPD use this type of approach.

Item 3: To what extent did this meeting with this supervisor make you feel more respected than in past supervisory meetings?

< Table 20C about here >

These responses indicate that for the most part, officers did not feel that this approach made them feel that much more respected than in the past. However, it may be that officers already felt a certain level of respect with the existing approach, so that not much improvement was needed.

Item 4: To what extent did this meeting with this supervisor feel like a fair and just system for providing feedback to officers?

< Table 20D about here >

These responses indicate that officers predominantly felt that these types of supervisory meetings feel like a fair and just feedback system.

Item 5: To what extent did this meeting with this supervisor help to stimulate your memory of the event(s)?

< Table 20E about here >

These responses indicate that for the most part, officers felt that these meetings stimulated their memory of the events to a moderately high degree.

Item 6: To what extent did this meeting with this supervisor allow you to recall specific details of the event(s)?

< Table 20F about here >

Again, these responses indicate that for the most part, officers felt that these meetings allowed them to recall specific details of the event to a moderate extent.

Item 7: To what extent did this meeting with this supervisor allow you the chance to explain your perspective without feeling judged?

< Table 20G about here >

These responses indicate that officers felt to a fairly strong extent that they were able to explain their perspectives without feeling judged.

Item 8: To what extent did this meeting with this supervisor make you feel you were being unduly criticized?

< Table 20H about here >

In reviewing these responses, it is unclear as to whether all officers read the question accurately. Because this item was reverse coded (higher score is a more negative response), and the fact that seven respondents had anomalous responses, this was the only question where a higher score corresponded to a more negative response, meaning that the actual mean may have been slightly lower). In any event, the mean of 1.42 and median of 1 suggests that most officers did not feel at all unduly criticized by the supervisor(s) in these meetings.

Item 9: To what extent did this meeting with this supervisor help you to understand ways in which you could improve future encounters?

< Table 20I about here >

These responses indicate that for the most part, officers only felt that the meetings helped them to understand ways to improve their behavior to a small degree. This could also be explained by the fact that a number of individuals felt that the incidents selected were of low priority, stress, and importance, and/or that individual debriefings are not as helpful as squad or larger group debriefings in terms of individual performance (see below).

Item 10: Please provide comments about this meeting or suggested improvements for future supervisory meetings.

There were 32 total comments provided in this section. Many (15) of the responses had to do with the selection of incidents for debriefing in these sessions. All of these respondents felt that the selection of mundane incidents of low criticality made the exercise less effective/ineffective. In reviewing the content of all 32 responses, we were able to code each response into at least one or sometimes two categories, resulting in 34 total responses across categories.

Responses. A description of each category, the total similar responses, and in some cases [an] example(s) are provided below:

1. Incidents too mundane, low stress, or not relevant (n = 15)

"…better to choose a significant incident"

"…more productive if used for a 'high risk' type of incident"

"[events were] nothing calls….mundane"

"incidents selected should not be done at random"; narrow to more dynamic situations

2. Good/great idea or tool (n = 7)

3. Need entire squad/all officers on scene (n = 4)

"…should also debrief as a group"

4. Current supervisors already do this (n = 3)

"[our precinct commanders] know our character and trust our judgment…"

5. Not beneficial (n = 2)

"…colossal waste of time…"

6. Confidentiality concerns (n = 2)

"…being cornered in a room … does not lead to authentic answers"

7. Neutral (n = 1)

Comparison of West precinct to other identified precincts. At the outset, we noted that among the precincts identified by officers, there was a higher rate of return for those from the West Precinct. While it would be highly speculative to indicate the reasons for this, it nevertheless raised our concern that perhaps the West precinct differed in some way from the others (perhaps higher degree of trust in supervisors, being encouraged/pushed to respond to a greater degree, having a better organizational climate, etc.).

As such, we decided to analyze the means and standard deviations of the item responses from the West Precinct as compared to those from the other identified precincts (North, South, East, and Southwest). We present those results in Table 21 below:

< *Table 21 about here* >

The results indicate that in general, those in the West Precinct had more favorable views of the intervention, the supervisory approach, and its benefits except for item 4 where the responses were almost identical (West 3.71 vs. other 3.73) and where the West had a slightly higher standard deviation, indicating that both groups felt similarly about the justness of the supervisory session. For item 8 where the coding was reversed, those from the West Precinct had only a slightly lower score and virtually the same standard deviation, indicating that neither group felt more unduly criticized than the other precincts. Again, while we are not able to provide an explanation for these differences, they were nevertheless present in this sample.

Conclusions

General Findings

Our results suggest that short, non-punitive, supervisory meetings can lead to officers resolving more incidents in non-punitive ways. Importantly, we find that this general de-escalation of citizen encounters is not associated with a reduction in officer activity. Indeed, we failed to find evidence that officers responded to fewer events, or initiated fewer citizen encounters. We also failed to find evidence that officers responded to these meetings by limiting the amount of time they spend with citizens. Instead, officers appear to check in and out of incidents with roughly the same frequency. Furthermore, we failed to find evidence that officers are less likely to file written reports about incidents after meetings.

What we do observe is a reduction in the frequency with which officers are involved in serious, potentially problematic, encounters with citizens. Specifically, officers who are reminded to think about their own thought process and how they incorporate new information *while* talking to the citizen rather than beforehand are less likely to decide on resolving an incident via arrest. These officers are also less likely to be involved in an incident where physical force is needed to control a situation.

Conditional on officers engaging the public and continuing to deter crime through their physical presence (Nagin, 2013), it is difficult to argue why a reduction in force or arrests is not a desirable outcome. To the extent that the marginal arrest affected by training could have been resolved via citation or warning, rather than a clearly criminal incident that warranted formal charges, and because citizens always prefer not to be arrested, these supervisory meetings should promote public perceptions of police integrity in the long run. In addition, the officers who participated in the engagements generally perceived them to be, at worst, relatively innocuous meetings and, at best, a chance to interact with their supervisors in a way that made them feel that their experiences and perceptions were valued in the department

Limitations of the Study

To the best of our knowledge, this study is one of the first to examine the impact of procedural justice training on officer behavior in the field, rather than through citizen surveys. Measuring officer integrity through citizen surveys has the advantage of directly tracking how people view the police. However, there are very real costs associated with relying on surveys to evaluate police training programs. Most importantly, surveys that are representative of the population of an area will not be representative of people who actually interact with the police. Attempts to evaluate the impact of a police training program by quantifying a contemporaneous

change in the perceptions of people who did not interact with the police will inevitably find that the program had no impact.

A better approach would be to survey people who interacted with police officers before and after the training program. This method was used in Mazerolle, et al. (2012), one of the few studies that did find substantive effects on police perceptions after traffic stops. Surveying a representative sample of all people who have interacted with the police, rather than a sample of people who were arrested, is critical for the internal validity of the results; our results suggest that the marginal arrestee is likely to be more criminal or culpable, and therefore likely to be more dissatisfied with the police, after procedural justice training.

Implications For Further Research, Policy, & Practice

There are several positive contributions of this evaluation to the understanding of policing integrity. First, we developed an EIS system that draws from recent criminological insights about behavioral hot spots, and there is some evidence that officers working in areas of "marginal" risk – the lowest risk of a high risk group- were more likely to benefit from these additional meetings. Second, we demonstrated that even short supervisory meetings that take place at relatively low frequencies can promote the use of procedural justice in the field in a practical and efficient manner. Third, we measured the effect of procedural justice training on policing using quantifiable field outcomes - officer activity and incident outcomes - rather than inherently subjective self-reports by the officers. Finally, we hope that this project will foster further partnerships between researchers and police professionals. Beyond simple feedback, members of the Seattle Police Department collaborated with the Co-PIs in the development of many components of this proposal, and contributed a significant part of the human capital necessary for the project to succeed.

This research has tremendous potential to add knowledge to our efforts to promote police integrity by increasing legitimacy in police citizen interactions. There are also multiple potential extensions of this research that could benefit other departments. Our HRC model is based on sound criminological theory and sophisticated statistical modeling; to the best of our knowledge, we are one of the first to use behavioral hot spots as a key factor in identifying police who face situations where negative encounters are likely to arise.

Our experimental engagement was not only appropriate to this early identification of officers, but it promoted procedural justice in two ways. Sergeants encouraged officers to use LEED concepts during citizen encounters, and provided positive feedback to those already doing so during the engagements. Second, in a dimension that we believe is unique to our study, sergeants modeled procedural justice for the officers by using LEED during the intervention under the supervision of lieutenants who were trained in procedural justice. Like our HRC model, this too was motivated by recent advances in police research, which argue that people are more receptive to authority when procedural justice is used (Hickman, Piquero, Lawton, & Greene, 2001).

A final important component of our study was that we experimentally evaluated the impact of LEED techniques on tractable and policy relevant outcomes. In practice, procedural justice initiatives are often met with skepticism by the public and are considered to be little more than "political correctness run amok."[11] We evaluated the impact of LEED using outcomes that, unlike self-reported "satisfaction with the police," are easily quantified in a cost-benefit analysis, and which are more frequent, and therefore easier to plausibly measure than officer shootings or use of excessive force.

[11] http://www.foxnews.com/us/2011/06/22/cop-killings-up-new-training-stresses-kindness/

Although randomized control trials produce unbiased estimates of treatment effects, one limitation is that replication of these experimental results in other cities is necessary before strong policy decisions are made. However, we have demonstrated proof of the concept that it is indeed possible for police department leadership to make relatively minor changes to supervisory strategies, and have real, measurable effects on the frequency with which police officers resort to potentially problematic tactics. Additional research can either confirm that these results hold in other cities and in other settings, or find that the Seattle context was unique.

REFERENCES

Alpert, G. P., & Walker, S. (2000). Police accountability and early warning systems: Developing policies and programs. *Justice Research and Policy*, *2*(2), 59-72.

Alpert, G., Rojek, J., & Noble, J. (2012). *The cognitive interview in policing: Negotiating control.* (Briefing paper). Queensland: Australian Research Council.

Amendola, K. L. (2006). *Overview of early warning and intervention systems: Focus on U.S. Department of Justice investigations.* Presentation at Early Warning System Conference, Phoenix, Arizona, Sponsored by Preferred Training Solutions in cooperation with the Phoenix Police Department. Retrieved from http://www.policefoundation.org/content/early-identification-and-intervention-systems-slide-show

Amendola, K. L. (2003). *Overview of early warning and intervention systems: Focus on U.S. Department of Justice investigations.* Washington, DC: Police Foundation.

Braga, A., & Weisburd, D. (2010). *Policing problem places: Crime hot spots and effective prevention*. New York City, NY: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Eck, J. & Weisburd, D. (1995). Crime places in crime theory. In J. E. Eck & D. Weisburd (Eds.), *Crime and place: Crime prevention studies, 4,* (pp. 1-33). Monsey, NY: Willow Tree Press.

Engel, R. S. (2005). Citizens' perceptions of procedural and distributive injustice during traffic stops with police. *Journal of Research in Crime and Delinquency 42*(4), 445-481.

Fisher, R. P. & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing*. Springfield, IL: Charles C. Thomas Publisher.

Fisher, R. P. & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. *International Journal of Law and Psychiatry, 33*(5-6), 321-328.

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of the actual victims and witnesses of crime. *Journal of Applied Psychology*, *74*(5), 722.

Gau, J. M. & Brunson, R. K. (2010). Procedural justice and order maintenance policing: A study of inner-city young men's perceptions of police legitimacy. *Justice Quarterly*, *27*(2), 252-279.

Hickman, M., Piquero, A., Lawton, B., & Greene, J. (2001). Applying Tittle's control balance theory to police deviance. *Policing: An International Journal of Police Strategies &*

Management, 24(4), 497-520.

Hilbe, J. (2009). *Logistic regression models*. Boca Raton, FL: Chapman and Hall.

Hinds, L., & Murphy, K. (2007). Public satisfaction with police: Using procedural justice to improve police legitimacy. *Australian & New Zealand Journal of Criminology, 40*(1), 27-42.

Kaplan, Z., Iancu, I., & Bodner, E. (2001). A review of psychological debriefing after extreme stress. *Psychiatric Services, 52*(6), 824-827.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490−517.

Lipsey., M. W.(1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D.J. Rog (Eds.), *Handbook of Applied Social Research Methods* (pp. 39-68). Thousand Oaks, CA: Sage.

Maguire, M., & Norris, C. (1994). Police investigations: Practice and malpractice. *Journal of Law and Society, 21*(1), 72-84.

Mazerolle, L., Bennett, S., Eggins, E., Antrobus, E., White, G., & Davis, J. (2011). *Testing police legitimacy one breath at a time: The Queensland community engagement trial.* (Technical report). Queensland: Australian Research Council.

Mazerolle, L., Bennett, S., Antrobus, E., & Eggins, E. (2012). Procedural justice, routine encounters and citizen perceptions of police: Main findings from the Queensland Community Engagement Trial (QCET). *Journal of experimental criminology, 8*(4), 343-367.

Melekian, B. (2012). *Values-based discipline: The key to organizational transformation in law enforcement agencies* (Doctoral dissertation, University of Southern California). Retrieved from http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll3/id/61802

Nagin, D. S. (2013). Deterrence: A review of the evidence by a criminologist for economists. *The Annual Review of Economics, 5*(1), 83-105.

Nagler, J. (1994). Scobit: An alternative estimator to logit and probit. *American Journal of Political Science, 38*(1), 230-255.

North, D. (1981). *Structure and change in economic history.* New York City, NY: Norton and Company.

Ortiz, C. W. & Amendola, K. L. (2005). Early warning systems. In L. E. Sullivan & M. S. Rosen (Eds.), *Encyclopedia of law enforcement* (pp. 162-164). Thousand Oaks, CA: Sage Publications.

Rahr, S., Diaz, J., & Hawe, J. (2011). *The four pillars of justice based policing: Listen and*

explain with equity and dignity. (Brief). Retrieved from http://www.portlandonline.com/auditor/index.cfm?a=368336&c=56523

Reisig, M. D. (2007). Continuing the discussion on community policing, procedural justice and community policing—What shapes residents' willingness to participate in crime prevention programs? *Policing, 1*(3), 356-369.

Reisig, M. D., Bratton, J., & Gertz, M. G. (2007). The construct validity and refinement of process-based policing measures. *Criminal Justice & Behavior, 34*(7), 1005-1028.

Ridgeway, G. & MacDonald, J. (2009). Methods for assessing racially biased policing. In S. K. Rice & M. D.White & (Eds.), *Race, ethnicity and policing: The issues, methods, research and future* (pp. 180-204). New York, NY: New York University press.

Rosenbaum, D. P., Schuck, A.M., Costello, S. K., Hawkins, D. F., & Ring, M. K. (2005). Attitudes toward the police: The effects of direct and vicarious experience. *Police Quarterly, 8*(3), 343-365.

Rosenbaum, D. P. & Lawrence, D. (2011). *Teaching respectful police-citizen encounters and good decision making: Results of a randomized control trial with police recruits.* Unpublished manuscript, National Police Research Platform. Retrieved from http://www.nationalpoliceresearch.org/the-chicago-quality-interactio/

Schulhofer, S. J., Tyler, T. R., & Huq, A. Z. (2011). American policing at a crossroads: Unsustainable policies and the procedural justice alternative. *The Journal of Criminal Law and Criminology, 101*(2), 335-374.

Sherman, L. (1998). American policing. In M. Tonry (Ed.), *The 1998 handbook of crime and punishment* (pp. 429 – 456). New York, NY: Oxford University Press.

Sherman, L. & Weisburd, D. (1995). General deterrent effects of police patrol in crime 'hot spots': A randomized study. *Justice Quarterly, 12*(4), 625-648.

Sherman, L., Gartin, P., & Buerger, M. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27-55.

Sunshine, J. & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review, 37*(3): 513-547.

Tyler, T. R. (1990). *Why people obey the law: Procedural justice, legitimacy, and compliance*. New Haven, CT: Yale University Press.

Tyler, T. R. (1988). What is procedural justice?: Criteria used by citizens to assess the fairness of legal procedures. *Law and Society Review, 22*(1), 103-135.

Tyler, T. R. (1998). Trust and democratic governance. In V. Braithwaite & M. Levi (Eds.), *Trust and governance* (pp. 269–294). New York: Russell Sage.

Tyler, T. R. (2001). Public trust and confidence in legal authorities: What do majority and minority group members want from the law and legal institutions? *Behavioral Sciences and the Law, 19*(2), 215 - 235.

Tyler, T. R. (2003). Procedural justice, legitimacy, and the effective rule of law. *Crime and Justice: A Review of Research*, 30, 431–505.

Tyler, T. R. (2005). *Readings in procedural justice*. Burlington, VT: Ashgate.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio State Journal of Criminal Law, 6*(1), 231-276.

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts.* New York City, NY: Russell-Sage.

Tyler, T. R. & Wakslak, C. J. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority. *Criminology, 42*(2), 253-282.

United States Department of Justice, Civil Rights Division. (2011). *Investigation of the Seattle Police Department.* Retrieved from http://www.justice.gov/crt/about/spl/seattlepd.php

United States Department of Justice, Bureau of Justice Statistics. (2003). *Law Enforcement Management and Administrative Statistics (LEMAS): Sample survey of law enforcement agencies* [Data file and codebook]. Retrieved from http://www.bjs.gov/index.cfm?ty=dcdetail&iid=248

United States Department of Justice, Bureau of Justice Statistics. (2007). *Law Enforcement Management and Administrative Statistics (LEMAS)* [Data file and codebook]. Retrieved from http://www.bjs.gov/index.cfm?ty=dcdetail&iid=248

Walker, S. (2000). Early warning systems for police: Concept, history, and issues. *Police Quarterly, 3*(2), 132-152.

Walker, S. (2001). Searching for the denominator: Problems with police traffic stop data and an early warning system solution. *Justice Research and Policy, 3*(1), 63-95.

Walker, S. (2003). *Early intervention systems for law enforcement agencies: A planning and management guide*. Washington, D.C.: U.S. Dept. of Justice, Office of Community Oriented Policing Services.

Walker, S., & Alpert, G. (2004). Early intervention systems: The new paradigm. In M. Hickman, A. Piquero & J. Greene (Eds.) *Police integrity and ethics* (pp. 221 – 235). Belmont, CA: Wadsworth Group.

Walker, S., Alpert, G., & Kenney, D. (2001). *Early warning systems: Responding to the problem police officer* (National Institute of Justice Research in Brief). Washington D.C.: U.S. Department of Justice.

Walker, S., & Graham, N. (1998). Citizen complaints in response to police misconduct: the results of a victimization survey. *Police Quarterly*, *1*(1), 65-89.

Walsh, M. (2011). *Use of force by Seattle police officers, 2010.* Seattle, WA: Seattle Police Department.

Weisburd, D. (2008). *Policing places* (Ideas in American Policing Monograph). Washington D.C.: Police Foundation.

APPENDICES

Appendix A: Training SPD Sergeants To Conduct Experimental Engagement

Background on the Training

Our supervisor training program, designed strictly for the purpose of overseeing the experimental intervention, was modeled off of the training programs laid out in Mazerolle et al. (2011), Rosenbaum and Lawrence (2011) and Fisher and Geiselman (1992). Mazerolle et al. (2011) and Rosenbaum and Lawrence (2011), hereafter referred to as MAZ and RL, evaluated the impact of procedural justice on citizen perceptions of fairness and respect for police by randomly assigning officers to use LEED concepts in their civilian interactions, either during random breathalyzer tests (MAZ) or during general police work (RL). There were some differences in the outcomes of these studies, with MAZ finding significant improvements in citizens' attitudes towards police, but RL finding that while some officers in the experimental group appeared to be more likely to use LEED concepts in the field, there was little effect on officers' beliefs about procedural justice. These disparate findings are potentially explained by differences in the project implementation. Initially, the MAZ research team held a group meeting with the officers working in the jurisdiction of interest, in which they explained the purpose of their study, the basic ideas behind procedural justice, and how they expected these ideas to be implemented in practice. Officers were given the chance to ask the MAZ research team questions about the experiment, and their participation, directly.

During the MAZ experiment, officers in the LEED group were provided with a procedural justice cue card that they kept with them in the field. This cue card not only reminded officers of important phrases such as "*We are pulling cars over today at random. You were not specifically singled out for this test*," but also reminded officers to begin the encounter with neutral, non-judgmental greetings, and to end the citizen encounter by giving positive feedback

to the driver. Members of the MAZ research team also rode along with officers in both the treatment and control groups, in order to confirm that the experiment was being implemented properly.

RL, on the other hand, added twenty additional hours of classroom training in procedural justice techniques to the standard academy training of police recruits. The training began with a formal lecture on procedural justice from one officer and one university professor, and the remaining hours of training were more interactive. Trainees in the treated group watched pre-recorded or live re-enactments of police-citizen interactions where police did or did not use LEED concepts. They then participated in role-playing exercises and provided feedback on the role-playing of others.

The most obvious difference in the design of the MAZ and RL studies is the timing of the LEED training. While officers in the RL treatment almost certainly spent more time learning about procedural justice concepts, this training ended once they left the academy. MAZ officers, in contrast, were continually reminded of LEED concepts, by having both the cue card with them, and having the researchers present during their stops.

On the one hand, the MAZ experiment has a higher level of internal validity; for example, the research team was able to confirm that treated and control officers behaved as intended. On the other hand, the steps taken to increase internal validity arguably detracted from the external validity of MAZ relative to RL. In practice, officers will not necessarily carry a cue card around with them all the time, and will not have their behavior monitored by outside researchers. The RL treatment, which simply added classroom time to standard training, is almost certainly a better approximation of a policy that would actually be implemented in a police department.

Training Lieutenants to Model LEED for Sergeants

With the experiences of MAZ and RL in mind, our experimental implementation was designed to combine a relatively short period of classroom training for lieutenants, who then instructed sergeants under their command on how to model LEED techniques. With an eye towards external validity, lieutenants were told that research team members were not to be present during the engagements. Even as observers, the presence of the researchers alone would make the LEED engagement different from one that would be implemented in practice.

The presence of the lieutenant, however, ensured the validity of the project implementation, and also reinforced the institutionalization of the procedural justice training. In practice, the sustained involvement of lieutenants in an officer training program is less likely, but we believed that temporary lieutenant supervision could reasonably be part of the implementation of such a methodology in another police department.

Cognitive Interview Model

The LEED debrief training was developed after in-depth discussions about how using cognitive interviews could assist in the development of honest, complete and worthwhile information. Essentially, information received from officers, suspects, and witnesses is used to reconstruct a scene, situation or encounter and to form a description of what is known about a specific situation. A critical goal of the cognitive interview is to get each witness to provide accurate information, but there is no standard method or "best practice" to achieve this outcome (Maguire and Norris, 1994). There are many ways investigators attempt to elicit information and often, it is a "gut feeling" rather than a proven strategy based on evidence that is used by them to interact with individuals. The cognitive interview process provides the interviewee with an opportunity to recall facts more effectively and efficiently.

Thus, interviewers need to learn how to listen, ask questions and prompt subjects without being biased, leading or prejudiced. In this type of interview, it is the job of the interviewer to encourage the participant to do most of the mental work as he/she has the information sought by the interviewer. It is the interviewer's job to help the participant provide the information in a complete and unbiased manner. For example, memory concerning a threat should be more easily obtained when the participant is thinking about when he or she first perceived it, rather than when he or she is talking about the person posing the threat.

The Experimental Oversight Training Session

Using the tenets of cognitive interviewing, a deputy chief conducted a series of trial interviews with a small group of senior lieutenants, as well as some sergeants and officers. The train-the-trainer workbook and other informational materials were developed, tested, and modified in two follow-up rounds of mock interviews. The deputy chief met and trained all the participating lieutenants, who in turn trained the sergeants. Additionally, the lieutenants were tasked with maintaining the quality of the interviews conducted by the sergeants. The deputy chief was available to assist in maintaining quality control throughout the experimental period.

The program to train and implement the experimental engagement model ultimately involved the entire department, which was a rather momentous and important undertaking in and of itself. Among the significant transformational objectives of the project undertaken was involving the entirety of the officers in the Seattle Police Department – numbering about 1300 – in learning the principles of LEED, and the protocols of the experimental engagement.

The deputy chief and Co-PI Dr. Geoffrey Alpert conducted the initial supervisor training session for the twenty SPD lieutenants assigned to oversee the experimental intervention engagement sessions. The deputy chief introduced Dr. Alpert, and explained to the selected

supervisors that the SPD and Police Foundation were jointly conducting groundbreaking research on supervisory techniques, that might help both the department and the policing community at large. The deputy chief further emphasized that, while no employee was required to participate in this experiment, the SPD not only fully supported the research but was also an active participant in the development of the experiment. The lieutenants were then tasked with explaining and demonstrating these principles for the 3-4 sergeants under their command.

The process began with the introduction of the model by the Chief of Police and command staff to the department as a whole, and in-depth with department captains and lieutenants. The lieutenant rank was assigned to manage the program, and train sergeants in its structure and protocols. Sergeants then served as the interviewers and interlocutors, with their respective lieutenants in attendance as observers. To accomplish this, lieutenants and captains were asked to become familiar with the interview model (described below) through meetings and role-playing with the command staff, which included assistant and deputy chiefs, and were assisted by LEED instructors from the SPD Training Section. This helped to lay the foundation for a "train-the-trainer" approach. The lieutenants then imparted the interview model to their assigned sergeants, who were participated as interlocutors with officers randomly selected for experimental engagements.

This top down approach was utilized in order to ensure buy-in at every command and supervisory level of the SPD, and was efficient in its span and scope. The experimental engagements were focused on patrol officers, who were commanded by five captains and twenty lieutenants. The command staff of seven assistant or deputy chiefs was responsible for conducting the familiarization discussion with these 25 commanders, which roughly constituted a 1:4 instructional ratio. The 20 lieutenants, in turn, were responsible for training the

approximately seventy-two sergeants in Patrol, the Anti-Crime Team, Community Police Team and Footbeat/Bicycle Squads, again, roughly constituting a 1:4 instructional ratio. The patrol force itself, at the time, was comprised of both 911 and proactive officers distributed among 5 precincts, numbering approximately 680 sworn officers.

Training and implementation steps based upon the experimental engagement model. As stated above, the training/implementation process *was* the model and vice versa. The steps towards completing this crucial process were as follows:

A. The initial process of familiarization and project objectives proceeded from the Chief to all employees;

B. The Command Staff (Assistant and Deputy Chiefs) assisted by LEED specialists from the SPD Training Section, followed the process described above to familiarize and practice the model with their Captains and Lieutenants; and

C. The Lieutenants, following this "train the trainer" format, implemented the overall familiarization program with all precinct Sergeants.

These conversations were projected to take no more than two hours, and included opportunities for role-playing based upon the criteria of random assignment of officers to an EE. For example, there are abundant use-of-force reports, which could have been employed to illustrate the engagement model when random assignment is triggered by a TASER application or other trigger, for example.

Appendix B: The LEED Engagement Script

Training Module for the Experimental Engagement (EE) Model

The methodology followed "adult learning" principles: candid, respectful conversation; comprehensive review of the topics and learning points, with comprehension acknowledged at each point; and scenario based role-playing to examine and validate the model.

The "training" then consisted of modeling the engagement model itself. Briefly, the elements of the 6- part engagement – or conversation - were as follows:

I. Introduction and Overview of the LEED Engagement

II. Statement of Purpose: The supervisor explained that the officer was randomly selected for the engagement, and what the engagement consisted of. At the outset, the officer was reassured that the forthcoming engagement was not going to result in discipline, that the objectives included an examination from the officer's perspective, and that respectful, candid facilitation was to govern the discussion. The sergeant was encouraged to use the processes and devices of conversational clarity – paraphrasing, echoing, presenting or responding to hypotheticals, and above all, to ask clarifying questions at any point of the engagement. The officer was informed that LEED principles would be introduced throughout his/her recapitulation of the incident in question, and moreover, that the sergeant was bound to employ those same LEED principles during the engagement itself. It was then disclosed that the observance of these overarching principles was one reason why the lieutenant was to be present.

III. Initial Narrative: The officer was to be asked to "walkthrough" the incident, which was chosen jointly by the sergeant and officer, with the aim being that it involved some interaction with a citizen but was not necessarily an incident that would otherwise be discussed in a supervisory setting. There could have been the option, dependent upon practical opportunity,

to physically return to the site of the incident itself (although this was not done). This initial narrative was to be in the officer's own words, and should have been relatively unimpeded. The sergeant was directed at this point to facilitate the full airing of the incident by the officer in his/her own words but could ask clarifying questions. The sergeant's principle task at this juncture was to identify and organize the follow-on interview, which is the crux of the engagement, as described below. Once the "walkthrough" was completed to the likes of both the officer and the sergeant, the actual engagement could commence.

IV. Cognitive Debrief: Following the walkthrough, the sergeant was tasked with facilitating a discussion of the officer's thought process, emotional state, assessment of threats and opportunities, and other factors in the officer's view which underpinned the decisions and outcomes of the incident. This first interplay between sergeant and officer was aimed at mutually arriving at a comprehensive understanding of the reason the incident resolved as it did, and set up the process to benchmark the elements of the incident against LEED principles.

V. LEED Debrief: Following the cognitive interview/debrief was the re-examination of the incident actions, decision points and factors in light of LEED principles. Specifically, any point in the interaction which may have allowed opportunities to provide explanation, or establish affirmations of respect and fairness, or simply involved listening, was to be discussed with a view towards answering a three prong test:

1) Were these opportunities exploited and successful? What contributed to the success?

2) Were these opportunities impossible to exploit because of other factors? What were those impediments or precluding factors?

3) Were there missed opportunities? What were the causal factors?

The sergeant was to then move through the engagement towards a seminal moment; that is, the determination of whether there were unexploited options to further the LEED principles of *Listening* and *Explaining* governed by the principles of *Equity* and *Dignity*.

VI. Reconciliation and Summary: At this point, the sergeant was to recede from the facilitator role exercised during the cognitive and LEED debriefs. The officer was to be encouraged to provide a summary assessment of the entire engagement, with specific attention paid to any lessons learned. The officer was also to be given an unimpeded opportunity to assess the quality of the interview according to LEED principles; namely, to respectfully provide impressions of the sergeant's adherence to LEED in his/her interaction with the officer. Again, the presence of the lieutenant, who previously acted as the sergeant's trainer and who was responsible to his/her captain and chief for management of the project, was to ensure that this reconciliation phase was governed by LEED principles.

Overview of the LEED Engagement

At the beginning of the engagement, the sergeant was asked to follow standard introductory instructions, adapted from Rosenbaum and Lawrence (2011) and Fisher and Geiselman (1992). These consisted of:

(1) Thanking the officer for coming in to help you test this new supervisory training program: reminding them that they were selected randomly, participant identity and comments are confidential and non-attributive, and that this meeting is entirely meant to help examine new supervisory techniques.

(2) Transferring control to the officer: telling the officer that you would like to discuss an interaction that they had on (date of flagged encounter) at around (time of flagged encounter), as reported in G.O. # ____. Letting the officer choose the starting point for

the narrative and give the account at his or her own speed and in his or her own words, not interrupting the officer, if at all possible. Listening actively to what he or she has to say. Allowing for pauses.

[The following script was provided to sergeants to help guide them through the debrief process].

Debrief Model
Sample Script

Introduction script: Welcome officer and confirm baseline understandings of process and protections.

- Incident # _____ has been randomly selected for a debrief of the circumstances of the incident as perceived and reported by [you] the officer.
- In addition to being a random assignment, officer identity is anonymous and confidential. The department has no access to the identity of the participants in the debrief, and the involved officer will be coded by a code assigned and retained by Cornell university, which again will not be accessible to the department.
- No notes will be taken by the sergeant and lieutenant, and any subsequent documentation of training and equipment needs, tactics, best practices and other valuable insights intended to reduce officer risk in difficult encounters will not include personal identifying information. The watch/operations commander is charged with ensuring that confidentiality and officer anonymity is maintained at all points of this debriefing process.

At this point, the sergeant should inquire if the officer is aware of any complaint or investigation into the incident being examined, and terminate the debrief if the answer is yes.

- The debrief will take no more than 30 minutes, unless by mutual agreement between [you] the officer and [myself] the sergeant.
- Do you have any questions up to this point?

Sergeants received training on how to answer questions such as: 1) why are we doing it? 2) why all the protections? 3) why is the Lieutenant here? 4) what is LEED? And other questions as they arose regarding confidentiality and anonymity, what constituted a high-risk circumstance, why there was a random and control group, what the union jurisdiction was in this process etc.

- If there are not any questions [or any more questions] I will use some key questions to facilitate your analysis of G.O. #_____.

Series one script: Basic incident information, risk assessment and tactics

Officer orientation: *Describing the incident and thought process in the real time framework of the incident as it unfolded*

- How did the incident begin? What were your initial thoughts based on the information you had, or what you initially observed, about the risk factors of this scenario?
- What did you observe about the behavior, mental or emotional state of the subject you encountered? Were there other risk factors observed (drugs, alcohol)?
- Were there external or environmental factors you included in your risk assessment and preliminary tactical thought process? These could include anything which you factored into your size up of the incident [such as] weather conditions, lighting, traffic, presence of hostile (or supportive) bystanders, availability of back-up, confidence in the

capabilities of your secondary officers or partner, prior adverse publicity about the location, scenario or involved parties, etc.

- Describe your tactical approach. What was your thought process around how to address the scenario, and the behavior of the subject?
- Did your tactical approach work to address the risks you identified?
- Did your initial risk assessment change? In what way? Were you required to change tactics in response to a change in behavior or risk? What were those changes?
- [Did anything] occur during the encounter, which you would describe as surprising and unpredictable? [If so, what?] How did you respond to these unforeseen events?
- Are there other factors, indicators, risks or other information you consider important in your recollection of the incident?
- Please share your professional assessment of how this incident [was] resolved.

Series two: Post incident assessment and examination of options and needs

Officer orientation: *analysis of the incident in retrospect, with an emphasis on identifying ways to mitigate risk and improve department training and support*

- In hindsight, do you believe you had the information you needed to size up the situation, identify the risk factors and employ the tactics you needed to address the situation? What, if anything, was missing?
- Do you consider the outcome of the incident to [have been] successful?
- How do you personally and professionally measure success?
- In hindsight, were there things you might have done differently? To put [it] another way, were there options to resolve the incident, which might have been effective?
- Were there options you wish you had? What are they?

- There is a lot of talk in the department around Procedural Justice, otherwise known as "LEED". What do you know about it? How would you describe it? Based on what you know, how do you assess its utility, relevance or general value?
- How would you size up the options or opportunities to employ LEED tactics in the incident we are discussing?
- Using as your reference point the incident we are debriefing, what can or should the department do to improve your ability to carry out your duties? This is an open question [for example] training, staffing, deployment, equipment, supervision, command, communication, accountability etc. Nothing is off limits.

Note to sergeant: it might be wise to set a time parameter for this question. Remember the 30 minute debrief objective, although this can be exceeded by mutual agreement

- In your professional judgment, what are the greatest risks you confront in your role as patrol officer? And what are the most effective and significant ways the department can reduce those risks?

Series three: Debriefing the debrief

Officer orientation: *candid and constructively critical.*

- What questions were missed?
- What information do I [the sergeant] need to know to better meet my role and responsibility to supervise?
- Did this debrief meet your standards and expectations for respectfulness? Relevance? Value?
- How would you change this process? Should this debriefing model be continued? Expanded?

- Any other feedback is welcome.

[END OF SCRIPT]

Appendix C: The HRC Model

In the development of the High Risk Circumstance (HRC) model we applied insights from criminology and statistics to develop a new kind of Early Intervention System (EIS) focused on identifying a larger pool of officers who were statistically more likely to engage in potentially problematic encounters in the future based on individual and contextual factors rather than on identifying specific officers who had engaged in problematic encounters in the past.

Specifically, the HRC model identifies officers working in behavioral "hot spots," or in other words, small geographic areas where police officers are more likely to be involved in problematic citizen encounters such as assaults on police officers, intoxicated persons, etc.

The HRC model developed for this experiment was calibrated using geographically identified, incident-level data and human resource records from 2009 collected by the Seattle Police Department (SPD). The first step in developing the HRC model was to define exactly what types of behaviors we wanted officers to avoid. These behaviors were the dependent variables used to calibrate the HRC statistical model. In the case of the SPD, our primary definition of such behaviors consisted of incidents in which an officer filed a use of force report, incidents in which an officer was named in a citizen complaint, and incidents in which the officer was injured.

The model was used to assign a "risk score" to every incident recorded in the SPD's Computer Aided Dispatch (CAD) system, based on a 30 day history of events in close geographic and temporal (time of day and day of week) proximity to the incident in question.

In order to construct the predicted risk score based on these location and officer-specific elements, we used data on all CAD incidents from 2010 to predict the likelihood that a potentially problematic event occurred. A potentially problematic event was defined as an

incident in which an officer used force, an officer was injured, or a complaint was in the process of being filed or was filed against an SPD officer who was working in that area on that day. In collaboration with SPD employees, we identified the following data fields, which were used to predict risk and uncertainty associated with an encounter. All of these data fields are collected by the SPD as part of their normal operations. The final data fields included in the HRC model were:

1. The number of times officers had been dispatched to that area
2. The number of 911 calls initiated from that area
3. The number of events in that area involving someone in mental distress
4. The number of times events in that area have involved citizens with firearms
5. The number of times officers have responded to domestic violence incidents in that area
6. The number of times dispatchers have chosen to include officer safety warnings for officers deployed to that area
7. The fraction of events that are correctly described by the dispatcher to the officer
8. The fraction of events that are described by the officer as more urgent, or a higher priority call, than initially described by the dispatcher
9. The number of officers injured on that street segment during the previous month
10. Whether or not the involved officer had more than the average (median) number of sick days in the previous month
11. Whether or not the involved officer worked more than the average (median) amount of overtime in the previous month
12. Whether or not the involved officer had more than the average (median) number of second jobs during the previous month

Data fields 1 through 8 were defined at the census block and street segment level and at the day of the week and time of day (morning, afternoon, evening, and late night) period, in the 30 days prior to the current incident.

Appendix D: SPD COMMENT CARD

As part of a research project being conducted by the Police Foundation in Washington, DC, to examine various forms of feedback and supervisory training, we hope you will agree to participate in this study by completing and returning this comment card regarding this supervisory meeting. Your participation in this research is anonymous, and will only be used for these research purposes. None of your responses will be connected to you and are being requested without your name or any identifying information.

Please CIRCLE the numerical response that bests represents your reaction to the meeting you had with a Seattle Police Department Supervisor today in the most candid fashion. Please note that this survey is being read and tallied only by the PF research team.

	N/A	**1**	**2**	**3**	**4**	**5**
	Not at all	to a small degree		to a modest degree		to a great degree

To what extent did this meeting with this supervisor…

1.	help to stimulate your memory of the event(s)?	N/A	1	2	3	4	5
2.	allow you to recall specific details of the event(s)?	N/A	1	2	3	4	5
3.	feel like a fair and just system for providing feedback to officers?	N/A	1	2	3	4	5
4.	make you feel more respected than in past supervisory meetings?	N/A	1	2	3	4	5
5.	make you feel you were being unduly criticized?	N/A	1	2	3	4	5
6.	help you to understand ways in which you could improve future encounters?	N/A	1	2	3	4	5
7.	allow you the chance to explain your perspective without feeling judged?	N/A	1	2	3	4	5
8.	make you feel like what you had to say actually mattered to this supervisor?	N/A	1	2	3	4	5
9.	make you wish more supervisors in the SPD used this type of feedback approach?	N/A	1	2	3	4	5

10. Please provide any comments about this meeting or suggested improvements for future supervisory meetings:

Appendix E: Procedural Justice Engagement Monitoring Checklist

Purpose:

The purpose of this checklist is to ensure that the procedural justice experiment is being properly implemented during the engagement. The purpose of the engagement is to remind officers how they can use LEED concepts on the job, while simultaneously modeling procedural justice for them in practice.

Date:_____

Supervisor Name: _____

Beginning of Engagement

Start time: _____ End time: _____

Was the **officer** on time?

- ☐ Yes
- ☐ No

Was the **supervisor** on time?

- ☐ Yes
- ☐ No

At the beginning of the engagement, the sergeant will follow standard introductory instructions, adapted from established cognitive interview techniques. Please make note of the behaviors listed below that took place during your observation of the engagement.

A. Standard Introductory Instructions
   - ☐ Lieutenant was present to observe engagement between sergeant and selected officer.
   - ☐ Purpose of engagement was explained to officer (e.g. officer is helping test new supervisory training program).
   - ☐ Officer was reminded that he/she was selected at random.
   - ☐ Officer was ensured of the confidentiality of the meeting.
   - ☐ Officer was given opportunity to ask questions and he/she received satisfactory answers to his/her questions.
   - ☐ Officer was asked if he/she would be willing to participate.

B. Transfer of Control

- ☐ Flagged incident to be discussed identified to officer with minimal detail (only date and time of flagged encounter detailed).
- ☐ Officer was allowed to take the lead in explaining the identified incident without interruption.

Engagement (Incident Debrief)

Start time: _____ End time: _____

C. Incident walkthrough

- ☐ The officer was allowed to recall information about the identified incident on his/her own terms without pressure to speculate about details or observations about which they were unsure or confused.
- ☐ The officer explained what happened, what he/she observed or believed, and how he/she responded.
- ☐ After giving his/her account, the sergeant conducting the interview walked through the incident with the participant exploring whether there were opportunities to undertake OR impediments to introducing the actions of LISTENING, EXPLAINING, EQUITY and DIGNITY (LEED).
- ☐ The sergeant utilized the following questions to help the officer facilitate his/her walkthrough of the identified G.O.
  - ☐ How did the incident begin?
  - ☐ What were your initial thoughts based on the information you had, or what you initially observed about the risk factors of this scenario?
  - ☐ What did you observe about the behavior, mental or emotional state of the subject you encountered?
  - ☐ Were there other risk factors observed (drugs, alcohol)?
  - ☐ Were there external or environmental factors you included in your risk assessment and preliminary tactical though process? These could include anything which you factored into your size up of the incident: weather conditions, lighting, traffic, presence of hostile (or supportive) bystanders, availability of back-up, confidence in the capabilities of your secondary

officers or partner, prior adverse publicity about the location, scenario or involved parties, etc.

- ☐ Describe your tactical approach. What was your thought process around how to address the scenario, and the behavior of the subject?
- ☐ Did your tactical approach work to address the risks you identified?
- ☐ Did your initial risk assessment change? In what way?
- ☐ Were you required to change tactics in response to a change in behavior or risk? What were those changes?
- ☐ What occurred during the encounter, which you would describe as surprising and unpredictable?
- ☐ How did you respond to these unforeseen events?
- ☐ Are there other factors, indicators, risks or other information you consider important in your recollection of the incident?
- ☐ Please share your professional assessment of how this incident resolved.

D. Post incident assessment and examination of options and needs

- ☐ Officer was given the opportunity to analyze the incident in retrospect with an emphasis on identifying ways to mitigate risk and improve department training and support.
- ☐ The sergeant utilized the following questions to help the officer facilitate his/her analysis of the identified G.O.
  - ☐ In hindsight, do you believe you had the information you needed to size up the situation, identify the risk factors and employ the tactics you needed to address the situation?
  - ☐ What, if anything, was missing?
  - ☐ Do you consider the outcome of the incident to be successful
  - ☐ How do you personally and professionally measure success?
  - ☐ In hindsight, were there things you might have done differently? To put in another way, were there options to resolve the incident, which might have been effective?
  - ☐ Were there options you wish you had? What are they?

- ☐ There is a lot of talk in the department around Procedural Justice, otherwise known as "LEED."
- ☐ What do you know about it?
- ☐ How would you describe it?
- ☐ Based on what you know, how do you assess its utility, relevance or general value?
- ☐ How would you size up the options or opportunities to employ LEED tactics in the incident we are discussing?
- ☐ Using as your reference point the incident we are debriefing, what can or should the department do to improve your ability to carry out your duties? This is an open question: training, staffing, deployment, equipment, supervision, command, communication, accountability: nothing is off limits. *(Note to sergeant: it might be wise to set a time parameter for this question. Remember the 30 minute debrief objective, although this can be exceeded by mutual agreement)*
- ☐ In your professional judgment, what are the greatest risks you confront in your role as patrol officer?
- ☐ What are the most effective and significant ways the department can reduce those risks?

Post-Engagement

Start time: _____ End time: _____

E. Officer feedback and comment card

- ☐ The officer was given the opportunity to debrief the debrief and comment on the following:
  - ☐ What questions were missed?
  - ☐ What information does (the sergeant) need to know to better meet my role and responsibility to supervise?
  - ☐ Did this debrief meet your standards and expectations for respectfulness? Relevance? Value?

- ☐ How would you change this process? Should this debriefing model be continued? Expanded?
- ☐ Any other feedback is welcome.

☐ The officer was reaffirmed that all ideas, observations, proposals, and identified gaps and needs will be presented to department leadership without attribution or identifiers and that the lieutenant present was charged with ensuring that anonymity and confidentiality were preserved.

☐ Following the engagement the officer was asked to fill out a short confidential comment card on the engagement experience. The officer was advised that filling out the card was completely voluntary and that he/she could decline to answer any or all parts of the survey card.

☐ The officer was given the blue comment card cover sheet with the Police Foundation contact information on it to keep.

Body Language

Did supervisor cross his/her arms during the meeting (appearing closed off)?

Did the supervisor have sufficient eye contact with officer?

Did the officer cross his/her arms during the meeting (appearing defensive)?

Did the officer look away during much of the engagement?

Please make note of both the sergeant's and the officer's body language.

Additional Notes and Overall Observations

Did the supervisor pick out an incident that was of sufficient length for discussion?

Did the supervisor appear to take this seriously, and spend sufficient time in discussing the event? In other words, did the supervisor appear attentive?

Did the officer appear to take this seriously in terms of his/her responses?

Did the engagement appear non-disciplinary in nature?

Who appeared to be talking the most?

- ☐ Officer
- ☐ Supervisor
- ☐ About the same

FIGURES 1 - 4

Figure 1: Distribution of Time from Notification to Engagement

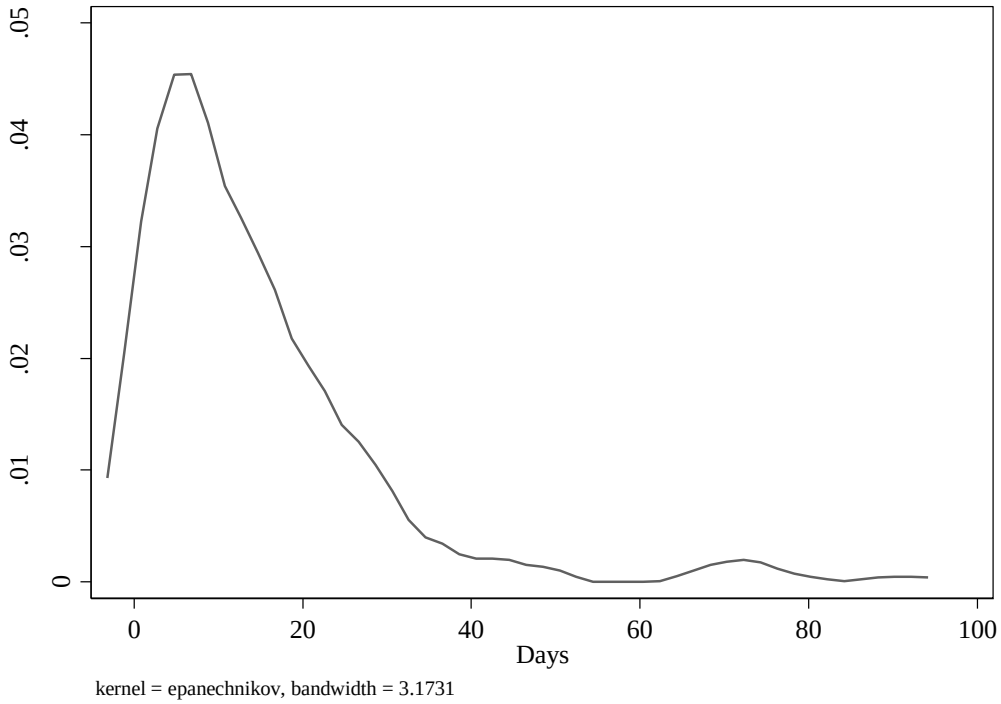
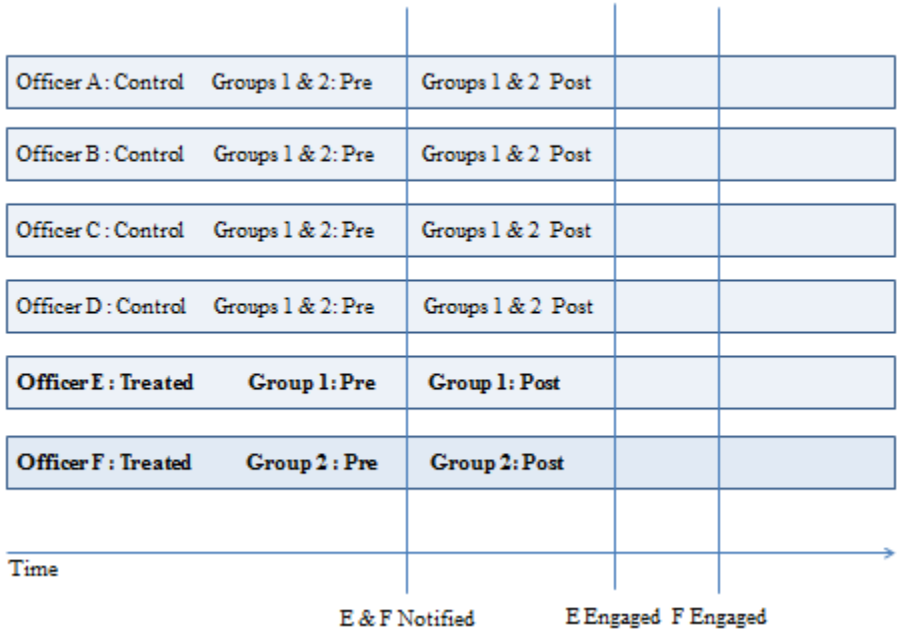


Figure 2: Construction of Experimental Groups, two treated officers and four control officers

Panel A: Time Period of Analysis: Before and After Notification



Panel B: Time Period of Analysis: Before Notification and After Engagement

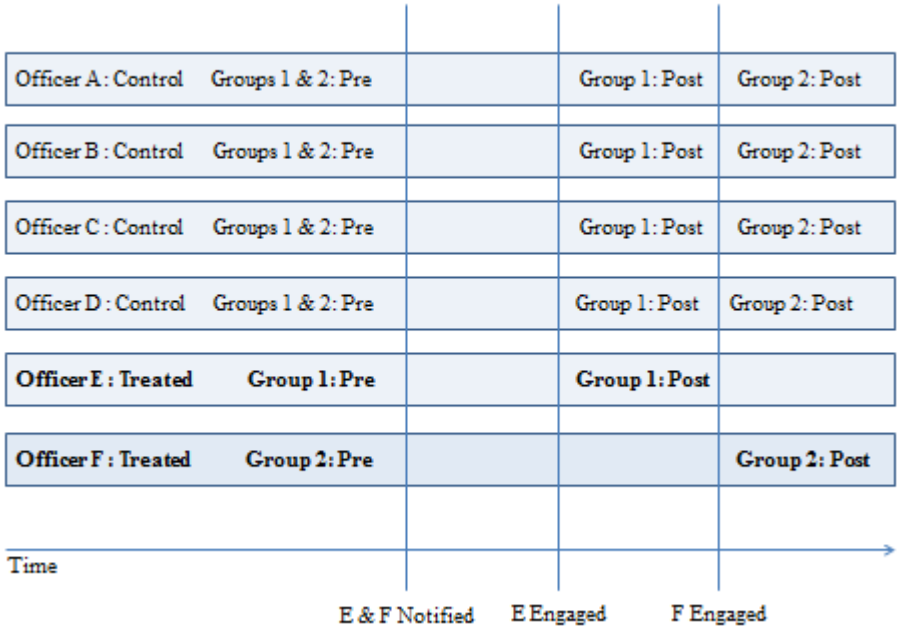


Figure 3: Probability Distribution of Force Incidents Prior to Notification

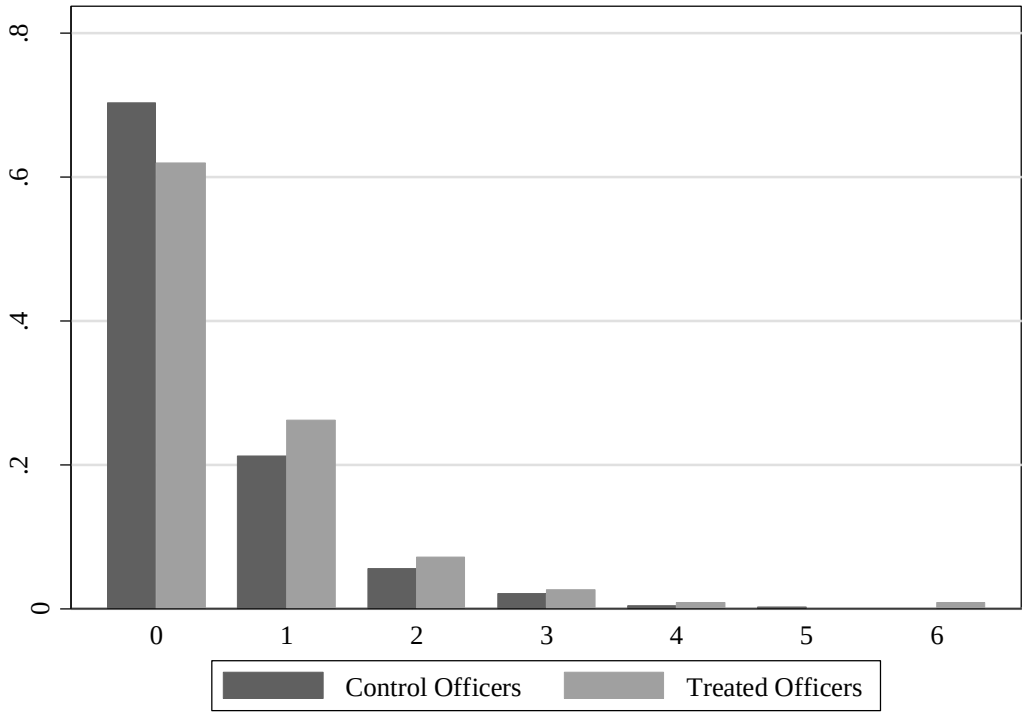
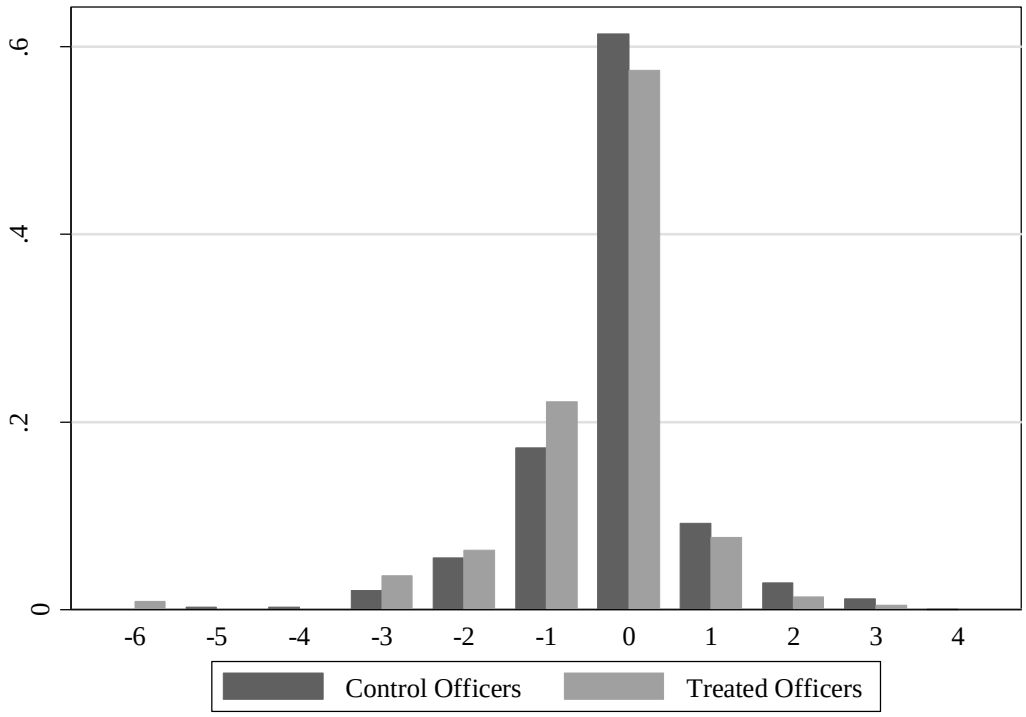


Figure 4: Probability Distribution of Change in Force Indents



TABLES 1 – 21

Table 1: Predicted Risk Scores and Actual Potentially Problematic Risk Rate in February and March, 2012 by Precinct and Sector

	Predicted Risk Score (x 1000)	Potentially Problematic Event Rate (x 1000)
North Precinct		
U	2.6155	3.4891
N	4.5568	3.0173
J	2.4954	1.4194
L	2.9134	0.9362
B	2.7077	0.6152
West Precinct		
K	3.1575	3.1177
Q	3.6729	2.6023
M	2.7027	2.0481
D	2.8757	1.718
East Precinct		
E	2.9366	3.8633
G	2.7828	1.1669
C	3.4295	0.948
South Precinct		
R	2.9115	2.1272
S	2.6894	1.8658
O	2.8435	1.7979
Southwest Precinct		
F	3.9204	1.4911
W	2.5005	0.5351

Table 2: HRC Model Logit Parameters

Parameter	Unit of Measurement	Logit Coefficient Weight
Fraction of Incidents assigned higher than typical priority codes	Census Block	-0.1670919
	Census Block, Day of the Week	-0.9552692
	Census Block, Time of Day	-0.1653143
	Census Block, Time of Day, Day of the Week	1.337767
	Street Segment	0.5420524
	Street Segment, Day of the Week	0.2930098
	Street Segment, Day of the Week, Time of Day	-0.2259336
	Street Segment, Time of Day	-0.4919433
Fraction of Incidents cleared as dispatched	Census Block	0.1067927
	Census Block, Day of the Week	0.339997
	Census Block, Time of Day	-0.2717843
	Census Block, Time of Day, Day of the Week	0.1583683
	Street Segment	0.2965052
	Street Segment, Day of the Week	-0.205414
	Street Segment, Day of the Week, Time of Day	-0.0021559
	Street Segment, Time of Day	-0.0608896
Fraction of Incidents cleared as more serious than as dispatched	Census Block	1.543565
	Census Block, Day of the Week	-3.413816
	Census Block, Time of Day	-2.016165
	Census Block, Time of Day, Day of the Week	4.614575
	Street Segment	-8.011918
	Street Segment, Day of the Week	4.912824
	Street Segment, Time of Day	1.467594
More than Average Overtime Hours	Officer	0.1796542
More than Average Second Jobs	Officer	0.0729175
More than Average Sick Days	Officer	-0.0505735
Previous Month's Injuries	Street Segment	3.690103
Number of 911 Calls	Census Block	-0.0155581
	Census Block, Day of the Week	0.022542
	Census Block, Day of the Week, Time of Day	-0.0477802
	Census Block, Time of Day	0.0557603
	Street Segment	-0.0320639
	Street Segment, Day of the Week	0.1481227
	Street Segment, Day of the Week, Time of Day	-0.2614739
	Street Segment, Time of Day	0.0030107

Number of CAD incidents

	Census Block	0.0100086
	Census Block, Day of the Week	-0.0625784
	Census Block, Day of the Week, Time of Day	-0.0036729
	Census Block, Time of Day	-0.0094977
	Street Segment	-0.0180991
	Street Segment, Day of the Week	0.1070731
	Street Segment, Day of the Week, Time of Day	0.0717182
	Street Segment, Time of Day	0.0028498
Number of Domestic Violence Incidents	Census Block	0.0285296
	Census Block, Day of the Week	0.7974532
	Census Block, Day of the Week, Time of Day	0.2102308
	Census Block, Time of Day	-0.5183722
	Street Segment	-0.0467877
	Street Segment, Day of the Week	-0.3350682
	Street Segment, Day of the Week, Time of Day	-1.035518
	Street Segment, Time of Day	0.6696225
Number of Firearm/Shots Fired Incidents	Census Block	-0.3283155
	Census Block, Day of the Week	-0.870897
	Census Block, Day of the Week, Time of Day	1.85194
	Census Block, Time of Day	0.7677631
	Street Segment	0.256366
	Street Segment, Day of the Week	1.078482
	Street Segment, Time of Day	-2.075535
Number of Incidents with Officer Safety Notes	Census Block	0.3484783
	Census Block, Day of the Week	0.2016767
	Census Block, Day of the Week, Time of Day	-1.879869
	Census Block, Time of Day	-0.5537603
	Street Segment	0.4296225
	Street Segment, Day of the Week	-2.067785
	Street Segment, Day of the Week, Time of Day	2.523242
	Street Segment, Time of Day	-0.0063393
Number of Mental Distress Incidents	Census Block	-0.1377365
	Census Block, Day of the Week	0.6080221
	Census Block, Day of the Week, Time of Day	-0.5015358
	Census Block, Time of Day	-0.0921126
	Street Segment	0.1380854
	Street Segment, Day of the Week	-0.6368511
	Street Segment, Day of the Week, Time of Day	0.4387145
	Street Segment, Time of Day	0.0920619

Table 3: Linear Probability Model Estimates of Predicted Risk

Location History (Past 30 Days)		Change in Predicted Risk Score x 1000	Probability Effect is Zero	*Estimated Total Effect of Events in Same Area, Same Time of Day Same Day of the Week x 1000*
Number of CAD incidents	Census Block	0.0214	50.40%	
	Census Block, Day of the Week	-0.1562	25.80%	
	Census Block, Day of the Week, Time of Day	0.00184	97.80%	
	Census Block, Time of Day	-0.0776	75.60%	*-0.21056*
	Street Segment	-0.0612	17.20%	
	Street Segment, Day of the Week	0.3766	5.40%	
	Street Segment, Day of the Week, Time of Day	0.0128	89.20%	
	Street Segment, Time of Day	0.1224	71.60%	*0.24*
Number of 911 Calls	Census Block	-0.0735	31.70%	
	Census Block, Day of the Week	0.1747	52.40%	
	Census Block, Day of the Week, Time of Day	0.1749	28.10%	
	Census Block, Time of Day	-0.2375	62.80%	*0.0386*
	Street Segment	-0.0532	58.50%	
	Street Segment, Day of the Week	0.2108	56.70%	
	Street Segment, Day of the Week, Time of Day	-0.033	87.50%	
	Street Segment, Time of Day	-0.4611	47.00%	*-0.3*
Number of Mental Distress Incidents	Census Block	-0.397	0.10%	
	Census Block, Day of the Week	1.8451	0.00%	
	Census Block, Day of the Week, Time of Day	-0.1309	60.10%	
	Census Block, Time of Day	-1.4595	4.20%	*0.1423*
	Street Segment	0.4896	0.50%	
	Street Segment, Day of the Week	-2.1107	0.00%	
	Street Segment, Day of the Week, Time of Day	0.1051	75.90%	
	Street Segment, Time of Day	1.4813	11.80%	*-0.18*
Number of Firearm/Shots Fired Incidents				

	Census Block	-0.7767	42.80%	
	Census Block, Day of the Week	-2.9475	22.90%	
	Census Block, Day of the Week, Time of Day	1.9895	27.20%	
	Census Block, Time of Day	10.3493	2.00%	*8.6146*
	Street Segment	1.1857	35.00%	
	Street Segment, Day of the Week	2.236	47.20%	
	Street Segment, Day of the Week, Time of Day	-5.148	2.70%	
	Street Segment, Time of Day	-7.9138	16.20%	*-1.03*
Number of Domestic Violence Incidents	Census Block	0.128	63.50%	
	Census Block, Day of the Week	5.4019	0.00%	
	Census Block, Day of the Week, Time of Day	-1.7444	0.20%	
	Census Block, Time of Day	-2.4638	7.50%	*1.3217*
	Street Segment	-0.223	53.10%	
	Street Segment, Day of the Week	-3.0044	0.20%	
	Street Segment, Day of the Week, Time of Day	2.4672	0.00%	
	Street Segment, Time of Day	-1.44	40.50%	*-0.88*
Number of Incidents with Officer Safety Notes	Census Block	1.7738	0.90%	
	Census Block, Day of the Week	-0.114	94.80%	
	Census Block, Day of the Week, Time of Day	-2.1445	8.30%	
	Census Block, Time of Day	-7.4123	2.20%	*-7.897*
	Street Segment	1.3156	14.20%	
	Street Segment, Day of the Week	-5.2203	1.90%	
	Street Segment, Day of the Week, Time of Day	-0.468	77.30%	
	Street Segment, Time of Day	10.288	1.30%	*-1.98*

Fraction of Incidents assigned higher than typical priority codes	Census Block	-0.1961	85.70%	
	Census Block, Day of the Week	-2.3205	2.70%	
	Census Block, Day of the Week, Time of Day	-0.5943	54.80%	
	Census Block, Time of Day	4.1159	0.10%	*1.005*
	Street Segment	1.4927	12.80%	
	Street Segment, Day of the Week	0.4331	70.20%	
	Street Segment, Day of the Week, Time of Day	-1.1806	24.60%	
	Street Segment, Time of Day	-0.5127	72.40%	*1.238*
Fraction of Incidents cleared as dispatched	Census Block	0.2897	58.70%	
	Census Block, Day of the Week	0.8864	5.60%	
	Census Block, Day of the Week, Time of Day	-0.61	18.70%	
	Census Block, Time of Day	0.6726	21.90%	*1.2387*
	Street Segment	0.6943	14.20%	
	Street Segment, Day of the Week	-0.5107	31.20%	
	Street Segment, Day of the Week, Time of Day	-0.2208	64.30%	
	Street Segment, Time of Day	-0.2445	70.70%	*0.957*
Fraction of Incidents cleared as more serious than as dispatched	Census Block	2.5998	75.30%	
	Census Block, Day of the Week	-3.7729	59.50%	
	Census Block, Day of the Week, Time of Day	-12.5011	9.60%	
	Census Block, Time of Day	21.9534	0.80%	*8.2792*
	Street Segment	-6.331	33.40%	
	Street Segment, Day of the Week	16.6936	2.40%	
	Street Segment, Day of the Week, Time of Day	0.2292	97.50%	
	Street Segment, Time of Day	-21.7095	1.70%	*-2.84*
HR Data				
More than Average Second Jobs	Officer	0.2075	74.80%	
More than Average Sick Days	Officer	-0.0928	86.70%	
More than Average Overtime Hours	Officer	0.5884	35.00%	
Previous Month's Injuries	Street Segment	117.17	0.00%	

Table 4: Pre-Notification Officer Activity and Incident Outcomes

	One Week Prior				Six Weeks Prior			
	Treated	Control	*Probability of*		Treated	Control	*Probability of*	
	N=221	*N=1213*	*treatment and control mean drawn from the same population*	*correctly identifying a 10% treatment effect*	*N=221*	*N=1213*	*treatment and control mean drawn from the same population*	*correctly identifying a 10% treatment effect*
Activity Measures								
Incidents	43.33 (18.48)	40.32 (22.90)	*0.232*	*0.89*	210.73 (87.81)	205.87 (90.91)	*0.801*	*0.94*
Report Taken	0.299 (0.128)	0.283 (0.158)	*0.188*	*0.89*	0.292 (0.100)	0.287 (0.104)	*0.378*	*0.99*
On Views	0.318 (0.198)	0.325 (0.223)	*0.378*	*0.71*	0.314 (0.173)	0.319 (0.189)	*0.404*	*0.80*
Minutes On Scene	39.61 (15.85)	39.74 (23.57)	*0.974*	*0.93*	37.80 (12.69)	39.69 (17.94)	*0.208*	*0.99*
Incident Outcomes								
Arrests	0.070 (0.071)	0.060 (0.077)	*0.391*	*0.31*	0.065 (0.053)	0.062 (0.052)	*0.818*	*0.48*
Force Reports (total)	0.018 (0.134)	0.018 (0.134)	*0.538*	*0.10*	0.213 (0.481)	0.149 (0.450)	*0.276*	*0.13*
Number of Complaints (total)	0.00 (0.00)	0.004 (0.064)	*0.307*	*0.11**	0.054 (0.280)	0.054 (0.227)	*0.923*	*0.11*
Citations	0.239 (0.117)	0.224 (0.131)	*0.083*	*0.82*	0.236 (0.086)	0.233 (0.090)	*0.354*	*0.98*
Warnings	0.051 (0.061)	0.052 (0.080)	*0.732*	*0.30*	0.050 (0.047)	0.049 (0.052)	*0.860*	*0.40*
Assistance Rendered	0.489 (0.139)	0.474 (0.181)	*0.364*	*1.0*	0.488 (0.113)	0.485 (0.105)	*0.730*	*1.0*

Standard Deviations in Parentheses.

* a standard error of 0.001 is assumed for the treatment group

Table 5: Number of CAD Events Involving Engaged Officers (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	2.846 [1.819]	2.149 [1.920]	0.72 [6.611]	0.935 [6.084]
P(Effect = 0)	0.119	0.264	0.913	0.878
Pre-Notification Control Mean	40.32		205.87	
R^2	0.161	0.175	0.199	0.273

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 6: Percent of CAD Events Initiated by Officer ("On-Views") (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	0.0268+ [0.0139]	-0.0107 [0.0129]	-0.00589 [0.00768]	-0.00931 [0.00917]
P(Effect = 0)	0.0545	0.407	0.444	0.311
Pre-Notification Control Mean	0.325		0.319	
R^2	0.218	0.203	0.277	0.314

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 7: Average Estimated Time on Scene, in Minutes (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	3.078 [2.752]	1.527 [2.115]	1.60+ [0.892]	0.962 [1.029]
P(Effect = 0)	0.264	0.723	0.074	0.350
Pre-Notification Control Mean	39.74		39.69	
R^2	0.172	0.219	0.171	0.228

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 8: Fraction of CAD Events with Written Report (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.00191 [0.0145]	-0.00417 [0.0137]	-0.00232 [0.00575]	-0.0036 [0.00609]
P(Effect = 0)	0.895	0.762	0.686	0.554
Pre-Notification Control Mean	0.283		0.287	
R^2	0.156	0.190	0.224	0.282

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 9: Fraction of CAD Events Resulting in Arrest (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.00112 [0.00645]	-0.0156* [0.00642]	-0.00572+ [0.00346]	-0.00749+ [0.00404]
P(Effect = 0)	0.862	0.0159	0.0995	0.0648
Pre-Notification Control Mean	0.060		0.062	
R^2	0.254	0.220	0.185	0.179

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 10: Use of Force Incidents (n=1,434)

	One Week		Six Weeks		Full Time Period	
	After Notification	After Engagement	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.0117	-0.0393	-0.0738	-0.062	-0.207*	-0.187*
	[0.0216]	[0.0435]	[0.0486]	[0.0489]	[0.0890]	[0.0904]
P(Effect = 0)	0.588	0.366	0.130	0.202	0.0207	0.0391
Pre-Notification Control Mean	0.0181		0.149		0.418	
R^2	0.213	0.325	0.169	0.348	0.366	0.368

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 11: Use of Force Incidents, Trimmed Sample (n=1,434)

	One Week		Six Weeks		Full Time Period	
	After Notification	After Engagement	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.00112 [0.00645]	-0.0393 [0.0435]	-0.0738 [0.0486]	-0.0625 [0.0489]	-0.173* [0.0795]	-0.158* [0.0784]
P(Effect = 0)	0.588	0.366	0.130	0.202	0.0302	0.0452
Pre-Notification Control Mean	0.0181		0.149		0.409	
R^2	0.213	0.325	0.166	0.184	0.368	0.367

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 12: Citizen Complaints (n=1,434)

	One Week		Six Weeks		Full Time Period	
	After Notification	After Engagement	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	0.003	0.0347	0.0122	0.031	0.0128	0.020
	[0.00923]	[0.0290]	[0.0315]	[0.0319]	[0.0483]	[0.0478]
P(Effect = 0)	0.746	0.231	0.700	0.337	0.791	0.676
Pre-Notification Control Mean	0.004		0.049		0.115	
R^2	0.328	0.216	0.218	0.197	0.25	0.25

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 13: Fraction of CAD Events Resulting in Non-Criminal Infractions (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.0068 [0.0133]	0.00344 [0.0129]	0.0056 [0.00573]	0.0062 [0.00617]
P(Effect = 0)	0.609	0.789	0.33	0.316
Pre-Notification Control Mean	0.224		0.233	
R^2	0.164	0.200	0.218	0.303

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 14: Fraction of CAD Events Resulting in Verbal Warnings (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	-0.00057 [0.0069]	0.00583 [0.00775]	-0.0027 [0.00287]	-0.000735 [0.00329]
P(Effect = 0)	0.935	0.453	0.347	0.823
Pre-Notification Control Mean	0.052		0.049	
R^2	0.136	0.153	0.133	0.154

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 15: Fraction of CAD Events Resulting in Assistance Rendered (n=1,434)

	One Week		Six Weeks	
	After Notification	After Engagement	After Notification	After Engagement
Effect of Supervisory Meeting	0.0229 [0.0151]	0.0295+ [0.0156]	0.00657 [0.00803]	-0.00186 [0.00804]
P(Effect = 0)	0.131	0.0594	0.414	0.817
Pre-Notification Control Mean	0.474		0.485	
R^2	0.183	0.203	0.208	0.354

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 16A: Effect Heterogeneity by Frequency of Meetings, One Week Outcomes (n=1,434)

	Panel A: Officer Activity							
	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	2.358	0.0249	4.131	-0.00482	2.872	-0.00142	2.98	-0.0113
	[2.313]	[0.0169]	[3.801]	[0.0171]	[2.235]	[0.0146]	[2.483]	[0.0147]
Supervisory Meeting x One Previous Notification	-1.655	0.0109	-6.657	0.0138	-4.678	0.0042	-10.08+	0.0476
	[6.770]	[0.0404]	[8.295]	[0.0434]	[5.906]	[0.0441]	[5.575]	[0.0363]
Supervisory Meeting x Two Previous Notifications	9.182	-0.0472	-1.424	0.034	-4.269	0.0224	3.072	0.0093
	[12.95]	[0.0743]	[7.510]	[0.0627]	[11.32]	[0.0643]	[8.438]	[0.0558]
R^2	0.169	0.221	0.175	0.157	0.199	0.209	0.227	0.195
	Panel B: Incident Outcomes							
	After Notification				After Engagement			
	Arrests	Force	Complaints		Arrests	Force	Complaints	
Effect of Supervisory Meeting	-0.0043	-0.00463	0.00729		-0.0238**	-0.0817	0.0278	
	[0.00767]	[0.0245]	[0.0110]		[0.00766]	[0.0540]	[0.0333]	
Supervisory Meeting x One Previous Notification	0.015	-0.00444	-0.0222		0.0367+	0.190*	0.06	
	[0.0231]	[0.0632]	[0.0168]		[0.0210]	[0.0743]	[0.0582]	
Supervisory Meeting x Two Previous Notifications	0.00131	-0.0642	-0.00842		0.0346	0.177*	-0.114	
	[0.0259]	[0.111]	[0.0135]		[0.0288]	[0.0887]	[0.186]	
R^2	0.264	0.249	0.330		0.223	0.328	0.220	

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects, and first order effects of previous notifications. Robust standard errors in outcomes within officer (320 clusters)
*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 16B: Effect Heterogeneity by Frequency of Meetings, Six Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	5.593	-0.0108	1.980+	-0.00301	4.706	-0.014	1.479	-0.00431
	[7.702]	[0.00864]	[1.100]	[0.00689]	[7.396]	[0.00989]	[1.208]	[0.00729]
Supervisory Meeting x One Previous Notification	-11.73	0.0204	-2.933	-0.00343	-4.406	0.0214	-2.6	-0.00656
	[22.85]	[0.0183]	[2.374]	[0.0153]	[22.65]	[0.0204]	[2.428]	[0.0165]
Supervisory Meeting x Two Previous Notifications	-45.06	0.0423	-0.0335	0.0269	-44.91	0.0542	-2.037	0.0385
	[48.94]	[0.0419]	[2.966]	[0.0248]	[45.08]	[0.0468]	[4.231]	[0.0293]
R^2	0.213	0.229	0.177	0.286	0.284	0.323	0.229	0.286

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.00599	-0.103	0.0158	-0.00646	-0.0994	0.0275
	[0.00399]	[0.0633]	[0.0345]	[0.00435]	[0.0630]	[0.0348]
Supervisory Meeting x One Previous Notification	-0.00574	0.343+	-0.0403	-0.0119	0.353+	-0.0249
	[0.00938]	[0.196]	[0.0644]	[0.0114]	[0.181]	[0.0689]
Supervisory Meeting x Two Previous Notifications	0.00666	-0.546+	-0.0922	0.00788	-0.533+	0.0081
	[0.0162]	[0.291]	[0.223]	[0.0190]	[0.288]	[0.293]
R^2	0.196	0.193	0.235	0.184	0.208	0.214

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects, and first order effects of previous notifications. Robust standard errors in outcomes within officer (320 clusters)
*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 17A: Effect Heterogeneity by Predicted Risk Circumstance, One Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	2.75 [1.812]	0.0269+ [0.0138]	3.227 [2.731]	-0.00172 [0.0144]	2.082 [1.907]	0.00357 [0.0124]	1.532 [2.109]	-0.00433 [0.0138]
Supervisory Meeting x Predicted Risk	-3.323* [1.684]	0.0184 [0.0198]	1.1 [2.667]	0.0206 [0.0142]	0.0137 [1.968]	-0.00755 [0.0142]	2.305 [1.596]	0.00868 [0.0117]
R^2	0.164	0.22	0.175	0.157	0.195	0.203	0.221	0.191

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.000747 [0.00643]	-0.0112 [0.0214]	0.00287 [0.00923]	-0.0153* [0.00639]	-0.0371 [0.0435]	0.0354 [0.0288]
Supervisory Meeting x Predicted Risk	0.0121 [0.00778]	0.0058 [0.0280]	-0.0102* [0.0051]	0.00841 [0.00775]	-0.014 [0.064]	0.046 [0.0290]
R^2	0.258	0.214	0.330	0.221	0.330	0.209

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects and first order effects of predicted risk. Robust standard errors in outcomes within officer (320 clusters)
*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 17B: Effect Heterogeneity by Predicted Risk Circumstance, Six Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	0.672	-0.00541	1.554+	-0.00264	1.046	-0.00889	0.877	-0.00376
	[6.575]	[0.00769]	[0.883]	[0.00572]	[6.046]	[0.00921]	[1.018]	[0.00608]
Supervisory Meeting x Predicted Risk	-1.09	0.00472	-0.127	0.00369	-0.228	0.00724	0.0184	0.00458
	[5.576]	[0.00825]	[1.248]	[0.00557]	[5.257]	[0.00833]	[1.287]	[0.00627]
R^2	0.199	0.281	0.174	0.231	0.273	0.316	0.235	0.284

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.00576+	-0.0701	0.0108	-0.00754+	-0.057	0.03
	[0.00345]	[0.0480]	[0.0316]	[0.00403]	[0.0486]	[0.0319]
Supervisory Meeting x Predicted Risk	0.00750*	-0.0558	0.0296	0.00859**	-0.0281	0.0529*
	[0.00293]	[0.0569]	[0.0270]	[0.00315]	[0.0777]	[0.0258]
R^2	0.192	0.181	0.230	0.184	0.208	0.172

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. Predicted risk scores are normalized to have a mean of zero and a standard deviation of 1. All regressions include wave by treatment group fixed effects and first order effects of predicted risk. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 18A: Effect Heterogeneity by Treatment Wave, One Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	11.02+	0.0237	0.728	-0.0504	13.42*	-0.0301	-4.639	-0.0414
	[6.012]	[0.0472]	[7.602]	[0.0486]	[5.459]	[0.0492]	[6.157]	[0.0428]
Supervisory Meeting x Wave 2	-11.82	-0.0475	-3.436	-0.00752	-15.32+	0.0294	-1.971	-0.015
	[9.320]	[0.0573]	[9.270]	[0.0588]	[7.793]	[0.0604]	[7.422]	[0.0552]
Supervisory Meeting x Wave 3	-2.567	0.0922	24.2	0.0401	-5.519	0.00076	11.74	-0.0335
	[8.686]	[0.0876]	[16.68]	[0.0770]	[7.479]	[0.0568]	[9.787]	[0.0589]
Supervisory Meeting x Wave 4	-11.47	0.0153	-11.25	-0.0243	-16.97*	0.0335	22.01+	-0.00194
	[7.767]	[0.0643]	[19.66]	[0.0850]	[8.446]	[0.0701]	[12.71]	[0.0706]
Supervisory Meeting x Wave 5	0.641	-0.0593	-2.3	0.0309	-8.293	-0.0029	8.274	0.0344
	[10.39]	[0.0628]	[12.66]	[0.0604]	[9.453]	[0.0705]	[11.31]	[0.0559]
Supervisory Meeting x Wave 6	-3.382	-0.0443	10.4	0.0521	-13.05	0.00184	1.072	0.0641
	[9.219]	[0.0597]	[12.98]	[0.0607]	[8.200]	[0.0598]	[7.445]	[0.0662]
Supervisory Meeting x Wave 7	-12.87	0.00417	9.96	0.114+	-11.02	-0.0211	5.57	0.118+
	[8.109]	[0.0656]	[9.209]	[0.0636]	[8.149]	[0.0755]	[10.60]	[0.0616]
Supervisory Meeting x Wave 8	-16.03	-0.0236	1.741	0.0647	-17.22	0.0515	5.331	0.059
	[10.80]	[0.0654]	[8.876]	[0.0670]	[11.12]	[0.0604]	[7.760]	[0.0601]
Supervisory Meeting x Wave 9	-17.00*	-0.0507	5.711	0.13	-24.27**	0.0438	5.061	0.0769
	[8.200]	[0.0681]	[9.011]	[0.0820]	[8.285]	[0.0758]	[9.670]	[0.0681]
Supervisory Meeting x Wave 10	-5.598	0.0236	-6.889	0.0524	-3.893	0.0903	2.503	0.044
	[7.982]	[0.0593]	[9.221]	[0.0649]	[7.961]	[0.0640]	[7.300]	[0.0590]
Supervisory Meeting x Wave 11	-13.61	-0.0195	1.33	0.0904	-10.97	0.0595	5.947	0.064
	[8.489]	[0.0688]	[8.781]	[0.0667]	[7.540]	[0.0719]	[7.249]	[0.0541]
Supervisory Meeting x Wave 12	-6.173	0.122+	-3.206	0.0461	-10.42	0.122*	6.862	0.0442
	[9.428]	[0.0725]	[12.53]	[0.0708]	[8.179]	[0.0596]	[7.822]	[0.0666]
R^2	0.170	0.231	0.182	0.163	0.202	0.209	0.225	0.197

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.0346+	-0.197*	0	-0.0325+	-0.440+	-0.159
	[0.0199]	[0.0917]	-	[0.0190]	[0.242]	[0.157]
Supervisory Meeting x Wave 2	0.0212	0.127	0	0.014	0.635*	0.101
	[0.0313]	[0.0967]	-	[0.0280]	[0.319]	[0.177]

Supervisory Meeting x Wave 3	0.0446	0.204+	0	-0.0189	0.173	0.181
	[0.0295]	[0.112]	-	[0.0315]	[0.306]	[0.178]
Supervisory Meeting x Wave 4	0.0797*	0.283*	0	-0.0033	0.267	0.197
	[0.0379]	[0.127]	-	[0.0317]	[0.266]	[0.190]
Supervisory Meeting x Wave 5	0.0122	0.197*	0	0.0218	0.448+	0.456*
	[0.0277]	[0.0917]	-	[0.0243]	[0.267]	[0.208]
Supervisory Meeting x Wave 6	0.0116	0.250+	0	0.0331	0.447	-0.0185
	[0.0287]	[0.145]	-	[0.0326]	[0.276]	[0.187]
Supervisory Meeting x Wave 7	0.0389	0.292+	0	0.0205	0.567*	0.144
	[0.0289]	[0.152]	-	[0.0335]	[0.269]	[0.167]
Supervisory Meeting x Wave 8	0.0520+	0.241*	0	0.0329	0.533*	0.410*
	[0.0294]	[0.118]	-	[0.0255]	[0.254]	[0.208]
Supervisory Meeting x Wave 9	0.0354	0.148	0	-0.00725	0.401	0.252
	[0.0290]	[0.107]	-	[0.0395]	[0.245]	[0.174]
Supervisory Meeting x Wave 10	0.0239	0.147	0	0.0640*	0.39	0.159
	[0.0310]	[0.102]	-	[0.0323]	[0.247]	[0.172]
Supervisory Meeting x Wave 11	0.0402	0.0951	0	0.0335	0.504+	0.275
	[0.0286]	[0.114]	-	[0.0256]	[0.260]	[0.184]
Supervisory Meeting x Wave 12	0.038	0.197*	0	0.0135	0.440+	0.159
	[0.0395]	[0.0917]	-	[0.0258]	[0.242]	[0.157]
R^2	0.263	0.225	0.360	0.226	0.337	0.240

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 18B: Effect Heterogeneity by Treatment Wave, Six Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	24.36	-0.00814	0.0495	-0.0331	38.66+	-0.0176	-0.58	-0.0366
	[19.81]	[0.0266]	[3.830]	[0.0256]	[21.82]	[0.0278]	[3.748]	[0.0282]
Supervisory Meeting x Wave 2	-39.33	0.0151	5.459	0.0490+	-53.09*	0.0308	5.299	0.0572+
	[26.05]	[0.0332]	[4.531]	[0.0296]	[26.51]	[0.0350]	[4.589]	[0.0323]
Supervisory Meeting x Wave 3	-8.615	-0.0127	0.269	0.0292	-19.45	0.00903	-0.684	0.0181
	[25.71]	[0.0358]	[4.653]	[0.0330]	[28.09]	[0.0367]	[4.481]	[0.0356]
Supervisory Meeting x Wave 4	-19.96	-0.00821	0.7	0.00451	-47.56	-0.00197	2.179	0.0484
	[34.43]	[0.0361]	[6.199]	[0.0319]	[34.91]	[0.0374]	[6.401]	[0.0328]
Supervisory Meeting x Wave 5	-22.15	-0.0116	2.561	0.0312	-44.13	-0.00439	3.92	0.0483
	[34.46]	[0.0359]	[4.574]	[0.0310]	[37.29]	[0.0389]	[4.392]	[0.0333]
Supervisory Meeting x Wave 6	-36.87	-0.0132	0.61	0.0471	-49.69	-0.00566	0.61	0.0456
	[29.37]	[0.0340]	[4.246]	[0.0317]	[31.59]	[0.0353]	[4.603]	[0.0387]
Supervisory Meeting x Wave 7	16.81	0.00485	-0.638	0.0295	6.346	-0.0251	-1.49	0.031
	[27.95]	[0.0437]	[5.980]	[0.0308]	[29.19]	[0.0534]	[6.101]	[0.0330]
Supervisory Meeting x Wave 8	-27.02	-0.0106	0.753	0.0311	-39.23	0.00356	-0.0399	0.021
	[35.41]	[0.0329]	[4.093]	[0.0295]	[34.79]	[0.0363]	[4.119]	[0.0337]
Supervisory Meeting x Wave 9	-66.37+	-0.0245	2.78	0.0334	-78.68*	0.0157	4.592	0.0264
	[34.23]	[0.0383]	[5.294]	[0.0344]	[34.02]	[0.0492]	[7.596]	[0.0417]
Supervisory Meeting x Wave 10	-18.57	0.0377	2.624	0.0415	-35.17	0.0285	2.434	0.0432
	[28.75]	[0.0343]	[4.443]	[0.0326]	[30.51]	[0.0373]	[4.347]	[0.0350]
Supervisory Meeting x Wave 11	-26.81	0.044	2.362	0.0302	-40.99	0.0606	2.51	0.0305
	[30.69]	[0.0385]	[4.309]	[0.0305]	[33.30]	[0.0413]	[4.208]	[0.0333]
Supervisory Meeting x Wave 12	-42.11	0.00817	1.464	0.0398	-56.99+	-0.00226	-0.0849	0.0242
	[29.28]	[0.0444]	[5.786]	[0.0306]	[31.34]	[0.0433]	[4.494]	[0.0350]
R^2	0.208	0.283	0.174	0.229	0.284	0.318	0.231	0.287

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.0217	-0.381+	-0.0238	-0.0296+	-0.238	-0.0119
	[0.0141]	[0.217]	[0.0276]	[0.0176]	[0.199]	[0.0158]
Supervisory Meeting x Wave 2	0.0315*	0.690+	-0.133*	0.0369+	0.702+	-0.0626
	[0.0153]	[0.365]	[0.0672]	[0.0189]	[0.376]	[0.0665]

Supervisory Meeting x Wave 3	0.0145	0.0906	0.0455	0.0222	-0.0483	0.0419
	[0.0165]	[0.275]	[0.0570]	[0.0198]	[0.297]	[0.0481]
Supervisory Meeting x Wave 4	0.0157	0.332	0.0563	0.0396*	0.114	0.0335
	[0.0168]	[0.299]	[0.0874]	[0.0196]	[0.253]	[0.0807]
Supervisory Meeting x Wave 5	0.011	0.0911	0.0194	0.029	-0.0128	0.078
	[0.0158]	[0.296]	[0.190]	[0.0189]	[0.257]	[0.194]
Supervisory Meeting x Wave 6	0.0147	0.562	0.0238	0.0207	0.355	0.0119
	[0.0197]	[0.373]	[0.0513]	[0.0214]	[0.301]	[0.0460]
Supervisory Meeting x Wave 7	0.00478	0.516	0.0392	0.00737	0.268	0.0799
	[0.0191]	[0.319]	[0.0637]	[0.0225]	[0.274]	[0.0760]
Supervisory Meeting x Wave 8	0.0191	0.341	0.183	0.0151	0.141	0.132
	[0.0167]	[0.244]	[0.120]	[0.0212]	[0.226]	[0.130]
Supervisory Meeting x Wave 9	0.0147	0.187	-0.00535	0.0275	0.093	-0.0173
	[0.0282]	[0.251]	[0.115]	[0.0368]	[0.229]	[0.113]
Supervisory Meeting x Wave 10	0.0295+	0.185	0.0212	0.0358	0.0424	-0.00519
	[0.0168]	[0.240]	[0.108]	[0.0222]	[0.228]	[0.103]
Supervisory Meeting x Wave 11	0.0221	0.32	0.0524	0.0261	0.257	0.105
	[0.0186]	[0.248]	[0.136]	[0.0220]	[0.229]	[0.133]
Supervisory Meeting x Wave 12	0.0142	0.371	0.117	0.00445	0.228	0.105
	[0.0207]	[0.239]	[0.0736]	[0.0227]	[0.222]	[0.0699]
R^2	0.190	0.180	0.226	0.185	0.200	0.202

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 19A: Effect Heterogeneity by Precinct, One Week Outcomes (n=1,434)

	Panel A: Officer Activity							
	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	0.077	0.0638	20.53*	-0.007	-1.457	0.0339	1.188	-0.0292
	[4.829]	[0.0400]	[9.926]	[0.0309]	[5.374]	[0.0289]	[4.984]	[0.0399]
Supervisory Meeting x North Precinct	3.436	-0.066	-24.65*	-0.0221	6.556	-0.0393	-0.155	-0.0135
	[5.741]	[0.0454]	[11.59]	[0.0457]	[6.041]	[0.0386]	[6.422]	[0.0504]
Supervisory Meeting x South Precinct	-5.468	-0.0352	-8.357	-0.0018	-4.428	-0.0112	6.224	0.0585
	[6.340]	[0.0580]	[11.45]	[0.0476]	[6.974]	[0.0437]	[7.824]	[0.0518]
Supervisory Meeting x Southwest Precinct	6.023	-0.0768	-19.94+	0.0572	2.929	-0.0029	-3.503	0.0847
	[6.200]	[0.0482]	[11.65]	[0.0534]	[6.363]	[0.0473]	[7.767]	[0.0584]
Supervisory Meeting x West Precinct	6.421	-0.0155	-22.44*	0.0203	6.609	-0.0531	-0.712	0.0381
	[6.126]	[0.0484]	[10.47]	[0.0374]	[6.861]	[0.0358]	[6.231]	[0.0438]
R^2	0.165	0.221	0.182	0.158	0.198	0.205	0.221	0.194

	Panel B: Incident Outcomes					
	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.00669	-0.0814	-0.0564+	-0.0248	-0.165	0.0457
	[0.0141]	[0.0495]	[0.0339]	[0.0187]	[0.109]	[0.0593]
Supervisory Meeting x North Precinct	0.00497	0.0752	0.0564+	0.00463	0.114	-0.0573
	[0.0201]	[0.0610]	[0.0339]	[0.0233]	[0.123]	[0.0753]
Supervisory Meeting x South Precinct	0.0052	0.0844	0.0564+	0.0061	0.0621	0.0353
	[0.0204]	[0.0586]	[0.0339]	[0.0226]	[0.127]	[0.0906]
Supervisory Meeting x Southwest Precinct	-0.00384	0.0342	0.102*	0.024	0.0911	-0.0353
	[0.0287]	[0.103]	[0.0499]	[0.0267]	[0.189]	[0.0909]
Supervisory Meeting x West Precinct	0.0118	0.101	0.0782+	0.0148	0.235	0.0126
	[0.0169]	[0.0671]	[0.0402]	[0.0212]	[0.148]	[0.0886]
R^2	0.255	0.215	0.344	0.22	0.328	0.218

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 19B: Effect Heterogeneity by Precinct, Six Week Outcomes (n=1,434)

Panel A: Officer Activity

	After Notification				After Engagement			
	Events	On Views	Time	Report	Events	On Views	Time	Report
Effect of Supervisory Meeting	19.53	0.0101	3.064+	-0.0168	19.75	0.00588	2.246	-0.0152
	[18.11]	[0.0188]	[1.843]	[0.0139]	[13.81]	[0.0238]	[3.031]	[0.0184]
Supervisory Meeting x North Precinct	-16.41	-0.0213	-3.539	0.0117	-14.84	-0.0173	-3.418	0.00513
	[21.24]	[0.0228]	[2.626]	[0.0173]	[17.59]	[0.0302]	[3.779]	[0.0221]
Supervisory Meeting x South Precinct	-47.20*	-0.0121	1.597	0.00906	-53.5**	-0.0052	0.684	0.00866
	[21.40]	[0.0262]	[2.815]	[0.0199]	[18.58]	[0.0320]	[3.704]	[0.0237]
Supervisory Meeting x Southwest Precinct	-21.57	-0.007	-3.174	0.023	-22.05	-0.0015	-2.308	0.02
	[21.30]	[0.0309]	[2.706]	[0.0238]	[17.82]	[0.0343]	[3.604]	[0.0268]
Supervisory Meeting x West Precinct	-15.17	-0.0229	-1.147	0.0238	-13.48	-0.0293	-0.51	0.0219
	[23.37]	[0.0246]	[2.461]	[0.0177]	[19.00]	[0.0287]	[3.457]	[0.0206]
R^2	0.203	0.278	0.174	0.226	0.278	0.315	0.229	0.284

Panel B: Incident Outcomes

	After Notification			After Engagement		
	Arrests	Force	Complaints	Arrests	Force	Complaints
Effect of Supervisory Meeting	-0.0185+	-0.239	-0.0839	-0.0139	-0.127	-0.0407
	[0.0110]	[0.166]	[0.0911]	[0.0136]	[0.220]	[0.0902]
Supervisory Meeting x North Precinct	0.0116	0.169	0.13	-0.0004	0.06	0.0993
	[0.0128]	[0.179]	[0.0948]	[0.0163]	[0.227]	[0.0946]
Supervisory Meeting x South Precinct	0.02	0.166	-0.0132	0.0134	0.0599	-0.0602
	[0.0131]	[0.187]	[0.139]	[0.0155]	[0.235]	[0.139]
Supervisory Meeting x Southwest Precinct	0.00938	0.075	0.104	0.00795	-0.0502	0.0605
	[0.0152]	[0.219]	[0.130]	[0.0172]	[0.245]	[0.129]
Supervisory Meeting x West Precinct	0.0175	0.266	0.159	0.0118	0.137	0.145
	[0.0125]	[0.193]	[0.111]	[0.0147]	[0.238]	[0.110]
R^2	0.188	0.169	0.226	0.180	0.186	0.205

Notes: The dependent variable is the linear difference in outcomes relative to pre-Notification period for each officer. All regressions include wave by treatment group fixed effects. Robust standard errors in outcomes within officer (320 clusters)

*+ p< .10 * p<.05 ** p<.01 *** p<.001*

Table 20A: Item Analysis (n= 61)

Item 1: To what extent did this meeting with this supervisor make you feel what you had to say actually mattered to this supervisor?			
N:	61	Median:	4
Range:	1 to 5	SD:	1.05
Mean:	4.0	Skew:	-1.08

Table 20B: Item Analysis (n= 61)

Item 2: To what extent did this meeting with this supervisor make you wish more supervisors in the SPD used this type of feedback approach?

N:	61	Median:	3
Range:	1 to 5	SD:	1.33
Mean:	3.07	Skew:	-.08

Table 20C: Item Analysis (n= 56)

Item 3: To what extent did this meeting with this supervisor make you feel more respected than in past supervisory meetings?			
N:	56	Median:	3
Range:	1 to 5	SD:	1.46
Mean:	2.75	Skew:	.05

Table 20D: Item Analysis (n= 56)

Item 4: To what extent did this meeting with this supervisor feel like a fair and just system for providing feedback to officers?

N:	56	Median:	4
Range:	1 to 5	SD:	1.13
Mean:	3.55	Skew:	-.85

Table 20E: Item Analysis (n= 65)

Item 5: To what extent did this meeting with this supervisor help to stimulate your memory of the event(s)?

N:	65	Median:	4
Range:	1 to 5	SD:	1.17
Mean:	3.48	Skew:	-.60

Table 20F: Item Analysis (n= 65)

Item 6: To what extent did this meeting with this supervisor allow you to recall specific details of the event(s)?

N:	65	Median:	4
Range:	1 to 5	SD:	1.20
Mean:	3.45	Skew:	-.60

Table 20G: Item Analysis (n= 64)

Item 7: To what extent did this meeting with this supervisor allow you the chance to explain your perspective without feeling judged?

N:	64	Median:	4
Range:	1 to 5	SD:	1.20
Mean:	3.89	Skew:	-1.04

Table 20H: Item Analysis (n= 66)

Item 8: To what extent did this meeting with this supervisor make you feel you were being unduly criticized?

N:	66	Median:	1
Range:	1 to 5	SD:	1.10
Mean:	1.42	Skew:	2.47

Table 20I: Item Analysis (n= 64)

Item 9: To what extent did this meeting with this supervisor help you to understand ways in which you could improve future encounters?

N:	64	Median:	3
Range:	1 to 5	SD:	1.26
Mean:	2.77	Skew:	.26

Table 21: West Precinct Responses Compared to All Others

Question #	WEST PRECINCT (n = 15)			ALL OTHER PRECINCTS (n=27)		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
1	4.27	4	.88	4.0	4	1.15
2	3.60	3	.99	3.15	3	1.38
3	3.47	4	1.46	3.04	3	1.17
4	3.71	4	1.14	3.73	4	1.02
5	3.73	4	1.28	3.48	4	1.23
6	3.80	4	1.15	3.41	3	1.23
7	4.50	5	.76	3.92	4	1.00
8	1.33	1	1.05	1.37	1	1.06
9	3.07	3	1.22	2.89	3	1.26