

Where did it all go wrong? Implementation failure—and more—in a field experiment of procedural justice policing

Sarah MacQueen¹ · Ben Bradford²

Published online: 15 December 2016

© Springer Science+Business Media Dordrecht 2016

Abstract

Objectives This paper presents the findings from a retrospective qualitative process evaluation to the Scottish Community Engagement Trial (ScotCET). The study explores the unanticipated results of a randomized field trial testing the effect of ‘procedurally just’ modes of road policing on public perceptions of police. The ScotCET intervention failed to produce the hypothesized results, producing instead significant, and unexplained, negative effects on key aspects of public perception. The present study seeks to examine, from the perspectives of officers implementing the experiment, what the impacts (intended or otherwise) of participation were.

Methods Group interviews were held within the ScotCET experiment ‘units’ to explore how officers had collectively interpreted and framed ScotCET, and responded as a group to its requirements/demands. Nine groups were held over a 5-month period post experiment completion.

Results Findings indicate that communication breakdowns during the ScotCET implementation led to misunderstandings of its aims and objectives, and of the requirements placed on officers. Within the context of organizational reform and perceived organizational ‘injustice’, commonly cited aspects of police culture were invoked to facilitate non-compliance with aspects of the experimental intervention, leading to implementation failures, and, possibly, a diffuse negative effect on the attitudes and behaviors of experiment officers.

Electronic supplementary material The online version of this article (doi:10.1007/s11292-016-9278-7) contains supplementary material, which is available to authorized users.

✉ Ben Bradford
ben.bradford@crim.ox.ac.uk

¹ Scottish Centre for Crime and Justice Research, University of Edinburgh Law School, Edinburgh, UK

² Centre for Criminology, Faculty of Law, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK

Conclusions Organizational structures and processes, and coercive top-down direction, are insufficient to ensure successful implementation of policing research, and, by implication, policing reforms, particularly those that demand alternative ways of ‘doing’ policing and ‘seeing’ citizens. Greater investment in organizational justice and encouraging openness to evidence-led knowledge is needed to promote change.

Keywords Experimental research · Implementation failure · Organizational justice · Police culture · Procedural justice

Introduction

The use of experimental methodologies within criminology is a growing phenomenon, yet there has been relatively little thought given as to how we might interpret the outcomes of experiments in the face of unexpected, null or negative findings. Often, there is little in the data gathered to suggest how or why the results emerged, with a focus on quantitative data coming at the detriment of understanding the context and processes of implementation. This has led to presentation of what are frequently decontextualized findings, with little space for learning from mistakes (Sherman and Strang 2004, see also Bonell et al. 2012; Moore et al. 2015). In this paper, we explore these issues in the context of one specific experimental study, conducted within one particular criminological field: policing. The Scottish Community Engagement Trial (ScotCET), in its attempts to replicate as closely as possible a previously successful randomized field trial testing the procedural justice antecedents of trust, confidence and police legitimacy within a different policing context, yielded a set of unexpected negative outcomes (see MacQueen and Bradford 2015). In the absence of critical explanatory data, there was little information available to determine how and why the outcomes arose, and how these potentially harmful effects might be avoided in future interventions and developing models of policing.

This paper presents the findings from a retrospectively conducted qualitative process evaluation, designed to explore possible explanations for these unexpected results. Examining the context in which the study took place, and engaging with officers’ accounts of what was happening ‘on the ground’ during ScotCET (in terms of its demands on officers and its timing within a context of wider organizational change), we explore previously undocumented aspects of the study to address these critical questions. Anticipating the discussion below, it is suggested that navigating organizational hierarchies and cultures to enable effective communication with officers implementing experimental studies is critical to gaining trust and ‘buy-in’ and to fielding a successful experimental intervention. The implications of our findings for future research, experimental and otherwise, and the development and implementation of alternative models of policing, are also considered.

The Scottish Community Engagement Trial

The Scottish Community Engagement Trial (ScotCET) was a large-scale randomized field experiment, conducted in 2013/14, which aimed to replicate the Queensland

Community Engagement Trial (QCET; see Mazerolle et al. 2012). Drawing on procedural justice theory, which emphasizes the importance of fairness in processes of interaction in shaping public trust and confidence in authority figures and judgments of their legitimacy (Tyler 2006; Tyler and Huo 2002), QCET demonstrated that altering police officers' style of interaction and communication during routine police-initiated encounters with members of the public could positively influence attitudes towards the police. Signaling dignity and respect, inviting citizen participation, and providing clear explanation during encounters was found to enhance public satisfaction with the police, to improve levels of trust and confidence, and to encourage the conferment of greater legitimacy upon police (Mazerolle et al. 2012). QCET contributed to an expanding evidence base supporting the importance of procedural justice in shaping perceptions of the police and, through the application of robust experimental methods, was the first study to demonstrate a causal link between implementation of procedurally just forms of policing and formation of public opinion and conferment of legitimacy.

Procedural justice theory is increasingly used to inform the development of policing policies and practice in Scotland and beyond, but hitherto only limited evidence existed on how its various elements should be operationalized across different policing contexts. ScotCET was funded to address this gap and inform the *Reassuring the Public* component of the Scottish Government's *Justice Strategy for Scotland* (Scottish Government 2012) by testing whether the implementation of communication practices like those adopted in QCET might result in similarly positive outcomes in the Scottish context.

Applying the same broad approach as QCET, the ScotCET intervention comprised a series of key messages to include within interactions with citizens (in this case drivers stopped during a national road safety campaign), with the intention that each encounter in the experimental condition should have included *all* of the key elements of the procedural justice model. Unlike QCET, the aim was not to provide officers with a verbatim script but to equip them with a series of prompts to include these messages within encounters that already included a significant level of interaction. As in QCET, however, a leaflet was designed for distribution by the experiment group to reinforce the perceived fairness of the interaction and the wider safety campaign (see MacQueen and Bradford 2015).

During the trial period, half of the road police units operating in Scotland were randomly assigned to the experimental condition via a 'matched pairs' or block randomized design. At the outset, both control and experimental units operated 'business-as-usual' whilst distributing a questionnaire to drivers (the pre-period). Midway through the road safety campaign the experimental units were instructed to work under experimental conditions, incorporating the key messages into their interactions and distributing leaflets in every encounter thereafter (the post-period). A substantial number of operational officers were actively engaged in the design of the intervention, and outreach to officers in the experimental units regarding the rationale for the study and its aims and objectives was undertaken at the inception and delivery stages. Prior to the 'post' period, the research team visited each of the experimental units to verbally brief officers on the study and the specific experimental intervention, but it was not possible to involve in this process all road police officers given responsibility for experiment delivery.

To counter this problem, the research team developed a communication strategy with officers from the road police Management Support Unit (which oversees road traffic policing in Scotland). A written briefing was prepared for distribution to all experimental officers. This was cascaded from the Management Support Unit to each inspector, and then to sergeants (PS) and police constables (PC). The briefing was intended to provide supplementary information about study aims and objectives, and instructions about the requirements of implementation. These highlighted the need to work in a set of key 'procedural justice' messages during encounters, provided an illustrative worked example, and suggested a structure for proceeding. An A6 aide memoir for officers to carry on duty was also included. The briefing asked inspectors to relay to officers that they could follow the structure or 'script' (deliberately placed in inverted commas in reference to critiques by inspectors and officers of the notion of a script during preliminary field and design work) provided to them in the briefing materials¹. As detailed above, the intention was not that the officers should follow a given script verbatim, as the preparatory work indicated a looser set of messages was more appropriate to bringing the essence of procedural justice 'naturally' into any given encounter. Throughout the trial period, control and experimental units continued to distribute the same questionnaire to measure public opinion and perception following encounters, and to allow comparison of the pre and post periods and any difference in outcomes for control and experimental units.

ScotCET tested a series of hypotheses that assumed the positive findings from the original Queensland study would be repeated. However, in the event, analysis demonstrated that the ScotCET experimental intervention had the opposite effect to that predicted. In the control areas, judgments of the fairness of police and overall satisfaction with encounters improved somewhat between the pre and post periods. However, no such improvement occurred in the experimental sites. Rather, the intervention appeared to inhibit such increases in levels of public satisfaction and undermine, to a small extent, the belief that the police had adhered to principles of procedural justice during encounters. The intervention had little impact on trust, confidence and judgments of legitimacy, although in all cases the effect of the treatment was negative, even if not statistically significant (MacQueen and Bradford 2015).

In view of the growing body of evidence that suggested the experimental intervention would be successful, and the careful design of the experiment in collaboration with experienced police officers attuned to public engagement, the results are puzzling. The data gathered successfully captured the outcome of the experimental intervention, and the robust design and internal validity of the ScotCET experiment assures us that the negative effects observed within our experiment group can be directly attributed to the intervention, or factors associated with the intervention (Farrington and Welsh 2005; MacQueen and Bradford 2015; Shadish et al. 2002). However, no data were gathered that could *explain* the perplexing outcome of the trial and exactly what it was about the intervention that had the unanticipated negative effect on public opinion.

¹ Copies of the key messages and the aide memoir are presented in Appendices 1 and 2.

Explaining the ScotCET findings

In the absence of sufficient explanatory data or information, we can only speculate on possible explanations for ScotCET's unanticipated findings. However, three possibilities suggest themselves. Firstly, we may accept the result of the trial at face value, concluding that procedurally just modes of policing, as implemented in ScotCET, can lead to detrimental or undesirable outcomes. That is, there may be unintended consequences of the particular model of communication implemented in the study. It is possible, for example, that the ScotCET leaflet produced the negative effect.² It was intended to provide for a sense of openness, honesty and explanation of actions—perhaps the drivers stopped did not value these things, or the leaflet failed to deliver them in a meaningful manner (cf. Hohl et al. 2010; Wünsch and Hohl 2009)?

Yet, while leaflet distribution appeared to be consistent across the experimental units, a small group of respondents in the experimental condition ($n = 20$) did not recall receiving the leaflet—and analysis suggested they were *more* likely to voice negative opinions than those who did recall receiving it ($n = 154$). To give three examples: first, 89% of respondents who recalled receiving the leaflet agreed that ‘the officer was doing the right thing’ during the stop, compared with 60% of those who did not recall the leaflet ($X^2 = 12.4$; $p < .0005$); second, 71% of the leafleted group felt they had been given an opportunity to voice their views, compared with 47% of the non-leafleted group ($X^2 = 4.4$; $p = .04$); third, 92% of the former were satisfied with the way they were treated during the stop, compared with 80% of the latter ($X^2 = 3.1$; $p = .08$). While this evidence is far from conclusive (not least because it is unclear whether drivers actually did not receive the leaflet or whether they simply did not remember it), it does at least suggest that the overall negative results cannot be attributed to the leaflet alone. It may, however, be that the wider communication model, or perhaps other key components of it, might hold the key; but without understanding more about how it affected those delivering it, we cannot pinpoint exactly what it was about our model for delivering procedurally fair policing that led to these consequences.

A second possibility is implementation failure. Experimental outcomes can be affected by the overall success of intervention implementation and the maintenance of ‘treatment fidelity’ (Farrington and Welsh 2005; Gartin 1995; Weisburd 2000). Five elements define the fidelity of implementation of an intervention, program or reform process (Hassall and Lovell 2015; Hasson 2010):

- (1) *adherence* to the planned design, (2) *exposure* or *dose* (amount delivered), (3) *quality* of the delivery, (4) *participant responsiveness* and (5) *program differentiation* (presence or absence of the essential elements of the reform/program. (Hassall and Lovell 2015: 508, emphasis in original)

It is feasible that ScotCET failed on some or all of these criteria. It is often suggested that locating experimental studies within hierarchical organizations like the police increases the likelihood of successful implementation (Weisburd 2000, 2005). With sufficient buy-in and support at senior levels, the power of the chain of command

² A leaflet was also provided to drivers in the experimental condition of QCET.

should ensure that officers deliver experiments as intended. However, Weisburd (2000) has noted the difficulties in securing buy-in and treatment fidelity from agencies where individuals responsible for intervention delivery possess considerable discretionary power. The example he provides is that of judges, but sufficient evidence exists to suggest police officers enjoy their own autonomy and discretion ‘on the street’. Thus, while senior road police officers were strongly supportive of ScotCET from inception to post-implementation, the reliance on organizational hierarchies to percolate the justification for the study and its requirements down to individual officers ran the risk of allowing treatment dilution to occur (Gartin 1995). Once communication reached the operational unit level the trial was effectively handed over to the ‘mid-’ and ‘street-level’ officers to ensure implementation.

Thus, two issues faced ScotCET in the implementation phase: a reliance on top-down command to motivate and regulate officer involvement; and ceding the task of implementation and ensuring treatment fidelity to street-level officers who exercised considerable discretion (operating across widespread geographies with minimal supervision from senior officers). However, while various implementation failures may have occurred, it is unclear whether these can explain the *negative* effects observed. Driver experiences in the experimental group did not simply ‘revert’ to the baseline of the control group as one might expect given failure to implement an experimental condition; they became worse (albeit that this effect arose because an improvement in opinions in the control group was not observed in the experimental group).

As such, we might thirdly consider a slight reframing of the question of implementation failure. At the threshold, it seems plausible to suggest that there may have been something about the trial itself, and/or the organizational climate within which it was implemented, which triggered a negative response among the officers in the experimental group; this response may have influenced the way they interacted with motorists, resulting in the observed outcome. Implementation might even have been broadly successful—or at least consistent—but inclusion in the experiment may have triggered an adverse response among the officers involved. As explained below, there are several reasons to suggest that within policing in general, and the context of Scottish policing at the time of ScotCET in particular, asking police officers to implement an intervention might cause them to react negatively and, in turn, produce unexpected results. This, then, is a distinct form of implementation failure, associated with the context within which the trial was fielded and which produced not a null, but a negative effect.

The importance of organizational context

Findings from two strands of policing research support that the issues detailed above merit serious consideration. First, research on police occupational culture has long stressed that a number of core characteristics are associated with police cultures across multiple jurisdictions. These include, among other things, suspicion and cynicism (particularly towards the policies, procedures and leadership of senior management) and a rigid distinction between ‘us’ and ‘them’ (police and ‘outsiders’ of various types) coupled with in-group solidarity (Loftus 2009; Reiner 2010; Waddington 1999). Notably, these characteristics have been blamed for failed police reform projects (Loftus 2010; Skogan 2008), in relation, for example, to engagement in the partnership

working fundamental to community policing programmes (Cordner 2000; Skogan 2008). While the universality and practical implications of ‘police culture’ viewed in these terms has been questioned—there are plainly different police cultures even within specific organizations (Fielding 1988, 1989), and the relationship between what police officers say and what they do is often unclear (Waddington 1999)—anyone who has spent time researching with, or on, the police cannot help but be struck by the regularity with which tropes such as cynicism and suspicion of senior management and outsiders (such as academic researchers) can arise.

Such aspects of police culture(s) would appear inimical to research projects such as ScotCET. Cynics may not see the need for, or simply scorn, efforts to improve police–public relationships, particularly if they perceive these relationships as inherently problematic or confrontational. Perhaps more importantly, suspicion of senior managers and outside researchers may be linked to a reluctance to take on new practices and engage in the experiment in the way required to make it work. It does not seem unreasonable to suggest that many police officers simply do not like being told what to do by outsiders—since, for example, ‘no-one knows the job like they do’. Moreover their high-discretion and low-visibility work environment means they are often in a position to subvert or ignore instructions from senior officers, let alone from people outside the organization with no substantive claim on their attention or resources.

A rather more recent body of work adds another layer of complexity. Studies of organizational justice (Colquitt 2001; Greenberg 2011) within police bodies have begun to draw out links between the way police officers feel treated by supervisors and managers, and their attitudes, and possibly behavior, towards those they police (Bradford et al. 2014; Haas et al. 2015; Myhill and Bradford 2013; Tankebe 2010; Trinkner et al. 2016; Wolfe and Piquero 2011). This work suggests that the readiness of police officers to comply with rules and directives, or to engage in extra-role activities, is shaped in an important sense by the fairness of both their immediate supervisors *and* of wider organizational practices and processes. Similar to the procedural justice model established by Tyler and colleagues (Tyler 2006; Tyler and Blader 2000; Tyler and Huo 2002), theories of organizational justice stress issues of neutrality, respect and accountability (Haas et al. 2015: 11), to which is commonly added the arguably more concrete notion of communication or informational justice (Colquitt 2001). In particular, people in organizations value clear communication about what is going on and why. They also value ‘voice’—usually considered part of procedural justice—and the sense that communication is a two-way process that involves managers listening to the concerns of staff (Folger and Bies 1989). What this research suggests is, of course, that existing relationships between the officers in the experimental group and their managers may have influenced the extent to which the former ‘bought into’ ScotCET, mandated as it very clearly was by senior management within the organization. More generally, their readiness to take on extra-role activity—that is, engage in delivering the intervention—may also have been affected by the way they experienced the behavior of senior officers and the organization as a whole.

These issues become particularly pertinent when one considers the wider context of Scottish policing during the ScotCET timeframe. On April 1, 2013, 8 months prior to the commencement of the trial, the key provisions of the Police and Fire Reform

(Scotland) Act 2012 were brought into force, merging eight police force areas into a single national force, Police Scotland, under the leadership of one Chief Constable. This process represents the biggest change to Scottish policing since the 19th Century (Terpstra and Fyfe 2015), leading to dramatic overhauling of organizational structures and management, and arrangements for governance and accountability.

The merger shone a new political spotlight on Scottish policing, sparking extensive debate and criticism involving local and national government, media, academics and police bodies. The national force has been mired in controversies over top-down changes to policing 'style', such as the proactive turn towards 'stop and search' tactics and its resultant excessive use (Murray 2014), the deployment of armed police officers on routine patrol duties, and the use of targets and key performance indicators to 'enhance' police activity. New structures have appeared insufficient to cope with the demands of national policing, evidenced by the deaths of a young couple trapped in the wreckage of their car for 3 days after police failed to respond to accident reports, apparently due to communication 'lost' between the national call centers and local stations. While the day-to-day activities of street-level officers have been scrutinized, the spotlight has focused particularly on the culture of leadership, the presence of personal agendas and cronyism at the top of the organization, and the apparent lack of accountability and responsibility. Unsurprisingly, the publicized results of the first national staff survey (conducted after fieldwork for the present study took place) revealed widespread discontent within Police Scotland. While working relations between staff, officers and their immediate supervisors/line management appear characterized by trust and respect, the survey results portray a force where officers feel under-resourced with a poor work/life balance, and, critically, see themselves as under-valued within the wider organization. Senior management are perceived as failing to recognize staff contribution or show concern for staff well-being, and as unwilling to respond to and address the issues raised. Moreover, the survey highlighted a lack of support within the organization for the reforms experienced and a lack of trust in senior management to act in line with the purported values of the national force³.

Thus, the extant literature, coupled with the particular context within which ScotCET was taking place (much of which has come to light since the experiment was conducted), offers us a series of 'sensitizing concepts' that may help to explain ScotCET's unexpected findings. The intervention may not have worked in the way it was intended, it may not have been fielded properly, or it may have triggered a more diffuse change in the practice of the officers involved, all of which are possible in view of the issues raised above. But it is unclear which, if any, of these factors are genuinely relevant. Without this knowledge we cannot draw appropriate conclusions. The 'black box' of the experiment (Weisburd 2000) must be opened, in other words, to explore what went wrong, as only through further study will it be possible to extract the trial's lessons for future research and policy development (Gartin 1995; Sherman and Strang 2004).

³ No official report is available but see <http://www.bbc.co.uk/news/uk-scotland-34424206> (last accessed 13th August 2016)

Methods and analytical approach

The aim of the present study was to address these gaps in our understanding and explore, from the perspective of those working ‘on the ground’, how the experiment was implemented and its influence (intended or otherwise) on officer practice. Gaining such insights required adoption of a qualitative approach. A group interview method was employed, as participation in ScotCET was a shared officer experience and the ways in which they had collectively interpreted and framed the study were likely to be critical to the ways in which they responded as a group to its requirements/demands. Moreover, this group oriented approach, taking place within officers’ own base stations, offered the closest reflection of the social context in which the original study took place, as well as providing an opportunity to redress any felt power or control imbalance between the research team and the participating officers that the original study methodology may have created (Lavie-Ajayi 2014; Wilkinson 1998).

Nine group interviews were held between January and May 2015, with 55 officers participating in total⁴. All ten original experimental units took part, but due to their relatively small size and geographic proximity, two ‘paired-up’ in one group. Inspectors at each unit were invited to nominate a suitable date and time for the groups to take place. Decisions were taken in view of the demands on the unit at any point in time, avoiding particularly busy days or time periods and clashes with other events. The typical approach of each inspector was to suggest a routine day and to schedule the group to take place at shift changeover times so as to maximize the availability of officers (although participation in the group did not take priority over operational demands). Officers present in the base stations at the scheduled times would attend, and were free to leave at any time. In the event, officers only left the discussion due to operational demands (noted below) or because they had not participated in the original study.

A full breakdown of group composition in each unit is provided in online Appendix 3. Group size varied across the units from two participants (this group started with four, but two left to respond to a road traffic accident) to 14, with the average size being six officers plus the facilitator. Groups typically comprised male officers with several years of police service. In the interests of the protection of participant anonymity, specific details as to the exact length of individual service are not provided here. However, given the pathways to becoming a specialized road police officer, including an initial probationary period of 2 years followed by an extensive recruitment, training and assessment process before qualification, the minimum length of service of any officer would be between 4 and 5 years. Many officers in this study surpassed this minimum by some distance.

Also demonstrated in Appendix 3 are the varied ranks of officers participating. While groups mostly comprised constables, some sergeants participated in the larger groups, and in two the unit inspector was present, observing in one and actively participating in the other. Care was taken by the authors to ensure that all officers, particularly those participating in the presence of their immediate supervisors, felt enabled to provide honest feedback on their experiences of participating in the experimental study. During initial communications, and at

⁴ Two groups had been completed at an earlier date in order to scope the feasibility of the evaluative study.

the outset of each group, it was made clear that the purpose of the discussions was to explore why the experimental study and its particular vision for operationalizing procedurally fair policing, rather than any actions taken by the officers themselves, produced unexpected results. In discussing the experimental results with officers, the authors made sure to emphasize that while they may sound dramatic, they were merely slightly less positive than expected given both previous research evidence and the trajectory of the control results. Officers were invited to participate and contribute as experts, whose voices and honest responses were to be valued as offering key insights as to the impact of the experiment on them and the drivers they encountered. The group context was intended to support the enabling of honest feedback and all of the senior officers involved acknowledged and promoted this during discussions. On the basis of what transpired during the groups and on the findings that emerged, the presence of higher ranked officers appears, as far as we can tell, to have had little influence on the honesty of the officer feedback.

In four groups, one or two officers who were not experimental participants were present⁵ (they had been working elsewhere at the time, or had been on leave). All were currently partnered with an officer who had participated and could not resume operational duties without their partner. Thus, the decision was taken to allow them to participate and offer their opinion on the experiment and its implementation. Most, including a 'new-in-post' unit inspector, simply listened rather than verbally contributing, but a couple offered interesting insight to how they believed participating in the experiment would impact on them and what their response was likely to be. In general, these opinions tended to follow those offered by participants, perhaps serving to underline that the experiment was experienced by the officers as a group rather than as individuals. It must be noted that the risk of a 'group effect' emerging in this particular research context is comparatively high, but we took steps to minimize this risk as far as possible, ensuring that as many officers as possible were invited and able to express an opinion, whether it differed or not from those of others present. Differences of opinions, both across and within groups, did emerge, and we highlight these in our presentation of the results.

The group interviews generally lasted for one hour. Discussion was usually limited to this time period due to other demands on officer time or room space, but in a couple of cases it carried over for longer. The approach of the facilitator was to allow officer-led discussion, but a topic guide was used to retain focus on the experiment and serve as a prompt if needed. Broad topics included the nature of road traffic stops and how these are approached/conducted by officers; how the experiment was implemented and the quality of communication with those involved; the impact of the experiment on officers and their 'delivery' of vehicle stops; and whether other changes over the course of the trial period had been observed by officers, such as in the composition of traffic and subsequent focus of stops over the trial period. In the event, little prompting was used, as officers had very particular memories of the experiment and often a great deal to say about it.

⁵ Including, in two units, senior officers as noted in Appendix 3.

All group interviews (with the exception of two, where a second facilitator was available to make detailed discussion notes) were digitally recorded and transcribed. Transcripts and notes were entered into NVivo 10 and a thematic analytical approach was adopted (see Braun and Clarke 2006). The broad identification of themes within and across groups was informed from the outset by the very specific focus of the interviews, which were oriented around issues of communication, implementation, and what happened ‘on the ground’ during the original study and why. Underneath these broad organizing headers, more specific themes were then identified during analysis, with an iterative approach to coding adopted, such that both authors coded and re-coded the transcripts before agreeing on a final thematic framework.

Themes were identified on the basis of relevance to the research question. The majority cross-cut extensively, with much commonality emerging, and there was broad agreement on key issues across the different groups. Where themes were not universally raised, we acknowledge this in our presentation of findings. Where issues emerged within a small number of units, these were counted as themes only where the points raised were particularly pertinent to the officers who raised them *and* where sufficient external evidence existed to suggest they represented issues of wider reach. As such it may be argued that a hybrid approach of inductive and deductive analysis was undertaken here (see Fereday and Muir-Cochrane 2006), with the authors drawing on a wider body of research evidence and knowledge in order to place officers’ comments in a broader context. The final thematic framework is reflected in the key findings below, with the broad themes identified in analysis organized under three overarching headings to guide the reader through what was reported as happening ‘on the ground’ during the study, the ways in which officers interpreted and reacted to this, and the explanations offered. It must be noted that while distinct themes concerning officers’ practice and interpretations of the study and the wider context in which it was located emerged, all are strongly interrelated in the narratives provided by officers.

Results

What was happening ‘on the ground’?

Communication failure

Despite the time lapse between the experiment and follow-up evaluative study, officers appeared to hold clear memories of what had happened during ScotCET. However, recollection of the study aims and requirements differed quite substantially from ScotCET as designed. It became clear that small, but important, breakdowns in communication had occurred.

Efforts had been made by the research team to ensure the written briefings distributed to the experimental units, which detailed the requirements and rationale of the experimental intervention, were short and to the point, and were supplemented by bullet points to facilitate easy transfer from Inspector to officers. In the event, officers appear to have received an instruction set distilled even further. For most (nine out of

ten units) this had been a mere three-point directive summary, delivered verbally, comprising questionnaire and leaflet distribution, and use of the aide memoire as a *script* for all stops:

... we definitely got the impression that this is what we had to do. Not a case of, “Well, you can make it as flexible as you want.” It was a case of “We’ve got to do this” (Officer, unit 1)

... maybe I’ve missed being told it, but I was under the impression that what you had typed was what we had to say almost word for word. (Officer, unit 3)

In one unit, the instruction provided was further reduced, such that any reference to key messages, aide memoires or scripts was dropped entirely:

Officer 1: All the things we got here was the questionnaires that we’d to hand out. We never got anything like that. We never had anything.

Officer 2: I’ve got to be honest. I don’t remember ever seeing these either. We had the small leaflets. We had the...

Officer 1: ...big envelopes...basically the packet that we gave to them. No the aide-memoire, no. (Unit 7)

Some attributed the delivery of this over-simplified set of directives, and their strict interpretation, to the nature of policing and the style of working that was encouraged (and at times necessary) within the organization:

It’s something that happens quite a lot, if you’re looking at briefing, somebody has decided that the information needs to be put on...but when you read it as well, you take the main points, and probably...the rest of it goes in the bin kind of thing (laughter). (Officer, unit 2)

Police officers, being police officers — and especially, I would say, traffic officers ... If you tell them on a piece of paper, “I want you to do this,” they will do exactly what that says on that bit of paper (laughter); they will not think outside the box, because of the nature of what we do ... you develop this very logical but almost at times immovable way of doing something, so if you read something on a bit of paper that says, “You do this, you do this,” that’s exactly what you will get ... (Officer, unit 3)

Implementation failure

The impact of the briefings delivered to officers varied. For some officers the script command made no difference, dismissed from the outset as something they ‘were doing anyway’:

I think everybody here covers everything in that just as a matter of fact. We don't refer to that, so in actual fact I think that what came back is just a reflection on how we do our job at a road check. It had no influence on that. I had a look through that and I thought, "I do that anyway," and just put it away. (Officer, unit 5)

Others appear to have point-blank refused to deliver the intervention—or, at least, the verbal component of it:

Well, I looked at it [the briefing/script] and said, "I'm not doing that." (Officer, unit 1)

Amongst these officers, there was simply no adherence to the verbal requirements of a) the intervention as designed or b) the intervention as communicated. Others initially attempted to comply with the script command, but met with limited success and abandoned compliance shortly thereafter:

I lasted a week of people sort of looking, and laughing, and thinking... (Officer, unit 1)

...I think for most folks reading of that, we kind of thought right, this is some sort of experimental script, I think as [my colleague] says, I think folk maybe strived to stick quite rigidly to it, but found it difficult ... Certainly road checks always, I felt it was, you kind of looked at it and tried it, and you invariably slipped back into what you were doing anyway. (Officer, unit 2)

...I've used it once and I felt uncomfortable doing that; I didn't use it again, primarily on that basis. (Officer, unit 3)

At face value then, verbal communication across the stops conducted by the experimental units failed to achieve the required level of consistency. However, due to inclusion of a check question for drivers on the leaflet receipt, we can be certain that there was *some* consistency across the experiment units. As noted above, among the 174 survey forms returned by the experimental group in the post period only 20 respondents did not recall receiving a leaflet, and these failures were distributed across five of the ten units involved. Thus, despite the apparent delivery failure in terms of the verbal component of the experiment, some element of the design survived and a clear, albeit small, difference between experimental and control units was retained—although again as noted this is not sufficient to explain the negative results overall.

Impact of failure

While the experimental units did not deliver the verbal messages with any consistency, there was a degree of consistency in the nature of the communication failure occurring. All units had received minimal and/or incorrect communication about the experiment's requirements and the rationale behind them, with nine out of ten receiving a direction to deliver a script without further explanation or rationale.

Officer interpretation and reaction

Discussion revealed much similarity in the ways officers across the experimental units interpreted and reacted to the demands placed upon them. Overall, the imposition of the study had not gone down well. Some (three units) had interpreted the intervention as being imposed from the top due to perceived failings in current practices:

It saddens me a bit, because clearly that's the reason the study was put in place, because they're wanting us to improve in the way we're doing things, but we're already doing that, I think, as part of the job. (Officer, unit 1)

... but it did kind of give me the impression that what was on the scripted test was best practice ... which then automatically makes you think, "If this is best practice, this is not how I'm speaking to the public at the moment, albeit I'm covering the same points but I'm not saying as those words, then I must be doing something wrong." (Officer, unit 3)

Similar beliefs were voiced in relation to the questionnaire distributed to drivers. Officers in three units had assumed the questionnaire was to gauge public satisfaction (not entirely incorrectly), but that this was driven by senior management in response to perceived inadequacies in practice and, critically, complaints against officers:

Probably on the back of the suspicion of, "Okay, we're asking these members of the public now to fill in a sheet about us." You know, so even though it was, "Yes, okay, whatever," there's also that little bit of suspicion at the very beginning in regard to that too ... yes, somebody's checking up on us. (Officer, unit 5)

These interpretations of the rationale behind the study, and the nature of the demands placed on officers, led to a very negative set of reactions across the experimental units *and* the different ranks of officers within units:

... see when you handed out the questionnaire and I read it. I found it insulting ... aye. Because it's like, "What do you think I do when I speak to people?" (Officer [PC], unit 4)

It came across more negatively I think, than anything. It was handing a cop with whatever service and saying, "Right, now you've been doing it wrong all these years. Hand this out, this is what you ask the motorist," or, "This is what you say to them." It's like, "Who do they think they are? Telling us how to speak to people." (Officer [PC], unit 5)

...handing officers who've got lots and lots of service a thing where they have to read out to people ... I think, probably, I don't know if other officers agree but, getting something like a handout is kinda undermining, you know? (Officer [PS], unit 8)

The negative response was further compounded when officers had made initial attempts to utilize the script verbatim. Broadly, officers reported that rigidly following a script had a damaging effect on the quality of their interactions with the public, often citing the removal of the natural flow of encounters and their ability to include a ‘personal touch’ (Officer, unit 1), and how this made them feel uncomfortable:

Everybody here is well used to getting out and stopping people; you deal with them conversationally going through the stuff, if you’re suddenly confronted with “right here, you need to remember all this” and you’re dealing with somebody and you’re going to breathalyse them, wait a minute, there’s something here [gestures to fake notepad] but that’s... you can sound stilted... (Officer, unit 8)

As for an officer reading that out, I wasn’t happy speaking to someone in that manner, because I wouldn’t want to get spoken to in that way ... I think, from the officer’s point of view of doing it, they actually felt embarrassed about having to say that to a member of the public. (Officer, unit 1)

Perceived driver reaction

Many officers believed that the experiment requirements, as they had understood them, would have a damaging effect on how they were perceived by drivers (six units in total). In general, they felt their experience in dealing with members of the public had taught them the ‘best’ way to handle potentially confrontational encounters, which differed from the envisaged intervention. This included avoiding unnecessary delay or hold up during encounters (a point supported in the evidence base, Mazerolle et al. 2014), and avoiding too much ‘lecturing’. The insistence in the intervention for clear explanation of *why* a road safety campaign was needed was described as something that drivers were not interested in:

If I were a member of the public, I would be like, “Oh, I can’t be bothered reading that.” ... bullet points, bullet points is all they’re going to read. They’ve just been stopped ... they’ve been inconvenienced, they’re late for work. That’s going in the bin...they’re no even going to read that.... (Officer, unit 8)

Officers argued providing too much explanation or rationale could be detrimental in situations where a driver had incurred a ticket or penalty:

... you’re doing one or the other, you’re either educating and encouraging them, and warning them about their actions and trying to show them what could go wrong, and they’re made aware of that, and you bear that in mind, and cheerio ... Or, it’s the enforcement ... I think when you start to mingle the two, you’re having two bites at the person, and that’s where they’re starting to feel, you’re just kicking me here. (Officer, unit 2)

Deviation from ‘standard’ practice also ran the risk of drivers perceiving the officers as less competent and confident, and as hindering professional duty and relationship building:

...if a cop stops you and you’re having to read from a script, then the driver is thinking “This chap doesn’t know what he’s doing.” (laughter) (Officer, unit 3)

By sticking to that, it could sometimes come across as quite robotic and patronizing sometimes, like “Thanks very much for wearing your seatbelt.” They would think you were being a bit of smartarse, kind of thing (laughter). (Officer, unit 3)

Thus the experimental intervention, as communicated, seems to have had a series of negative effects on officers. In the absence of sufficient explanation, officers across different units ‘filled in the gaps’ in very similar ways, notably assuming the experimental intervention to be the outcome of criticism from the top, or outside, of the organization. Moreover, reactions within and across the units, and the different ranks of officers, proved very similar. Firstly, the challenge to officers’ professional expertise and experience was perceived as disrespectful, insulting and undermining. Where officers attempted to comply in full with the directives received this reaction was reinforced, both by a loss of sense of professional competence and a perceived negative response from drivers, who were deemed likely to be, at best, bored and uninterested or, at worst, judgmental about officers’ intentions and professionalism. The effect appears to have been that officers rejected the directives received and retreated, somewhat defensively, to a professional ‘comfort zone’ wherein they retained control over how to handle drivers and encounters. This may be viewed as a series of unintended consequences of the experimental implementation, and, arguably, suggests a very particular negative reaction against the study across and within the experimental units.

Explaining organizational and individual responses to the study

Context of change and organizational ‘injustice’

Why, then, did experimental officers react to the experiment in what seems to have been a consistently negative fashion? A central issue in many of the groups was the wider organizational change that served as an important backdrop to ScotCET. One group discussed long-term erosions to officers’ discretion. The remainder focused specifically on change following the merger, which one group positioned as unprecedented in nature and extent; all commented on the shift in priorities and focus they had experienced. The changes were described as being imposed top-down, and as bearing little or no relation to ‘local’ knowledge and experience. Structural changes had occurred across all of the legacy force areas, with new priorities and focus introduced in tandem:

This department has been under a huge change, substantial change in a very short period of time, [which has] altered the focus of the department ... we are vastly altered from what we were in a very short period of time. (Officer, unit 8)

Key changes discussed elsewhere were new expectations of what officers should be doing and how they ought to prove their ‘worth’ in the organization. For example, officers in three separate units discussed the shift in emphasis from ‘stop and speak’ during road safety campaigns to ‘stop and book’, and the development of performance indicators based on issuing tickets:

Officer 1: It wasn’t time to be nicey-nicey to a lot of people.

Officer 2: Yes, warnings weren’t counted.

Officer 1: Warnings didn’t exist.

Officer 2: So pre-[police reform], if we’d stopped somebody where we used to be based, we would give them a warning depending on the nature of the offence or something like that, and that would be fine. As [my colleague] was saying, we used to have and we piloted a thing out in [local town] for that, however it didn’t matter. Warnings didn’t count. If you didn’t issue a ticket, if you didn’t fine that person that was the only way to show that you had done some work that day. If you hadn’t done a certain number- but then numbers didn’t count. (Unit 5)

Alongside resource cuts (departmental cutbacks and pay cuts in real terms), this combination of factors was described as damaging to morale and fuelling ‘suspicion’ of senior management, who were seen to be imposing new policing styles, particular to one legacy force area, but antithetical to operational wisdom in others. Officers reflected on this as affecting their perception of, and reaction to, the demands of the experimental study:

[The initial reaction was] “Is this because [X legacy force] don’t know how to speak to people?!” (Officer, unit 6)

Officer 1: At that time, our heads were in a different place.

Officer 2: It wasn’t the questionnaires, it was the whole build up at the end of a long slog, basically. (Unit 5)

...morale, morale is a huge impacting factor. So yes, timing is everything... there’s still conflict because you have people who want to maintain their safe zone being forced into somewhere...The rest of the forces are joining us and don’t like where we’re going, so you have hostility and a resentment to the path that we’re undertaking...although basically the job we do is the same...but if you’re looking for morale, then yeah, that’s different. (Officer, unit 8)

Some officers acknowledged that this may have affected the quality of interaction between them and the drivers stopped over the campaign period:

Officer 1: It has to, yes. Even if you don’t think it, I think the way you’re dealing with it. It’s like, “Right, next. Next”.

Officer 2: Subconsciously, possibly. Or even if you're not necessarily in a completely happy place, so the way you speak to anybody is not necessarily the same as you would if you were in your happy place. (Unit 5)

It would appear that ScotCET was viewed in light of—and maybe even as part of—the wider change process being imposed at the time. The imposition of the study (without, recall, sufficient explanation or justification) and its apparent challenges to existing operational practice and wisdom, reflected the imposition of wider changes to Scottish policing, and may have compounded existing low morale and suspicion. The discussion of these issues not only helps illuminate why officers reacted in the ways they did, but also supports the argument that circumstances and communication failures combined to have a detrimental impact on officers' working experience. As noted, many seem also to have found being asked to deliver the intervention disrespectful of their status as experienced officers. Another potential impact on the quality of interaction in some, if not all, of the encounters in the experimental units may therefore have been that the officers involved felt even more disgruntled than would otherwise have been the case—and that this came out in their interactions with drivers.

Dominant operational police cultures

A compounding issue was the strong sense of a shared knowledge and culture amongst the officers. Two components emerged: a particular sense of the role of the road police officer and the nature of the job, combined with the superiority of knowledge gained 'on the job'; and cynicism regarding the input and judgment of outside agencies and citizens who could not understand the former.

While some officers stressed the importance of engagement with members of the public, others were clear that their role was predominantly targeting road traffic offenders through detection and deterrence, rather than engaging in prolonged dialogue:

Our time is precious as well, don't forget, if we're engaged in a road safety drink drive campaign. That wasn't a drink driver, but there are drink drivers out there, we know that, so we want to get rid of them, in a polite way, of course, (laughter) and then focus on the people that are drink driving. This is taking up our valuable time as well, if we're going to introduce this kind of means of interaction with the public. (Officer, unit 1)

As regards interactions with members of the public, it was almost universally agreed that on-the-job experience was the most important factor shaping ability to manage encounters. Officers assumed a great deal about what citizens want and expect from encounters with the police, and stressed the importance of exceeding perceived motorist expectations in terms of professionalism and efficiency, whilst simultaneously avoiding conflict and confrontation through minimal interaction; somewhat, although not entirely, counter to the procedural justice model envisaged in ScotCET:

If we stop a vehicle we know if somebody has maybe been drinking we're going to breath test them. If not, "Yes, thanks very much. We're happy. Okay, we will

let you go on your way.” Away they go. They’re happy with the interaction they’ve had with us, because it’s been quite short, admittedly, but we were very polite with them, they know why we spoke to them, and they’ve gone on their way, and we’ve not taken up any more of their valuable time. (Officer, unit 1)

I think that we’re more professional than the beat, but that’s only because we deal with decent people, whereas the beat deal with scumbags. Do you know what I mean? (Officer, unit 5)

I’m wanting, you know, for the situation to diffuse as quickly as possible. Or, in layman’s terms, just get, well, rid of it. Aye, aye, move it along, not sit explaining something like that to them. (Officer, unit 7)

A number of officers expressed the opinion that the study was attempting to ‘fix something that wasn’t broken’, and had been imposed by those, whether senior management or outsiders, who failed to grasp the nature of the job and the nature of the people being dealt with:

... there’s an assumption that every conversation we have with the public is always going to be a nice, happy, polite conversation. Sometimes it isn’t ... for some people higher up the tree or, you know, sitting in a nice office somewhere, they don’t like to think that that happens but sometimes a member of the public just need to be told, “What you did was wrong.” ... and I’m not going to say, “Thanks,” to certain types of individual if it’s not appropriate to say thanks. (Officer, unit 5)

... if you’re not experienced in dealing with people in our arena, then you’re really not in a position to tell us how to speak to people. (Officer, unit 4)

Moreover, a number of officers across the units expressed a lack of trust or fear about inviting citizens to ‘judge’ their conduct. There were contradictory ideas and opinions here, with officers dismissing members of the public as ‘complainers’, while at the same time describing how happy motorists tended to be with their treatment. It was generally believed, however, that drivers who responded to surveys were likely those with an axe to grind:

... I think your feedback’s going to be skewed. Because if we stop people for a traffic stop they’ve probably done something that they’ve misbehaved or made a mistake and we’re pointing that out. So, they’re probably a little bit grumpy. (Officer, unit 4)

You’ve got, you’re given with three points and a £100 fine. Nobody’s going to be overly happy about it... So you’re giving them an opportunity, you know to vent any frustrations, via the survey. (Officer, unit 2)

Some went further, suggesting that questions on public opinion and satisfaction were leading, and querying the production of the questionnaire. When their concerns over a lack of police input to its design were addressed, this was dismissed:

Officer: Just that one there, “Still thinking about your relationship with the police. How much power do you think the police have over people like you?” I just think that’s kind of — you imagine somebody you’ve just given a ticket, you’ll be like “rant!” I don’t know. I think maybe if the police were involved in writing the — you know if there was like ___.

Facilitator: We did have — there were people from [study police force] ... officers who were helping us.

Officer: Aye, probably pen pushers that have never seen the light of day.

Other officers: *Laughter* (Unit 7)

The levels of cynicism expressed, while not uniform, provide another important contextual issue to consider when seeking to understand officers’ responses to, and interpretations of, ScotCET. The shared cultural factors raised in the group discussions, including the dismissal of the ScotCET intervention as a relevant or meaningful set of activities, and the discrediting of research or the input/ voice of citizens more widely, appear to have allowed or facilitated officers’ retreat from participation by providing a set of justifications for their lack of compliance with the intervention and the directives set by senior management.

Discussion/Conclusions

The findings presented above provide an important methodological and substantive learning opportunity. Despite the recent ‘experimental turn’ in criminology, experimental studies in policing, particularly involving procedural justice, remain comparatively rare, and investigation into the experience of the officers implementing ScotCET provides insight concerning the potential for success—and failure—of this particular method. Key conditions and barriers, some arguably specific to police organizations, may hinder the implementation of experimental trials, and indeed academic research more broadly, as well as inhibit meaningful reform of police practice and values.

To start, though, what do our findings mean for the ScotCET? Why did it ‘go wrong’? The current study was not designed to address formal hypotheses concerning why ScotCET failed to produce the predicted outcomes. Rather, we were concerned to explore with the officers involved what they thought of the experiment and their experience of delivering it—these views, we thought, could provide clues, rather than definitive answers, to the question at hand. It seems clear from the preceding analysis that ScotCET suffered various forms of implementation failure, particularly in relation to motorists’ exposure to the intervention (which will have varied as officers did not adhere to or dropped the guidelines after initially trying to apply them).

Implementation failure does not in itself appear to explain the *negative* results of the experiment, however. Not only did the intervention not have the anticipated positive effect on perceptions of procedural justice, trust and satisfaction, it actually made things worse. While we cannot know for certain why this happened, it seems plausible to suggest that inclusion in the experimental group had a diffuse negative effect on the

attitudes and behaviors of the officers involved, and that this contributed to the overall outcome of the trial. Perhaps because they lacked trust in the motivations and directions of senior management within the changing organizational landscape, because the experiment was not well communicated to them and/or because they experienced inclusion in the trial as an imposition and possibly even a slight on their professional ability and integrity, being asked to deliver a particular set of messages concerning procedural justice may actually have triggered a (small) negative shift in the way officers interacted with motorists. Organizational justice theory would specifically predict that under such conditions officers' attitudes towards their jobs would change, and that they would become less willing to engage in new activities or work with superiors—or outsiders—towards novel or different outcomes. The sense of injustice that seemed to exist at the time of ScotCET, which the trial itself seems to have compounded, may have made the officers somewhat less polite, less forgiving, or simply less communicative than would otherwise have been the case. What occurred, that is, may have not only been implementation failure but a backfire effect arising from the specific organizational context within which ScotCET was implemented, and the apparent gulf between operational officers and senior management.

The limitations of the current study should of course be noted. Its retrospective nature means we are exploring issues after the event, and are wholly dependent on participants' recollection of what was happening at the time. Moreover, our group interview method, while useful for facilitating understanding of collective experience, is likely to have brought shared cultural, rather than operational, aspects of policing to the fore. Individual interviews with officers may have raised different issues or offered alternative insights, although these would also have presented their own set of limitations. Regardless, we are unable to grapple with what was actually happening on the ground, and cannot determine which elements of the often-contradictory discussion during the interviews best reflected what 'really' happened. Finally, we did not speak to officers in the control condition to gauge their reactions and interpretations of the study. If we had, we would have undoubtedly found similar contextual issues. Given that officers in the control condition were left to business-as-usual without challenge, it seems unlikely that they would have reacted in quite the same way as those in the experimental condition. Indeed, it is possible that officers in the control units may have 'acted up' during the trial period if they had taken the same views about the purpose of the questionnaire, an interesting possibility that we were unable to explore.

On reflection, it seems apparent that high quality experimental research requires a strong qualitative component that is built in from the outset. As emphasized by Sherman and Strang (2004), only through an ethnographic approach to experimentation, with in-depth observation and ongoing group and individual discussion throughout the process, can one draw a confident set of conclusions and insights on what was happening and what did or did not influence the results achieved.

Notwithstanding, there are clear messages emerging for those trying to field large-scale RCTs in policing. First, internal structures and processes cannot be relied on to disseminate and communicate details of the intervention, nor advise on how to field it. Indeed, contrary to the idea that police as hierarchical quasi-military organizations should be able to facilitate the control necessary for a field experiment (Weisburd 2005), the nature of the hierarchy of Police Scotland at the time of ScotCET—and perhaps as importantly, street-level officers' perceptions of this hierarchy—may have

militated *against* successful implementation. Relevant guidance and materials appear not to have been distributed as intended and key messages were not always circulated. In the absence of meaningful communication between officers and the mid-level management distributing trial materials, there was suspicion about senior management and outside ‘experts’ and *why* these groups might be trying to tell them how to do their jobs.

Second, this lack of communication, and suspicion, operating within a very particular set of organizational conditions, appear to have fostered a series of barriers to success associated with what are often considered dominant aspects of police culture. Many officers were cynical about the experiment and its aims (in so far as they understood what these were), as well as the merits of research and citizen participation; were oriented toward enforcement aspects policing in a way that led them to downplay the relevance of key principles of procedural justice; and were convinced that ‘cop lore’, learned on the job, was the only form of knowledge they really required. It is important to reiterate, as policing scholars have over many years, that the nature of operational policing allows such attitudes to shape the practice of officers. Working in pairs, well away from supervisory oversight and vested with all the powers granted to a constable under the British system(s) of policing, officers in ScotCET felt able to, and obviously did, circumvent or simply ignore the instructions they had been given. This is not, it has to be said, entirely unreasonable—traffic police in a sparsely populated country such as Scotland are often operating many miles away from any potential back-up, and many of the officers we spoke to were keenly aware that they needed to manage each encounter in a way that avoided conflict to the greatest extent possible. Anything that they saw as interfering in this—such as following a checklist—was arguably always likely to be jettisoned at the first opportunity. Yet the ability of police officers to simply ignore instructions given as part of an implementation process is likely to be a key factor shaping the success (or failure) of that process.

Third, and in relation to the original aims of the experimental study, it seems that while roads police in Scotland did in a general sense accept that treating those they stopped in a procedurally fair manner was important, there was also resistance to aspects of the procedural justice model of policing. We heard, for example, that explaining one’s actions is not always necessary, or even appropriate, and that dialogue can get in the way of a successful encounter. At the heart of such comments may have been the idea that while fair process can be used to ‘oil the wheels’ of an encounter it is not an intrinsic good, and can be ditched as and when necessary or desirable. It is perhaps not surprising, therefore, that the checklist approach taken in ScotCET failed, since it would have been neither sufficient to change officers’ minds about the need to behave in a procedurally fair manner, nor have provided an enforceable framework capable of overcoming such resistance (and even if it had, forcing officers to behave in a procedurally fair manner seems ethically troubling, if not oxymoronic). Notwithstanding the results of QCET, then, a tool more robust than a checklist may be required to motivate officers to behave in a more procedurally just manner in many or most policing contexts. In particular, training in how to handle encounters, and in the ethics of policing, appears vital if interventions such as that prescribed by ScotCET and the wider procedural justice literature are to be successful (Skogan et al. 2014; Wheller et al. 2013).

To move research in this field forward, and to establish what ‘good’ or ‘best practice’ models of policing and communication might look like, it appears critical to create a policing environment in which ideas from within and without policing organizations can be meaningfully tested. Police organizations can ensure officers deliver basic tasks (such as issuing leaflets and questionnaires) to at least some degree—but it is quite another matter to ensure investment in the more qualitative, less tangible elements of an intervention designed to alter and enhance common ways of seeing and policing citizens (see also Skogan 2008). Sufficient ‘buy-in’ from officers on the ground is critical, and coercive direction inadequate to achieve this. Achieving buy-in and fostering open environments conducive to new practice requires substantial change in internal, and external, approaches to undertaking and implementing policing research, experimental or otherwise. The importance of good communication, and fair treatment of officers within their own organization, cannot be over-estimated, and the value of investing in procedural justice oriented training, backed up with reinforcement and feedback by fellow and senior officers in terms of espousing the need for positive public engagement *and* ensuring internal organizational justice, is clear (see Skogan et al. 2014).

Fostering a greater openness to outsiders, academic and otherwise, and to evidence led knowledge is another critical step (Weisburd and Neyroud 2011), and the risks perceived by officers when testing alternatives to existing policing models (such as loss of professional control and competence, and the risk of criticism) must be addressed. As Gartin (1995) argues, implementation failures, like those experienced in ScotCET, often provide a useful means of exploring the potential impacts of future research and policy implementation. Failures and unintended consequences highlight the critical gaps between policy mandates and street-level operations. Our examination of a single experiment serves to highlight a series of internal and external shortcomings that acted as barriers to success. Addressing these shortcomings will be neither easy nor straightforward, requiring considerable investment of time and resources, but is an undeniably important endeavor.

References

- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Social Science and Medicine*, 75(12), 2,299–2,306.
- Bradford, B., Quinton, P., Myhill, A., & Porter, G. (2014). Why do the law comply? Procedural justice, group identification and officer motivation in police organisations. *European Journal of Criminology*, 11(1), 110–113.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Colquitt, J. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400.
- Cordner, G. (2000). Community policing: Elements and effects. In G. Alpert & A. Piquero (Eds.), *Community policing: Contemporary readings*. Prospect Heights: Waveland Press.
- Farrington, D. P., & Welsh, B. C. (2005). Randomised experiments in criminology: what have we learned over the last two decades? *Journal of Experimental Criminology*, 1(1), 9–38.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 1–11.

- Fielding, N. (1988). *Joining forces: Police training, socialization and occupational competence*. London: Routledge.
- Fielding, N. (1989). Police culture and police practice. In M. Weatheritt (Ed.), *Police research: Some future prospects*. Aldershot: Avebury.
- Folger, R., & Bies, R. J. (1989). Managerial responsibilities and procedural justice. *Employee Responsibilities and Rights Journal*, 2, 79–89.
- Gartin, P. R. (1995). Dealing with design failures in randomised field experiments: analytic issues regarding the evaluation of treatment effects. *Journal of Research in Crime and Delinquency*, 32(4), 425–445.
- Greenberg, J. (2011). Organisational justice: The dynamics of fairness in the workplace. In Z. Sheldon (Ed.), *APA handbook of industrial and organisational psychology* (Maintaining, expanding and contracting the organisation, Vol. 3). Washington DC: American Psychological Association.
- Haas, N. E., Van Craen, M., Skogan, W. G., & Fleitas, D. M. (2015). Explaining officer compliance: the importance of procedural justice and trust inside a police organisation. *Criminology and Criminal Justice*, 15(4), 442–463.
- Hassall, K. D., & Lovell, R. D. (2015). Fidelity of implementation: important considerations for policing scholars. *Policing and Society*, 25(5), 504–520.
- Hasson, H. (2010). Systematic evaluation of implementation and fidelity of complex interventions in health and social care. *Implementation Science*, 5, 67–76.
- Hohl, K., Bradford, B., & Stanko, E. A. (2010). Influencing trust and confidence in the London Metropolitan Police: results from an experiment testing the effect of leaflet-drops on public opinion. *British Journal of Criminology*, 50(3), 491–513.
- Lavie-Ajayi, M. (2014). I didn't know I was isolated until I wasn't isolated anymore: the use of groups in qualitative psychology research. *Qualitative Inquiry*, 20(2), 175–182.
- Loftus, B. (2009). *Police culture in a changing world*. Oxford: Oxford University Press.
- Loftus, B. (2010). Police occupational culture: classic themes, altered times. *Policing and Society*, 20(1), 1–20.
- MacQueen, S., & Bradford, B. (2015). Enhancing public trust and police legitimacy during road traffic encounters: results from a randomised controlled trial in Scotland. *Journal of Experimental Criminology*, 11(3), 419–443.
- Mazerolle, L., Bennett, S., Antrobus, E., & Eggins, E. (2012). Procedural justice, routine encounters and citizen perceptions of the police: main findings from the Queensland Community Engagement Trial (QCET). *Journal of Experimental Criminology*, 8(4), 343–366.
- Mazerolle, L., Bates, L., Bennett, S., White, G., Ferris, J., & Antrobus, E. (2014). Optimising the length of random breath tests: results from the Queensland Community Engagement Trial. *Australian and New Zealand Journal of Criminology*, 48(2), 256–276.
- Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015) Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 350(h1258).
- Murray, K. (2014) *Stop and search in Scotland: An evaluation of police practice*, SCCJR Research Report: University of Glasgow.
- Myhill, A., & Bradford, B. (2013). Overcoming cop culture? Organisational justice and police officers' attitudes toward the public. *Policing: An international Journal of Police Strategies and Management*, 36(2), 338–356.
- Reiner, R. (2010). *The politics of the police*. Oxford: Oxford University Press.
- Scottish Government. (2012). *The strategy for justice in Scotland*. Edinburgh: Scottish Government.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalised causal inference*. Boston: Houghton Mifflin Company.
- Sherman, L. W., & Strang, H. (2004). Experimental ethnography: the marriage of qualitative and quantitative research. *The Annals of the American Academy in Political and Social Science*, 595, 204–222.
- Skogan, W. G. (2008). Why reforms fail. *Policing and Society*, 18(1), 23–34.
- Skogan, W. G., Van Craen, M., & Hennessy, C. (2014). Training police for procedural justice. *Journal of Experimental Criminology*, 11(3), 319–334.
- Tankebe, J. (2010). Identifying the correlates of police organisational commitment in Ghana. *Justice Quarterly*, 13(1), 79–91.
- Terpstra, J., & Fyfe, N. (2015). Mind the implementation gap: police reform and local policing in the Netherlands and Scotland. *Criminology and Criminal Justice*, 15(5), 527–544.
- Trinkner, R., Tyler, T. R., & Goff, P. A. (2016). Justice from within: the relations between a procedurally just organizational climate and police organizational efficiency, endorsement of democratic policing, and officer well-being. *Psychology, Public Policy, and Law*, 22(2), 158–172.
- Tyler, T. R. (2006). *Why people obey the law*. Princeton: Princeton University Press.

- Tyler, T. R., & Blader, S. (2000). Procedural justice, social identity, and cooperative behaviour. *Personality and Social Psychology Review*, 7, 349–361.
- Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage.
- Waddington, P. A. J. (1999). Police (canteen) sub-culture: an appreciation. *British Journal of Criminology*, 39(2), 287–309.
- Weisburd, D. (2000). Randomised experiments in criminal justice policy: prospects and problems. *Crime and Delinquency*, 46(2), 181–193.
- Weisburd, D. (2005). Hot spots policing experiments and criminal justice research: lessons from the field. *The Annals of the American Academy in Political and Social Science*, 599, 220–245.
- Weisburd, D., & Neyroud, P. (2011). *Police science: Towards a new paradigm*. Cambridge: Harvard University <https://www.ncjrs.gov/pdffiles1/nij/228922.pdf>.
- Wheller, L., Quinton, P., Fildes, A., & Mills, A. (2013). *The Greater Manchester Police procedural justice training experiment*. London: College of Policing.
- Wilkinson, S. (1998). Focus groups in feminist research: power, interaction and the co-construction of meaning. *Women's Studies International Forum*, 21(1), 111–125.
- Wolfe, S. E., & Piquero, A. R. (2011). Organisational justice and police misconduct. *Criminal Justice and Behaviour*, 38, 322–353.
- Wünsch, D., & Hohl, K. (2009). Evidencing a “good practice model” of police communication: the impact of local policing newsletters on public confidence. *Policing: A Journal of Policy and Practice*, 3(4), 331–339.

Sarah MacQueen is a Research Fellow with the Scottish Centre for Crime and Justice Research, based at the University of Edinburgh Law School. Her most recent research explores the impact of the quality of interaction between police and citizens on public perception and judgements on procedural fairness and police legitimacy, testing the possibility of operationalizing the procedural justice framework in routine police encounters. She is also exploring the barriers and pitfalls in experimental policing research. Other research includes an examination of differential risk and experience of domestic abuse, and the ways in which particular groups of victims respond.

Ben Bradford is Departmental Lecturer in Criminology at the Centre for Criminology, University of Oxford. His research revolves around issues of procedural justice, trust, legitimacy, identity and compliance in criminal justice contexts, particularly in regard to the relationship between police and public. He has conducted studies on these and related issues in the UK, Australia, South Africa, Pakistan and other countries, and cross-national comparison is a growing aspect of his work. He also has an interest in organizational justice and the internal dynamics of police organizations, and in elements of operational policing, particularly stop and search, community patrol and other elements of ‘public-facing’ police work.