

6-14-2018

What Do We Know About In-service Police Training? Results of a Failed Systematic Review

Laura Huey

University of Western Ontario, lhuey@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/sociologypub>



Part of the [Sociology Commons](#)

Citation of this paper:

Huey, Laura, "What Do We Know About In-service Police Training? Results of a Failed Systematic Review" (2018). *Sociology Publications*. 40.

<https://ir.lib.uwo.ca/sociologypub/40>

What Do We Know About In-service Police Training? Results of a Failed Systematic Review

Author:

Laura Huey
Department of Sociology
University of Western Ontario
Room 5401, Social Science Centre
London, Ontario Canada N6A 5C2
Tel: 1-519-661-2111 ext. 87689
Fax: 1-519-661-3200
Email: lhuey@uwo.ca

Abstract:

To learn more about what the social scientific research literature can tell us about ‘what works’ in the field of in-service police training, the author attempted to conduct a systematic review of the recent published research on this topic (2000-2015). After initially narrowing the search results to 21 studies, the review had to be abandoned because there was an insufficient number of studies on any one topic or training technique. The author reflects on what this failed review means from the standpoint of the possible economic and social costs of potentially ineffective and inefficient in-service training.

Keywords: police, training, in-service, costs, effectiveness.

Following events in Ferguson, Missouri in August 2014, it would not be unfair to say that a significant amount of public commentary was focused on issues related to police use of force and the relationships of police agencies to the communities they serve. Attracting far less attention was the issue of police training and, in particular, the adequacy of current training practices and standards. One notable exception was a brief article that appeared in wired.com. In the context of discussing the Brown case, and the issue of addressing racial bias in policing through training, the author observed, “But what kind of training really works—and whether it can be designed to reduce the influence of racial biases—is [an] open question” (Zhang 2015). Others have also made similar claims. Indeed, it has been suggested that what could be reliably said about police training in this or any other context would not amount to much, because huge gaps exist in our knowledge of the short and long-term effects of various forms of police training due to the fact that training techniques, courses and/or whole programs are rarely subject to internal or independent evaluation (Mugford, Corey and Bennell 2011; Stanko and Dawson 2015).

Despite such dreary assessments, it was thought that, in light of the recent surge in interest in evidence based policing, we might see a greater interest by researchers in not only evaluating police training, but in evaluating modes of training delivery. Further, it was hoped that such evaluations would be made in light of two important considerations: effectiveness and efficiency (ie. ‘what works’ (Sherman 2013)). To that end, the author began the process of undertaking a systematic review of the literature on in-service police training – that is, post-recruit training for already sworn officers – for the period of 2000-2015. This review was intended to establish the following: a. the types of in-service training for which there were sufficient studies to form an evidence base; b. the training techniques that have been shown to be moderately effective, and; c. whether any of those techniques identified as moderately to consistently effective in meeting

course objective have been assessed for long-term effectiveness. As is discussed in the pages that follow, this systematic review failed. The reasons for this failure are provided, as is a discussion of why the lack of evaluative research on police training remains a critical public policy issue.

The state of interest in the evaluation of police training

Training, whether it be for the police or any other group, is only useful if it produces desired objectives, whether they be attitudinal, cognitive and/or behavioral outcomes. To understand whether a mode of training is both effective and efficient in producing those outcomes, police organizations must rely on internal or external evaluation. This is hardly a new insight: the first major evaluation of police training facilities came in 1931 with the Wickersham Report from the U.S. National Commission on Law Observance and Enforcement (Brereton 1961). Unfortunately, since Wickersham, there have been relatively few major reviews of police training and a lamentable dearth of either internal or independent evaluations of different types and forms of training. One recent study of scientific collaboration in police training for the period of 1987-2011 found that while the number of journal articles on police training had increased many of these were largely descriptive rather than evaluative, or, as the researchers put it: “The results show a lack of scientific productivity despite an increase in articles published on this subject in recent years, a finding indicative of the growing need for new police training research” (Aguilar-Moya, Melero-Fuentes, Aleixandre-Benavent and Valderrama-Zurián 2013: 698). This conclusion was shared by earlier work: “what is currently absent in the police training literature are concrete, empirically supported instructional strategies that can be incorporated into training to promote the long-term retention and transfer of learned skills and knowledge” (Mugford, Corey and Bennell 2011: 314). Not only do we know little about what works in the long-term, but we also know officers are

subject to training decay, which can significantly erode any skills or knowledge acquired (Snook, House, Macdonald and Eastwood 2012).

How to explain the lack of evaluation in such a critical area of policing? One readily apparent explanation is that it reflects the historical divide between police and researchers and a predominant view within policing that treats experience as vital to decision-making and often sees research as less useful or, worse yet, unnecessary (Bradley and Nixon 2009; Griffiths 2014; Telep and Lum 2015). This explanation echoes in recent comments by Stanko and Dawson (2015: 66) who view police training as “severely limited because it presumes it is training a craft,” therefore there is no need to draw on “knowledge that also exists in journals, systematic reviews and other scientific learning on policing.” A further explanation from the literature is more centred on practical concerns: that evaluation can be both expensive and disruptive. Thus, like colleges and universities, training evaluations are frequently limited to “‘smile sheets,’ the course evaluation forms that trainees complete at the end of instruction” (Glenn, Raymond et al. 2003: 62)

While it is the case that, relative to other areas of policing research, training has generated fewer evaluative studies, this is not to say that none exist, or that no previous attempts have been made to synthesize research findings. I was able to locate two reviews of police training. The first was Peter Neyroud’s (2011) review of police leadership, which specifically attempted a systematic review of the literature on police leadership training in order to distill some best practices under the banner of ‘what works.’ “The evidence base on effective training practice was very limited across all sectors,” Neyroud concluded, “[and] no systematic reviews in the policing context were found.” Thus, in order to answer the research questions posed, he drew instead from training in the fields of healthcare and education, suggesting that collaborative learning and training that integrates classroom and practice tend to be effective. A second study was conducted by Aguilar-

Moya et al. (2013) which used bibliometrics in order to determine the extent to which scientific research was being conducted in the area of police training during the period of 1987-2011. Unlike Neyroud's review, or the present study, the Aguilar-Moya study did not focus on synthesizing findings from evaluative studies, but instead sought to identify themes present in training research and which disciplines were contributing to that body of work. What these researchers observed is that the volume of research on training was not only diverse, but had been increasing over previous years (ibid.). Perhaps not surprisingly, they also identified 'skills' and 'management development' as "the most salient themes in the field of police training" (ibid.: 703).

Method of inquiry

Systematic review

To address the research questions posed in the introduction, the method selected was a systematic review (SR) of the research literature on police in-service training. A SR is a technique for "finding, sifting, sorting and synthesizing the findings of primary evaluations relevant to particular interventions" (Johnson et al. 2015: 460). In advance of beginning the finding phase of research, the researcher constructs a set of questions to be answered, as well as a set of well-defined inclusion and exclusion criteria (Akobeng 2005; Pawson 2006; Neyroud 2011). As Uman (2011: 57) explains, "Systematic reviews, as the name implies, typically involve a detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic." Although SRs may include a meta-analysis component, they differ in that the purpose of the systematic review is to summarize all empirical evidence that fits the inclusion criteria, whereas a meta-analysis entails the use of statistical methods to measure effect size by pooling results from multiple

primary studies (Huf, Kulcher, Pail, Friedrich, Filzmoser and Kasper 2011; Hofler and Hoyer 2014).

Despite the lamentable relative lack of available research in the area, this remains the most appropriate method for the task of improving our understanding of the state of the current research literature, as well as the extent to which researchers can abstract from this body of work any particular tools, techniques or principles that may be shown to have greater efficacy in the realm of police training or, failing that, that may be worth further study. To that end, I developed the following research questions to guide this effort:

1. What forms of in-service training have been empirically evaluated by researchers (by topic, technique)?
2. Beyond classroom lectures, can we identify any techniques (ie. training tools, teaching methods) from the research literature that are shown to be consistently effective, moderately effective or ineffective in meeting the objectives of police in-service training programs?
3. Of those techniques identified as moderately to consistently effective through systematic review, have they been assessed for long-term effectiveness?

Search strategy

Following standard SR practice, before commencing searches a specific set of inclusion and exclusion criteria were established (Akobeng 2005; Uman 2011). The inclusion criteria for this SR were: any study conducted on a specific form of in-service training (course or program) that used empirical techniques to measure the effects of that training on police officers. Articles excluded from the review were theoretical and/or descriptive only or measured attitudes towards the training rather than the effects of training itself.

In advance of our searches, I also identified which databases I would use and thus what types of research would be included (Akobeng 2005). Although an obvious preference would have

been to include only peer-reviewed research from academic journals, the lack of such articles in academic databases necessitated a decision beforehand to also include ‘grey literature’ – non-published studies from either independent researchers or internal evaluations within police organizations. Using ‘grey literature’ within SRs is not uncommon and, as in the instant case, may be necessary to achieving a larger number of studies (Savoie, Helmer, Green, Kazanjian 2003). To search for articles to include, two researchers independently searched for relevant studies from four sources. First, I looked for academic journal articles on police training using a library search engine against academic research databases (ie. criminal justice abstracts, sociological abstracts, etc.). In order to capture more recent evaluations, the search period was set at 2000-2015. Keywords used included: police, policing, training, education. Search engine results, which included both graduate theses and published research articles, were then assessed for possible inclusion based on whether the study reported contained results of an evaluation of an in-service course or programme of in-service training. Evaluative studies of in-service training were included regardless of research methodology, course content and/or instructional technique. To increase the number of studies for inclusion, I also searched for both academic and ‘grey literature’ using Public Safety Canada’s policing research portal, the University of Queensland’s Global Policing Database and Google’s search engine. Lastly, I turned to the Google search engine and, using the same key words, sought to identify any further published or unpublished studies. In total, fifty (n= 50) studies were initially collected. However, once I verified the topic and methodology used, and excluded studies that did not meet criteria set, the final number was of studies subject to initial coding – coding for topic and technique – was twenty-one (**n=21**).

Coding and analysis

The research questions posed require each of the **21** studies to be coded in two different ways. To answer the first question, concerning the in-service topics and technique(s) studied, I developed a simple inductive coding scheme while reading the studies for content. This allowed me to identify which topics received greater or lesser attention from researchers.

Coding studies to answer the second and third research questions required a more complex method. Further, as Lum et al. (2015: 3) recently observed, certain types of questions in a SR may require drawing on “many types of research.” This is particularly the case in relation to police training research; the **21** studies collected for this SR included diverse methodologies, from focus groups to pretest/posttest surveys. Therefore, what was required was a SR tool or technique that could be used to code findings irrespective of the methodology used in a given study. For this project, I chose one devised specifically for diverse methodologies, that was oriented towards evaluating research with a crime or crime prevention focus and that had been used by the U.K. College of Policing’s ‘what works’ program: EMMIE (Johnson et al. 2015).

While being able to determine effects across studies is important, an equally important consideration is the quality of the original researchⁱⁱ (Uman 2011). EMMIE is a rating system that allows researchers to look at both the reported effects of a study (‘the evidence’), as well as making an assessment of the quality (‘quality’) of that study based on standard methodological criteria. Its utility lies not only in the fact that it works for mixed methodological studiesⁱⁱⁱ, but it can be used to assess both individual studies as well as studies in aggregate. The EMMIE acronym refers to the core dimensions to be evaluated: effect, mechanism, moderator, implementation and economic (costs/benefits) (Johnson et al. 2015: 463). In short, with EMMIE, a study is assessed not only for methodological rigor, or to determine overall effects, but also for other considerations that are

likely of significant value to researchers, practitioners and policy-makers, such as the presence or absence of confounding variables and the economic or benefits costs of an intervention.

For each of the dimensions above, researchers are required to score a study from 0 to 5 using a scaling system intended to assess the quality of the study on a particular dimension. For example, 'implementation' requires raters to assess the quality of four study characteristics: 1. the description of the program of practice delivered; 2. the identification of facilitators and obstacles encountered during implementation; 3. description of key elements for a successful implementation, and; 4. specification of what would be required to replicate the intervention (Johnson et al. 2015). A score of 0 indicates a significant lack in relation to a particular category, whereas 5 indicates a significant degree of confidence that the study meets or exceeds expectations in that area (Johnson et al. 2015).

To answer the second research questions posed, each of the selected studies were assessed according to both two components of the EMMIE rating system. First, I examined the results of the study – the 'evidence' – in order to parse out any reported effects. In this case, I was looking at whether various forms of training were reported as successful and with what degree of confidence. Then each study was coded using the EMMIE-Q scoring system – which assesses the 'quality' of the study. For this step, I coded across only the first four^{iv} dimensions – E, M, M, and I – which allowed me to identify any training content, techniques, tools that showed both moderate to high effects and met a cut-off of 3 across each of the selected EMMIE-Q dimensions. The whole EMMIE coding process was greatly facilitated by the use of the EMMIE coding tool (Tompson, Bowers, Johnson and Belur 2015).

To address the third research question, I took any common interventions that showed significant positive effects *and* scored from 3 (moderate) to 5 (high) across a majority of the Q

criteria, and then re-examined the relevant studies to look at the length of the purported effects. In short, I looked at whether the study findings had measured short term effects only, or if there was a minimum follow up period of 6 months to a year.

Lastly, to ensure reliability in the coding for this project, and thus in the overall conclusions, all coding was independently verified by a member of my research team.

Results

Q1. What forms of in-service training have been empirically evaluated by researchers (by topic, technique)?

In relation to topic matter, I was left with 21 studies covering 11 different topics, from training on specific diversity issues to threat assessment courses. As can be seen in Table 1 below, four topics yielded more attention within the studies analyzed: interviewing techniques, diversity issues, mental health issues and domestic violence training. As can also be seen, a number of other important topics generated little to no attention from evaluative researchers. For example, not included among the remaining studies are topics such as general investigation or officer driving safety. Others, such as use of force training, were systematically evaluated within far too few studies (n=1).

Table 1: Training topics

Topic	Total studies	Author(s)/year
Interviewing	4	Smets & Pauwells (2010); Clark, Milne & Bull (2011); Snook, House, MacDonald & Eastwood (2012); Cederborg, Alm, Nises & Lamb (2013)
Mental health	4	Pinfold, Huxley, Thornicraft, Farmer et al. (2003); Mishara & Martin (2012); Krammedine, DeMarco, Hassel & Silverstone (2013); Herrington & Pope (2014)
Diversity issues	3	Bailey, Barr & Bunting (2001); Lamotte, Ouellette, Sanderson, Anderson et al. (2010); Israel, Harkness, Delucio, Ledbetter et al. (2014)
Domestic violence	3	Smithey, Green & Giacomazzi (2004); Blaney (2010); Ruff (2012)
Use of force	1	Armstrong, Clare & Plecas (2014)

Sex offenders	1	Darwinkle, Powell & Tidmarsh (2013)
Public health issues	1	Beletsky, Agrawal, Moreau, Kumar et al. (2012)
Procedural justice	1	Skogan, Van Craen & Henessy (2015)
Threat assessment	1	Storey, Gibas, Reeves & Hart (2011)
Crisis negotiation	1	Van Hasselt, Baker, Roman, Schlessinger et al. (2006)
Problem oriented policing	1	Quinet, Nunn & Kincaid (2003)

Whereas identifying course subject matter was relatively straightforward, I had greater difficulties with identifying specific training techniques analyzed within the studies examined. One problem was that in many instances researchers did not state what training techniques were used in their evaluation (see Table 2 below). Another problem observed was that several forms of police learning use multiple methods of teaching and student engagement, from assigning homework to employing interdisciplinary training teams to post-training coaching. Unfortunately, studies that did utilize multiple teaching methods did not typically parse out one or more techniques to evaluate from the others and so I do not know whether positive results were from one or more methods or a cumulative effect. To try to remedy this problem, what I have done is identify each technique from each study separately to see whether there were common methods across studies reporting significant positive effects.

Table 2: Techniques used

Technique used	Total studies	Author(s)/year
Practice exercises (including role play, mock interviewing, computer and other simulation exercises)	6	Van Hasselt, Baker, Roman, Schlessinger et al. (2006); Beletsky, Agrawal, Moreau, Kumar et al. (2012); Bailey, Barr & Bunting (2001); Krammedine, DeMarco, Hassel & Silverstone (2013);); Darwinkle, Powell & Tidmarsh (2013); Israel, Harkness, Delucio, Ledbetter et al. (2014); Armstrong, Clare & Plecas (2014)
Group work	5	Quinet, Nunn & Kincaid (2003); Lamotte, Ouellette, Sanderson, Anderson et al. (2010); Storey, Gibas, Reeves & Hart (2011); Israel, Harkness, Delucio, Ledbetter et al. (2014); Skogan, Van Craen & Henessy (2015)

Individual activities (personal reflections, completing assignments, etc)	5	Lamotte, Ouellette, Sanderson, Anderson et al. (2010); Storey, Gibas, Reeves & Hart (2011); Darwinkle, Powell & Tidmarsh (2013); Israel, Harkness, Delucio, Ledbetter et al. (2014)
Feedback and supervision on class exercises	4	Van Hasselt, Baker, Roman, Schlessinger et al. (2006); Cederborg, Alm, Nises & Lamb (2013); Krammedine, DeMarco, Hassel & Silverstone (2013);); Israel, Harkness, Delucio, Ledbetter et al. (2014)
Sequencing class days over time (6 months to 1 year)	2	Pinfold, Huxley, Thornicraft, Farmer et al. (2003); Cederborg, Alm, Nises & Lamb (2013)
Post-training coaching	1	Smets & Pauwells (2010)
Providing ‘take home’ information packs	1	Krammedine, DeMarco, Hassel & Silverstone (2013)
Team teaching	1	Israel, Harkness, Delucio, Ledbetter et al. (2014)
Assigning home work	1	Cederborg, Alm, Nises & Lamb (2013)
Instructor behavior modelling	1	Van Hasselt, Baker, Roman, Schlessinger et al. (2006)
No specific training technique identified	5	Smithey, Green & Giacomazzi (2004); Blaney (2010); Ruff (2012); Clarke, Milne & Bull (2011); Herrington & Pope (2014)

2. Using the EMMIE rating scale, can we identify any techniques (ie. training tools, teaching methods) from the research literature that are shown to be consistently effective, moderately effective or ineffective in meeting the objectives of police in-service training programs?

This question proved impossible to answer with the remaining studies. First, five (n=5) of them had to be removed from consideration because they failed to identify a training technique beyond conventional classroom lectures. Then, once one study (n=1) that identified only one technique was removed (Pinfold et al. 2003), only twelve (n=12) remained. There were, however, several problems with relying on these twelve to construct a systematic review – beyond the low number of possibly included studies. First, most of the studies (n=11) were not designed to test a particular technique. Instead, they were intended to assess whether a given training course produced a desired outcome, thus I had no means of assessing real or purported effects for any specific training method. Second, six (n=6) of the studies referenced the use of multiple training

techniques, but did not attempt to measure effects of individual training strategies, instead lumping them all together. Thus, there was no point in coding the one study (Armstrong et al. 2014) in which reassures measured the effects of a training intervention, particularly as in this instance they were measuring physiological effects on participants in simulation training and not training effects such as information retention and so on).

Q3. Of those techniques identified as moderately to consistently effective through systematic review, have they been assessed for long-term effectiveness?

The result discussed in relation to Question 2 above also meant that the third research question could not be answered. If there is an insufficient evidence base upon which to form a systematic review of ‘what works’ in police training, more generally, then there is clearly also an insufficient evidence base upon which to understand what might work in terms of producing long-term training effects.

Discussion

My attempt at conducting a systematic review of recent (2000-2015) research on in-service training failed. It failed, quite simply, because of a lack of peer-reviewed, published research in the area. Without a solid evidence base, there is little we can say, at least with any degree of confidence, concerning the efficacy of existing models of training for sworn police members.

Given the significant social issues with which researchers, practitioners and public policy-makers must grapple on a daily basis, there is, perhaps, an understandable tendency to shrug one’s shoulders and say ‘so what?’ There are at least two primary responses to this question: one is economic, the other social.

Most countries are currently experiencing a period of economic austerity at a time when public service costs are steadily rising. In response to growing awareness of shrinking public budgets and escalating service demands on police, policy-makers and police leaders have been

struggling with how to cut policing costs while maintaining current levels of service. In Canada, this quandary has generated renewed interest in looking more carefully at operational costs, leading to, among other outcomes, a recent analysis by Public Safety Canada of all training costs across Canadian police services, an estimated \$815 million annually (Maslov 2015). Comparable figures for the U.S., Australia and the U.K. could not be located^v; however, given relative population sizes, we can reasonably estimate that training costs will be higher in at least the U.S. and the U.K. compared to the figure for Canada. From an economic standpoint, it is somewhat disconcerting that significant public investments are made each year in this area when we know very little about ‘what works’ in police training, including how effectively training objectives are being met and what particular techniques work best (Neyroud 2011; Mugford et al. 2011; McKenna 2015; Stanko & Dawson 2015).

Lack of rigorous evaluation research on in-service training becomes an even more important issue when one considers the diverse array of social issues with which police face and to which we expect them to fairly, effectively and efficiently respond. These issues can range from new and increasingly deadly drugs hitting the streets (W-18 and fentanyl, to name two) to individuals in various states of crisis (with a range of emotional, physical, mental, communication and other issues) to a panoply of technologies and techno-aided social interactions that can generate safety hazards (‘driving while playing Pokemon’), as well as facilitate opportunities for criminal conduct (anything from ‘revenge porn’ and other forms of cyberbullying to online and other techno-assisted frauds). Each of these, and the myriad of others not mentioned here, not only create demands for services, but frequently real or perceived demands for police in-service training. As I write this, the Chicago Police is in the process of rolling out an ambitious plan to require each of its 12,500 sworn officers to attend a two day course on ‘de-escalation’ techniques

(Sweeney 2016). Clearly this training was developed in response to perceptions that a police legitimacy crisis could be brewing both in Chicago, and in the wider U.S., following a series of high-profile police shootings. While laudably being implemented as acknowledge of a problem needing to be addressed, as is often the case, there has been no public discussion of plans to independently evaluate the training program to ensure effective delivery, and thus that it meets organizational objectives. Without quality control in the form of rigorous, independent evaluations, such ‘fixes’ run the risk of being little more than ‘Band-Aid’ solutions. Further, to the extent that police agencies frequently borrow ideas, strategies and programs from each other that may appear to be unsuccessful, we run the risk of replicating programs that are, from an empirical standpoint, untested. Given the social implications of failures in this area, this is not a risk that policy-makers, practitioners and other civic stakeholders should want to take. And yet, as we see from the failure of the present study, it happens.

Academic papers invariably end with an exhortation for more research. Thus it feels a bit trite to offer this very same conclusion. And yet, in the instant case, there can be no other recommendation. I do, however, offer some thoughts on how future researchers might transcend some of the limitations of the present study. These future researchers, who will hopefully be blessed with a greater volume of studies upon which to draw, will find ways to improve upon my methodology by using, as an example, a better search strategy or different rating tools. I admit that the rating scale I attempted to employ was thorough, but overly complex for the task. Similar work could likely be done using any number of other rating schemes. There is also likely some repositories of police research that were missed, despite best efforts, and thus some potential studies that might have been included but were not located. If the search strategy described above

can be improved upon, this would greatly help advance the cause of understanding ‘what works’ in the field of in-service training.

References

- Aguilar-Moya, R., Melero-Fuentes, D., Aleixandre-Benavent, R. and Valderrama-Zurián, J. 2013. ‘Production and scientific collaboration in police training (1987-2011)’ *Policing: An International Journal of Police Strategies and Management*, 36(4): 768-786.
- Akobeng, A. 2005. ‘Principles of Evidence Based Medicine.’ *Archives of Disease in Childhood*, 90(8): 837-840.
- Alpert, G., Dunham, R., and Strohshine, M. 2006. *Policing: Continuity and Change*. Waveland, Long Grove, Ill.
- Arthur, W., Bennett, W., Stanush, P. and McNelly, T. 1998. ‘Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis.’ *Human Performance* 11(1):57-101.
- Bintliff, R. 1990. *Training Manual for Law Enforcement Officers*. Englewood Cliffs, NJ.: Prentice-Hall.
- Bradley, D. and Nixon, C. 2009. ‘Ending the “Dialogue of the Deaf”: Evidence and Policing Policies and Practices. An Australian Case Study.’ *Police Practice and Research*, 10(5–6): 423–435.
- Brereton, G. 1961. ‘The Importance of Training and Education in the Professionalization of Law Enforcement.’ *Journal of Criminal Law and Criminology*, 52(1): 111-121.
- Chan, J., Devery, C. and Doran, S. 2003. *Fair Cop: Learning the Art of Policing*. University of Toronto Press, Toronto.
- Cordner, G. 2014. *Police Administration*, 8th ed. Waltham, Mass.: Elsevier.
- Etter, G. and Griffin, R. 2011. ‘In-service training of older law enforcement officers: an andragogical argument.’ *Policing: An International Journal of Police Strategies & Management*, 34(2): 233 – 245.
- Glenn, R., Raymond, B., Barnes-Proby, D., and Williams, E. et al. 2003. *Training the 21st Century Police Officer: Redefining Police Professionalism for the Los Angeles Police Department*. Santa Monica, CA: RAND Corp.
- Griffiths, C. 2014. ‘Economics of Policing: Baseline for Policing Research in Canada.’ Report Commissioned by Public Safety Canada. Available online at: <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/bsln-plcng-rsrch/index-eng.aspx>

- Griffiths, C., Parent, R., & Whitelaw, B. 2001. *Community Policing in Canada*. Scarborough, ON: Nelson Thomson Learning.
- Hofler, M. and Hoyer, J. 2014. ‘Population size matters: bias in conventional meta-analysis.’ *International Journal of Social Research Methodology*, 17(6): 585–597.
- Huf, W., Kalcher, K., Pail, G., Friedrich, M-E., Filzmoser, P. and Kasper, S. 2011. ‘Meta-analysis: Fact or fiction? How to interpret meta-analyses.’ *The World Journal of Biological Psychiatry*, 12(1): 188–200
- Jewell, L. 2013. Models and Structures of Corrections and Police Training and Research Activities in Canadian and International Jurisdictions. Report for the Centre for Forensic Behavioural Science and Justice Studies, University of Saskatchewan. Available at: http://www.usask.ca/cfbsjs/research/pdf/annual_reports/CFBSJS%20Annual%20Report%202013-14.pdf
- Johnson, S., Tilley, N., and Bowers, K. 2015. “Introducing EMMIE: an evidence rating scale to encourage mixed-method crime prevention synthesis reviews.” *Journal of Experimental Criminology*, 11(4): 459–473.
- Litmanovitz, Y and Montgomery, P. 2012. “Training programs to improve attitudes, behaviours and skills of police officers regarding democratic values, civil rights, and human rights and to reduce unlawful use of force by police officers.” Report of the Campbell Collaboration. Available at: <http://www.campbellcollaboration.org/lib/project/201>
- Lum, C., Koper, C., Merola, L., Scherer, A., Reieux, A. 2015. ‘Existing and Ongoing Body Worn Camera Research: Knowledge Gaps and Opportunities: A Research Agenda for the Laura and John Arnold Foundation (Phase I Report).’ Available at: <http://cebcp.org/wp-content/technology/BodyWornCameraResearch.pdf>
- Mugford, R., Corey, S., and Bennell, C. 2013. ‘Improving police training from a cognitive load Perspective.’ *Policing: An International Journal of Police Strategies & Management*, 36(2): 312 -337.
- Neyroud, P. 2011. Review of Police Leadership and Training. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/118222/report.pdf
- Nijkerk, K. 1982. ‘The Development of Training and the Need for In-Service Training,’ in J. Yuille (ed.) *Police Selection and Training: The Role of Psychology*. Boston, Mass.: Martinus Nijhoff Publishers. Pp. 43-56.

- Pawson, R. 2006. 'Digging for Nuggets: How 'Bad' Research Can Yield 'Good Evidence.' *International Journal of Social Research Methodology*, 9(2): 127-142.
- Perrier, D. 1979. 'Is Policing a Profession?' *Canadian Journal of Criminology*, 21(1): 52-70.
- Savoie, I., Helmer, D., Green, C., Kazanjian, A. 2003. 'Beyond Medline: reducing bias through extended systematic review search.' *International Journal of Technology Assessment in Health Care*, 19(1): 168-78.
- Schulenberg, J. and Warren, D. 2009. 'Content and Adequacy of Specialized Police Training to Handle Youth-Related Incidents Perceptions of Trainers, Supervisors, and Frontline Officers.' *International Criminal Justice Review*, 19(4): 456-477.
- Sherman, L. 2013. 'The Rise of Evidence-Based Policing: Targeting, Testing, and Tracking.' *Crime and Justice*, 42(1): 377-451.
- Sloane, C. 1951. 'Police Professionalization.' *Journal of Criminal Law and Criminology*, 45(1): 77-79.
- Snook, B., House, J., Macdonald, S., and Eastwood, J. 2012. 'Police Witness Interview Training, Supervision, and Feedback: A Survey of Canadian Police Officers.' *Canadian Journal of Criminology and Criminal Justice*, 54(3): 363-372.
- Stanislas, P. 2014. 'The Challenges and Dilemmas Facing University-based Police Education in Britain' in P. Stanislas (ed.) *International Perspectives on Police Education and Training*. London: Routledge. Pp. 57-71.
- Stanko, E. and Dawson, P. 2015. *Police Use of Research Evidence: Recommendations for Improvement*. New York: Springer.
- Telep, C., and Lum, C. 2014. 'The Receptivity of Officers to Empirical Research and Evidence-Based Policing: An Examination of Survey Data From Three Agencies.' *Police Quarterly*, 17(4): 359-385.
- Tompson, L., Bowers, K., Johnson, S., Belur, J. 2015. 'EMMIE evidence appraisal coding tool.' [Dataset]. UCL Department of Security and Crime Science: London, UK. Available at: <http://discovery.ucl.ac.uk/1462093>
- Uman, L. 2011. 'Systematic Reviews and Meta-Analysis.' *Canadian Academy of Child and Adolescent Psychiatry*, 20(1): 57-59.
- Vollmer, A. 1932. 'Abstract of the Wickersham Police Report.' *Journal of Criminal Law and Criminology*, 22(5): 716-723.

White, S. 1972. 'A Perspective on Police Professionalization.' *Law & Society Review*, 7(1): 61-86.

Endnotes

ⁱ Research reports not found within mainstream academic journals and databases, such as results of studies produced within police agencies or public or other commissioned reports.

ⁱⁱ Some researchers, notably Pawson (2006) have, however, argued for the inclusion of both methodologically rigorous and methodologically weak studies on the ground that 'bad' research can yield useful insights. This view appears to remain a minority opinion.

ⁱⁱⁱ While some researchers conducting SRs eschew any study that is not a randomized control trial (see Pawson 2006), doing so is not feasible in this instance given the lack of RCTs in this area.

^{iv} I did not assess studies in relation to whether they address real or potential economic costs/benefits for two reasons. First, none of the studies looked at these issues. Second, they were not relevant to the research questions posed.

^v I could locate state/provincial or county estimates for recruit training, but no figures that would include in-service training and there was no way to reliably combine individual amounts across all areas to create a national estimate.