



# Measuring Police Integrity: Futile Exercise or Worthwhile Effort in Personnel Management? Revisiting Survey Data from Two Previous Studies in Order to Assess the Psychometric Qualities of the Klockars Questionnaire

Marc Alain<sup>1</sup>  · Michel Rousseau<sup>1</sup> · Francesco Carrer<sup>2</sup>

Published online: 26 July 2018  
© Society for Police and Criminal Psychology 2018

## Abstract

In this paper, we attempt to determine if the psychometric features of the 11-scenario survey designed by Klockars et al. provide valid results for survey responses concerning police integrity and ethics. To do this, we used the multi-facet Rasch statistical model on two samples from police respondents in two different countries (one in North America and the other in Continental Europe). Our results demonstrate that the survey, when combined with a specific statistical model, was able to assess the level of integrity of police officers in these two different samples. We conclude that the 11-scenario survey and the methodology we have developed can be considered very robust and is therefore useful for managerial purposes, particularly in candidate selection and promotion.

**Keywords** Police integrity · Police management · Police ethics · Police surveys

## Introduction

Police integrity and ethics have been, are, and probably always will be the subject of a significant debate that raises important questions: (1) what are police integrity and police ethics and why should we try to assess and measure these concepts? (2) If these concepts are measurable, how can they be assessed and what is the relative value of the available tools and methods? These questions are the founding blocks of the sociology of policing pioneered by Egon Bittner (1970). Questions about integrity and ethics in policing began as the consideration of philosophical and legal principles surrounding police interactions in modern societies (Kleinig 1996; Kadish 1962) and have gradually changed to concern operational and measurable elements of police training (Terrill 2001), police supervision, (Engel and Worden 2003; More and Miller 2014), and police officers' duties and rights

(Stoebuck 1980; Torres 1996; Karkkainen 2005). Following Ivković (2015), we propose that police integrity and ethics should be understood not as abstract elements of police work but as operational and empirically measurable concepts.

## Police Ethics: A Short Historical Perspective

Corruption and abuse of power involving police officers in both our English-American adaptations of the Peel model and their Continental European counterparts are probably as old as these systems themselves (Haller 1971, 1996; Walker 1976). Fogelson (1977) shows how North American police organizations first attempted to deal with such problems through reforms and the progressive adoption of standardized professional norms. However the Serpico affair, which led to the Knapp commission in 1972, clearly established that these so-called professional norms were neither strong enough nor sufficiently integrated among individual members to prevent the corruption of whole precincts in the NYPD. The publication of the Knapp commission report introduced the idea of “rotten apples,” police officers who avoided official control measures and policies to achieve their own interests. The 1972 report also introduced the first ranking of breaches of police ethics by referring to low-level corrupt officers as “grass-

---

✉ Marc Alain  
Marc.alain@uqtr.ca

<sup>1</sup> Department of psychoeducation, Université du Québec à Trois-Rivières, 850, avenue de Vimy, porte 10 C.P. 32, Québec, QC G1S 0B7, Canada

<sup>2</sup> Genova, Italy

eat-ers” and more highly corrupt officers as “meat-eaters.” Not until 22 years later, in 1994, with the publication of the Mollen Commission report, which also dealt with NYPD police officers, was the idea of rotten apples abandoned in favor of looking at the barrel itself as a source of rot. Punch (2003), reflecting on the Dutroux scandal in Belgium, went even further in proposing a “rotten orchard,” suggesting that a systemic emphasis on the whole organization is necessary to understand ethical problems in police organizations.

This being said, however, the challenge of changing, superficially or profoundly, the police professional culture remains nowadays not completely addressed. Studying this professional culture reveal two important characteristics that might impede planned changes. Often referred to as “canteen culture” (Waddington 1999; Kingshott and Prinsloo 2004), police professional culture contains both negative and positive aspects. As Van Hulst (2013) shows, if canteen culture can be quite effective at pushing aside the recruits’ high sense of morale in favor of what can be understood as negative behaviors (accepting gifts and privileges because of being a police officer, for example), it also have important potential positive effects. Kingshott and Prinsloo (2004) were quite revealing in showing that canteen culture constitutes a very important mechanism of forging and maintaining professional bonds between officers, these bonds constituting essential reinforcement of the necessary mutual trust among those who choose to “serve and protect.” The mere presence of this fundamental professional culture, then, must be counted for and evaluated for any police organization wishing to go through reforms in favor of higher levels of integrity among its ranks, a task for which our own researches might eventually contribute.

Recommendations for how to assess police ethics and integrity have always been part of such inquiries and reports (Chappell and Piquero 2004; Skolnick 2002). How to undertake such assessments remains, however, the subject of a still-evolving sphere of police sociology. Both the Knapp and the Mollen commissions (as well as the Christofer Commission on the events related to the beating of Rodney King in Los Angeles in 1991) identified the “blue code of silence” (also referred as the “blue curtain”) as the greatest obstacle to ethical assessment of any police organization (Neyroud and Beckley 2001). The problem of serious bias that this code introduces into any attempt to assess the levels of conformity to ethical rules by asking police officers for their opinions has been thoroughly addressed by Westmarland (2005) in an article that stresses the need for further investigation into police ethics and integrity. Simply asking officers about the conduct of their fellow officers generally results in answers that bear little resemblance to events related in the media:

Investigations into police corruption, lying and brutality have uncovered the most striking examples of the Code.

Absent the inquiry, the powerful mandate of silence usually lies dormant, almost like an incubating disease. In such situations, the Code manifests itself as a refusal to offer information that might incriminate or embarrass a fellow officer. ... The Blue Code is not restricted to law enforcement in the United States; its existence has been documented in police departments in other westernized countries. — Skolnick 2002: pp. 8–9

Investigating levels of ethical commitment among the officers of a given police organization by analyzing reports of breaches, corruption, or unjustified use of force does not necessarily provide better results. As Ivkovic et al. (2012; see also Ivković 2015) note, it is not clear that police misconduct reports are useful in measuring the general level of ethics within any given police agency as the number of incidents of misconduct or their reported seriousness may be deceiving. An increase in the number of reported incidents could be caused by tougher policies against police deviance while a decrease in seriousness might be the result of more efficient cover-up strategies by either corrupt officers or their administration (Punch 2003). Neither would necessarily involve any change in the underlying level of ethics. Ivkovic and Haberfeld (2015b) also discuss how difficult it is to use inquiry commission reports as a basis for determining the level of ethics in a police department. These inquiries are rare and in some countries, such as Region B (Carrer and Alain 2011, 2012; Carrer 2004), do not take place even when problems of corruption are clearly apparent. As well, careful examination of all the incidents that led to the inquiry is a difficult and time-consuming task for any researcher (Ivkovic et al. 2012, p. 150). Researchers as well as administrators have therefore adopted another approach to assessing levels of ethical commitment in police organizations: surveys.

### The Use of Surveys for Assessing Ethical Commitment among Police Officers

In the USA, using surveys to explore police officers’ attitudes and integrity began with Milton Rokeach (Rokeach et al. 1971) who, with his team, attempted to determine whether there were differences in values and attitudes between police officers (Caldero 1997). In the same period, police managers began to accept the idea that surveys of officers in their organizations could be useful in understanding how officers responded to change, such as the shift from traditional policing to the community policing model (Walker and Katz 2012; Klockars et al. 2007; Greene et al. 1994). However, such surveys are usually undertaken for managerial purposes and are not available to the public or scientists. As well, because they serve particular managerial purposes, it cannot be assumed that they represent the values and attitudes of a true random sample of a given population of police officers.

There are further problems in trying to assess the ethical commitment of police officers, challenges well described by Ivkovic:

Consistent in these studies is the phenomenon that whenever direct questions about police misconduct are posed, be it about police corruption, use of excessive force, or police testifying, the researchers are bound to experience similar forms of opposition: police administrators are reluctant to open their doors to researchers raising questions about police misconduct, possibly fearing that any misconduct uncovered will be interpreted negatively for the administration/agency; police officers fearing ostracism from their colleagues if they reveal anything about the misconduct of their fellow officers or fearing disciplinary and/or criminal consequences if their own misconduct is uncovered. In addition, typical witnesses and victims of police misconduct, such as prostitutes, drug dealers, and other career criminals, may not be credible witnesses in court. — Ivković 2015, p. 2)

The problem of the code of silence was addressed by Weisburd et al. (2001) in their study of the attitudes of US police officers regarding abuse of authority. Their survey of 925 randomly selected American police officers in 121 departments confirmed that police officers' attitudes regarding abuse of authority are clearly influenced by the code of silence. Respondents who believe that abuses of authority are related to a "tough on crime" attitude also think that, in some instances, such abuses are permissible and should not be reported (Weisburd et al. 2001).

Trying to assess ethical commitment by asking police respondents to react to even hypothetical questions about police misconduct remains a major challenge, in part because ethics is never completely black or white. The ethical commitment, however defined, depends largely on individual philosophy, whether this philosophical position is derived from Kant's logical propositions or a more humanistic position of care (Noddings 1984; Gilligan 1982). These two ethical standpoints separate officers who favor a "crime-fighter" attitude from those who believe actions should arise from a more caring and protecting attitude (Fitzgerald et al. 2002; Neyroud 2003). While it might be expected that police organizations will include officers who endorse both of these roles, circumstances in a particular police action might mean that police officers who support one role would find themselves working with officers who support the other, with each group proposing different responses. The presence of conflicting views on the role of the police increases the problem of assessing any given ethical standpoint within a whole police department.

These difficulties led to the work of Carl Klockars and his team, who in 1994 suggested that the idea of police integrity should be used instead of police ethics. According to these authors, referring to integrity resolves most of the obstacles involved in asking police officers to talk about their colleagues' misconduct:

Whereas the respondents were generally eager to respond to most questions, none was willing to provide any examples of graft or corruption in their police agencies. In 1994, Klockars and Kutnjak Ivković developed an alternative methodology to the study of police misconduct, an approach that does not face such serious challenges because it measures its opposite—police integrity. ... Klockars et al. define police integrity as 'the normative inclination among police to resist temptations to abuse the rights and privileges of their occupation.' — Klockars et al. 2007

### Exploring the Psychometric Qualities of a Widely Used Instrument

Our study used the survey tool developed by Klockars et al. (Klockars and Ivkovic 1999; Klockars et al. 2003), which has been used in comparative studies in several countries (Ivkovic et al. 2016; Ivkovic and Klockars 1995; Klockars et al. 1997; Punch et al. 2000). This survey asks respondents to assess 11 hypothetical cases of police misconduct with regard to the seriousness of each case as well as the degree to which they would be willing to report such conduct.

The Klockars questionnaire is aimed at investigating practices (to report or not to report a colleague's unethical behavior, to comply or not to comply with the employer's integrity policy, and so on) that must be dealt with by police officers across the world, however different their cultural environment. With a few notable exceptions (see, for instance, Hickman et al. 2016; Brown and Reed Benedict 2002; Hickman et al. 2004), the 11-vignette questionnaire has been empirically validated as able to overcome the specificities of national/regional police organizational cultures (Ivkovic and Haberfeld 2015a) by being submitted to more or less the same methodological treatment: the results obtained through national or regional samples are presented in ranked order (see, for example, Westmarland 2005), which makes it possible to determine if sample responses are similar to responses from one of the original US samples (as is done in Klockars et al.'s edited 2004 work, *The contours of police integrity*).

## Contributing to Further Enhancing the Validity of the 11 Vignette Questionnaire

In this article, we establish the psychometric qualities of Klockars' 11-vignette questionnaire and explore different ways in which answers can be combined to establish respondent profiles on a unique scale. Establishing the psychometric qualities of the 11-vignette questionnaire makes possible additional analyses, such as more precise assessments of integrity levels in sub-cultures within one police agency (patrolmen versus detectives, recruits versus experienced officers, women versus men, and so on). We expect the results of such work to both further enhance the theoretical value of Klockars' work and help police managers to tailor use of this questionnaire to their particular needs. The treatment of the questionnaire's answers proposed here is of benefit to both the science of police integrity assessment and the managerial use of the questionnaire. In the first instance, we are confident that the establishment of the psychometric potential and qualities of the Klockars questionnaire will open new paths for further explorations of both old data sets—as is done here—and new data, as well as, more importantly, making possible more robust treatment of the data being collected. And, since we should be able to obtain more precise measurements of answers provided by police officers, police managers will be in a much better position to rely on these measurements in dealing with levels of integrity in the organizations they manage.

## Surveying Police Ethics in Two Regions from Two Different Countries, Field Work, and Contexts

In the case of region A—in North America—a large police organization felt that important reforms involving ethics and integrity were needed in both the police code of conduct and the continuous training officers receive. In early 2002, we were asked to conduct a survey to assess these elements, not only among a representative sample of officers from this organization but also among officers from other jurisdictions in the same region to provide some basis for comparison. A comparable situation occurred few years later, in June 2006, when police management in region B—in continental Europe—asked us to conduct the same kind of assessment before they began an in-depth management reform. In these two instances, we were able to provide precise assessments of base rates of ethics and integrity.

Police organizations in both regions are responsible for providing services to important cities as well as more rural communities and villages. The distribution of population in cities, rural communities, and villages is quite similar in both regions.

In both regions, organizations leaders had asked for such an evaluation in order to have a clear picture of existing levels of staff integrity and ethical commitment before beginning planned reforms. Our survey was intended as a first step before reforms were undertaken at all levels of the police organization, beginning with basic training, hiring standards and procedures, rank upgrade practices, and career evaluations.

### Region A Sample

The total police force in region A is estimated to be nearly 15,000 men and women (both part- and full-time employees). Once permissions were granted and cooperation attained, 600 questionnaires were distributed to police managers for further distribution to individuals in their organization. Each questionnaire was accompanied by a letter, signed by both the director of the police organization and the director of the police union that explained the purpose of the survey and guaranteed the complete anonymity of responses and respondents' identities.<sup>1</sup> The questionnaires were distributed by research assistants to a sample created in a collaboration between researchers and managers to make sure that quotas for rank, duties, and number of years' experience for the entire police population would be reached. In each precinct, respondents were invited by their supervising officers (who were generally at the first level of the organizations' management ladder) to complete the response sheet at their convenience and then put it in a sealed cardboard box to be picked up by the research assistants. Respondents were clearly told in the accompanying letter that they were absolutely free to respond or not to the questionnaire. The whole operation was conducted over a 4-month period; of the 600 questionnaires initially distributed, 455 were returned, a response rate of 75.8%. Chi-square test computed on the respondents' distribution on the three stratification criteria (rank, duties, and number of years of experience) showed non-significant differences (Alain 2003, 2004).

### Region B Sample

At the time of the study (June of 2006), authorities in region B were embarking on a vast project to reorganize the police to allow them to operate in a more coordinated fashion. This plan had repercussions for the command structure, basic and continuous training of police officers, and prevention and

<sup>1</sup> At the time of this first study, as well as before the second study conducted in region B, the two police organizations did not ask neither us nor any other authorities to provide an ethical certificate delivered by a recognized ethical board. At that time, the principal researcher was head of the research center of the only police academy from his region, where no procedure whatsoever regarding these precautions were in place. It must be said, however, that in Canada all three major research policies specifically mention that ethical certification is not required when a research is conducted under administrative request, which was the case for both regions A and B.



treatment of police misconduct.<sup>2</sup> We were approached to provide information and base rates to give the management team a point of reference before they began the projected reform. Relying on the procedure tested in region A (a letter signed by the police and the union directors and emphasis that respondents could answer at their convenience and at will), 2000 questionnaires were distributed by research assistants throughout the region (which had a total of 13,000 officers), and 1520 valid and completed questionnaires were returned, a response rate of 76%; in this case, the operation took place over a period of 6 months. Relying on the same three stratification criteria, data showed a small but significant overrepresentation of ranked officer among the respondents.<sup>3</sup>

## A Short Presentation of the Survey Tool

As noted above, the 11 scenario survey tool developed by Klockars et al. has been thoroughly used across the world and has been the subject of at least two edited books (Klockars et al. 2003; Ivkovic and Haberfeld 2015a) and a complete issue of *Policing: An International Journal of Police Strategies and Management* (May 2016). The survey presents 11 vignettes, each of which describes a story that is considered to be, in the majority of the countries where it has been used, police misconduct; these vignettes are presented in the table below (Table 1).

According to Klockars et al. (2003), none of stories of misconduct involves extreme behavior. That is, none is so benign or so dramatic that all respondents would put it in roughly the same place on a ranked scale (e.g., totally wrong or totally benign). Klockars' team also attempted to produce culturally neutral scenarios:

Those who study instances of police abuse of authority for gain encounter their first problem in attempting to achieve cultural neutrality in the different meaning of money in different cultures. A bribe of U.S.\$50 clearly has a different meaning for a U.S. police officer who earns \$40,000 per year and a Eastern European police officer who earns the annual equivalent of U.S.\$4,000. ... In the scenarios we created that involved monetary gain, we attempted to resolve this problem by expressing the gain in terms of some local value equivalent. — Klockars et al. 2003, p. 8.

<sup>2</sup> As we write this article, these planned reforms have not yet been implemented, more than 10 years after the survey operation.

<sup>3</sup> This question of over/under representation of specific quotas in the two samples must be understood here as irrelevant, since our purpose is not to assess how the integrity score might differ from one strata to another, but merely to verify the psychometric qualities of the tool based on responses of the whole two samples.

Scenario 1 can, however, be seen as culturally biased, since it involves a police practice found in the USA but rarely elsewhere, at least officially: “a police officer runs a private business in which he sells and installs security devices ...”. For our study, this specific element did not cause any discrepancies since the laws that govern police activities and duties in both regions specifically prohibit any police officer from engaging in private business related to private security.

For each of the 11 scenarios, respondent were asked to reply to the same set of seven questions, which can be divided into three different but interrelated themes. The first theme involves the idea of seriousness, the second involves the question of disciplinary measures, while the third one explores willingness to report the depicted behavior. These seven questions are outlined in the next table (Table 2).

The questionnaire was pretested among small groups of volunteers in police organizations in both regions A and B to make sure that all vignettes and assessment questions were easily understood and unambiguous. Before discussing the data analysis and methodological assessment we propose, we should point out again that the importance of revisiting two sets of older data should be judged in relation to our contribution to the development of a survey tool in ways that have, as far as we know, never been previously used. Our work is important for its methodology and explorative value in providing a psychometric evaluation of this survey tool.

## Data Analysis

The basic assumption in psychometric modeling is that the concept being measured, in this case police integrity, is distributed on a continuum that goes from a very small to a very high level of integrity. The goal of the model is therefore to provide scores, based on respondent answers to different items of a survey tool that represent the integrity level of that specific respondent. Psychometric models are then used to compute the score and are also used to assess the quality of the survey tool.

Analysis of the survey tool was conducted using the Item Response Theory (IRT) measurement modeling (Bond and Fox 2001). More specifically, the Rasch one-parameter logistic (1PL) model was chosen and is set out in the following equation:

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} \quad i = 1, 2, \dots, n.$$

The Rasch model specifies that the probability ( $P_i$ ) of a response to a question (in this case 1 or 0 in a dichotomous survey tool) is dependent on the level of integrity of the respondent (called ability parameter or  $\theta$ ) and the level of integrity needed for a respondent to endorse an item (called

**Table 1** Case Scenario

Case 1.	A police officer runs a private business in which he sells and installs security devices, such as alarms, special locks, etc. He does not work during his off-duty hours.
Case 2.	A police officer routinely accepts free meals, cigarettes, and other items of small value from merchant on his beat. He does not solicit these gifts and is careful not to abuse the generosity of those who give gifts to him.
Case 3.	A police officer stops a motorist for speeding. The officer agrees to accept a personal gift of half of the amount of the fine in exchange for not issuing a citation.
Case 4.	A police officer is widely liked in the community, and on holidays local merchants and restaurant and bar owners show their appreciation for his attention by giving him gifts or food and liquor.
Case 5.	A police officer discovers a burglary of a jewelry shop. The display cases are smashed and it is obvious that many items have been taken. While searching the shop, he takes a watch, worth about two days’ pay for that officer. He reports that the watch had been stolen during the burglary.
Case 6.	A police officer has a private arrangement with a local auto body shop to refer the owners of cars damaged in the accidents to the shop. In exchange for each referral, he receives a payment of 5% of the repair bill from the shop owner.
Case 7.	A police officer, who happens to be a very good auto mechanic, is scheduled to work during coming holidays. A supervisor offers to give him these days off, if he agrees to tune-up his supervisor’s personal car. Evaluate the supervisor’s behavior.
Case 8.	At 2 A. M. a police officer, who is on duty, is driving his patrol car on a deserted road. He sees a vehicle that has been driven off the road and is stuck in a ditch. He approaches the vehicle and observes that the driver is obviously intoxicated. He also finds that the driver is a police officer. Instead of reporting this accident, he transports the driver to his home.
Case 9.	A police officer finds a bar on his beat which is still serving drinks a half hour past his legal closing time. Instead of reporting this violation, the police officer agrees to accept a couple of free drinks from the owner.
Case 10.	Two police officers on foot patrol surprise a man who is attempting to break into an automobile. The man flees. They chase him for about two blocks before apprehending him by tacking him and wrestling him to the ground. After he is under control both officer punch him a couple of times in the stomach as punishment for fleeing and resisting.
Case 11.	A police officer finds a wallet in a parking lot. It contains the amount of money to a full day’s pay for that officer. He reports the wallet as lost property, but keep the money for himself.

difficulty parameter or  $b_i$ ) (Engelhard 2013). A subject with a low level of integrity will be less likely to endorse items that require the respondent to have a high level of integrity. The respondent’s level of integrity (ability parameter) and level of integrity needed for an item to be endorsed (difficulty parameter) are on the same scale of value and therefore the values of those two parameters can be directly compared. A subject will be more likely to endorse items whose difficulty parameters on the level of integrity scale are lower than the values of their ability parameters. The scale of values provided by the model is an equal-interval scale with a mean of zero and a standard deviation of 1. Those values are determined before the analysis based on either mean or standard deviation of the distribution of items or respondents. Given this relation, the model could be restated:

$$\log \left[ \frac{P_{ni}}{1-P_{ni}} \right] = B_n - D_i$$

where,

- $P_{ni}$  the probability that respondent  $n$  will give an answer of 1
- $B_n$  the ability parameter of the respondent
- $D_i$  the difficulty parameter of item  $i$ .

Ordinal scales, like the Likert scale, can also be analyzed with this model using the Rating scale model. The Rating scale model adds contrast to the dichotomous model that represents the likelihood that the answer of a respondent will be observed in a category of the Likert scale compared to the

preceding category of answers. The model is then represented by the following equation:

$$\log \left[ \frac{P_{nik}}{P_{ni(k-1)}} \right] = B_n - D_i - F_k$$

where,

- $P_{nik}$  the probability that respondent  $n$  will give an answer of  $k$  for an item  $i$
- $P_{ni(k-1)}$  the probability that respondent  $n$  will give an answer of  $k - 1$
- $B_n$  the ability parameter of the respondent
- $D_i$  the difficulty parameter of item  $i$
- $F_k$  the difficulty of the answer option  $k$  compared to answer option  $k - 1$

In the dichotomous or Rating scale models, endorsement of an item option depends only on the ability parameter of the respondent and the difficulty parameter of the item. Typical psychometric analysis, including factor analysis, Cronbach’s alpha, and the Rasch model, relies on the assumption of local independence. This assumption specifies that endorsement of an item option for a specific item depends only on the ability parameters of the subject.

The survey tool we analyzed could not meet the local independence assumption because the same group of items is repeated across 11 possible situations. An answer to an item regarding a specific situation could not only be dependent on that situation but also on answers concerning the same item in

**Table 2** Case scenario assessment options

1. How serious do YOU consider this behavior to be?				
Not at all serious		Very serious		
1	2	3	4	5
2. How serious do MOST POLICE OFFICERS IN YOUR AGENCY consider this behavior to be?				
Not at all serious		Very serious		
1	2	3	4	5
3. Would this behavior be regarded as a violation of official policy in your agency?				
Definitely not		Definitely yes		
1	2	3	4	5
4. If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think SHOULD follow?				
1. NONE		4. PERIOD OF SUSPENSION		
2. VERBAL REPRIMAND		5. DEMOTION IN RANK		
3. WRITTEN REPRIMAND		6. DISMISSAL		
5. If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think WOULD follow?				
1. NONE		4. PERIOD OF SUSPENSION		
2. VERBAL REPRIMAND		5. DEMOTION IN RANK		
3. WRITTEN REPRIMAND		6. DISMISSAL		
6. Do you think YOU would report a fellow police officer who engaged in this behavior?				
Definitely not		Definitely yes		
1	2	3	4	5
7. Do you think MOST POLICE OFFICERS IN YOUR AGENCY would report a fellow police officer who engaged in this behavior?				
Definitely not		Definitely yes		
1	2	3	4	5

other situations. Classic Rasch measurement models could therefore not be applied to this kind of data. The multi-facet Rasch model (MFRM) has been developed precisely to deal with this kind of measurement situation. The model is called multi-faceted because it can deal with a more complex testing design. In a more classic situation, there are only respondents and items—a two-facet design. In the situation discussed here, there are three coexisting facets: the respondents, the 11 scenarios, and the 7 assessment items.

The MFRM model is represented by the following equation:

$$\log \left[ \frac{P_{nij(k)}}{P_{nij(k-1)}} \right] = B_n - D_i - C_j - F_k$$

where,

$P_{nij(k)}$  the probability that respondent  $n$  will give an answer of  $k$  for an item  $i$  to a situation  $j$ ,

$P_{nij(k-1)}$  the probability that respondent  $n$  will give an answer of  $k - 1$  for an item  $i$  to a situation  $j$ ,

$B_n$  the ability parameter of the respondent  
 $D_i$  the difficulty parameter of item  $i$ .  
 $C_j$  the difficulty parameter of situation  $j$ , and  
 $F_k$  the difficulty of the answer option  $k$  compared to answer option  $k - 1$ .

The MFRM model logistically transforms the respondents' ordinal answers on the Likert scale to an equal-interval logit scale of measures. When a multi-faceted analysis is conducted, the various facets are analyzed simultaneously and independently and calibrated onto a single linear scale that represents the continuum of the level of integrity measured by the instrument. This approach unites all facets of the measurement design on a single scale of values, which makes it possible to look at individual-level effects of the facets (how the respondents, the situations, and the items in the analysis performed together) and to look at the interaction between the three facets of the design (at what level of integrity respondents, situations, and items were located on the assessment scale). The analyses were all conducted using Facets software version 3.71.4.

### Misfit Evaluation

The Facets software has different fit indices, which provide information on how well the data fit the Rasch model. The appearance of large differences between the observed and the expected answers (expressed as standardized residuals) suggests surprising or unexpected results. Unexpected results are considered to be detrimental to the measurement process and for the instrument's psychometric qualities. The residuals are transformed into mean-square (MnSq) error statistics called outfit MnSq and infit MnSq. The outfit MnSq are unweighted MnSq residual statistics and are less sensitive to outliers. The expected value of the MnSq is 1 and the range goes from 0 to infinity. An item with MnSq value of between 0.5 and 1.5 is usually considered to be adequate for the measurement. An MnSq with a value less than 0.5 indicates little variation in the pattern of answers while an MnSq with a value greater than 1.5 shows a set of answers with one or more unexpected or surprising answers that do not seem to fit with the other answers. The present article presents results for the final set of data, once it had been cleared of misfit cases.

### Results

Three-facet MFRM analyses were run separately for region A and B samples. For the region A data, 30,030 potential answers were observed. Missing responses (where respondents had failed to provide answers to specific items (the seven questions) for specific situations (the 11 vignettes)) accounted for 4.9% of total answers ( $n = 1467$ ). For the region B sample, there were 100,320 complete answers with 16.1% ( $n =$

16,207) missing. Since the model can estimate latent traits without complete data, we decided not to impute missing data or to remove subjects with incomplete answers. After first analysis, an examination of MnSq outfit and infit was conducted and item 3 of the instrument (“Would this behaviour be regarded as a violation of official policy in your agency?” with answers ranging from 1 “Definitely not” to 5 “Definitely yes”) was identified as a major source of noise in the model with a majority of unexpected answers linked to this particular item. We decided to remove it and the new analysis produced no MnSq outfit and infit with values higher than 1.5.

### Fitting Information and Performance for the Scenarios Facet

Table 3 presents goodness-of-fit information and the location of the difficulty parameter for each scenario on the scale of integrity. The location of the difficulty parameter provides information on the level of integrity associated with the scenarios. A scenario with a negative and rather low value of difficulty parameter is a scenario for which a respondent does not need a high level of integrity to think that there is some problem with the situation presented in the scenario, while a scenario with a positive and high value needs a higher level of integrity for a respondent to express that there is a problem associated with the scenario. For both region A and B samples, no scenario shows too high or low infit and outfit MnSq values. For both samples, situation 1 shows the highest level of disturbance, meaning that more unexpected values were observed for this situation than for the others. For region A, there is good variability in the value of the difficulty parameter, the range going from -2.06 (situation 5: “A police officer discovers a burglary of a jewelry store ...”) to 1.80 (situation 1: “A police officer runs a private business in which he sells and installs security devices ...”) with scenarios found at every level of the continuum of the integrity scale. Variability of the value of the difficulty parameter is lower along the continuum for the region B sample. The range goes from -1.09 (situation 5) to 1.28 (situation 4: “A police officer is widely liked in the community ...”). For this sample, we also observed a good distribution of the values along the scale. Since the parameters of the MFRM are calibrated separately for each sample, the location of difficulty parameter values cannot be compared directly but comparison of the relative position of each situation is possible. For both samples situation 5 showed the lowest value on the location parameter. This means that for both groups respondents with low levels of integrity, understand that this scenario raises an ethical problem. Situations 3, 6, and 11<sup>4</sup> are also associated with a low

<sup>4</sup> Scenario 3: “A police officer stops a motorist for speeding ...”; scenario 6: “A police officer has a private arrangement with a local auto body shop ...”; scenario 11: “A police officer finds a wallet in a parking lot ...”

**Table 3** Location parameters and goodness-of-fit indices for the scenario facet by region A and B samples

Scenario number	Region A			Region B		
	<i>b</i>	Infit	Outfit	<i>b</i>	Infit	Outfit
Scenario 1	1.80	1.43	1.47	0.92	1.28	1.51
Scenario 2	1.25	0.91	0.92	0.50	0.98	1.00
Scenario 3	-1.12	1.00	0.93	-1.03	1.10	0.95
Scenario 4	1.14	0.92	0.94	1.28	1.12	1.21
Scenario 5	-2.06	1.20	0.89	-1.09	1.09	0.95
Scenario 6	-0.62	0.88	0.86	-0.72	0.92	0.86
Scenario 7	0.10	1.25	1.31	-0.08	1.07	1.06
Scenario 8	0.59	1.06	1.07	0.57	0.97	0.97
Scenario 9	-0.13	0.83	0.83	0.14	0.77	0.77
Scenario 10	0.06	0.89	0.87	0.16	1.06	1.06
Scenario 11	-1.01	-1.01	0.92	-0.62	0.95	0.94

value on the difficulty parameter (e.g., these scenarios are more clearly associated with ethical problems). On the other side of the continuum, scenarios 1, 2, 4, and 8 are those which only respondents with the highest levels of integrity consider to be ethically problematic. It is worth noting that the relative positions of each scenario are quite similar for both region A and B samples. Overall, results suggest that this survey tool is useful to measure levels of integrity of police officers, especially considering that the two samples are from very different types and traditions of police organizational culture.

### Fitting Information and Performance for the Items Facet

Goodness of fit and values of the location parameters are presented in Table 4. Once item 3 was removed, no further items show too high or too low values of MnSq infit or outfit. All items show goodness-of-fit values close to their expected value of 1. The variability of their difficulty parameter on the integrity scale is lower for both samples than what was observed for the scenarios facet. For region A respondents, the range goes from -0.59 (item 1: “How serious do YOU consider this behaviour to be?” with answers ranging from 1 “Not at all serious” to 5 “Very serious”) to 0.61 (item 7: “Do you think MOST POLICE OFFICERS IN YOUR AGENCY would report a fellow officer who engaged in this behaviour?” with answers ranging from 1 “Definitely not” to 5 “Definitely yes”) and from -0.67 (item 1) to 0.60 (item 7) for the region B respondents. For both regions A and B, the relative position of the difficulty parameter of each item is almost the same: item 1 is associated with the lowest level of integrity while item 7 is associated with the highest. This means that for item 1, it takes only the lowest level of integrity to tell that there is a



**Table 4** Location parameter and goodness-of-fit indices for the items facet by region A and B samples

Item number	Region A			Region B		
	<i>b</i>	Infit	Outfit	<i>b</i>	Infit	Outfit
Item 1	−0.59	1.08	0.91	−0.67	1.04	0.97
Item 2	−0.25	0.91	0.86	−0.45	0.92	0.89
Item 4	0.15	0.77	0.80	0.01	0.94	0.98
Item 5	−0.07	1.00	1.04	0.16	1.04	1.06
Item 6	0.14	1.30	1.19	0.35	1.12	1.14
Item 7	0.61	1.11	1.20	0.60	1.05	1.11

problem related to the specific item, while item 7 picks out the respondents who show the highest level of integrity.

### Fitting Information and Performance of the Likert Scales

The different ways respondents used the Likert scales to provide their answer in the 11 vignettes evaluation sheet was also analyzed and Table 5 presents information regarding these scales. The table indicates that most respondents favored option 4 or 5—62% in region A and 52% in region B. If the Likert scales are functioning as intended, the average measures should increase in their location parameter as the category of the Likert scale increases. That is the case here, suggesting that respondents with higher scores on an item are exhibiting a higher level of integrity than those with lower scores on that item. Outfit MnSq should be interpreted like the other goodness-of-fit measures, where values closer to 1 mean that the answers expected by the MFRM for the rating category are close to what is observed. This element also reveals the general quality of the instrument, as the response scale works as the model predicts. As illustrated in Table 5, the outfit MnSq are all very close to 1, suggesting that the Likert scales work very well for both samples.

### Variable Maps for Regions A and B

One of the most interesting features of the MFRM is its capacity to present every facet on the same scale of values—the level of integrity—producing a clear representation of the relationships between all the facets being considered by the model. Figure 1 presents the variable map for the region A sample, while Fig. 2 presents the same figure for the region B sample. The first column of the variable map displays the scale of integrity, which is an equal interval scale ranging from −3 to 4, on which all facets were positioned. The second column displays the integrity level of respondents, ordered from higher levels of integrity at the top to lower levels at the bottom. As indicated, an asterisk represents 6 respondents for the region A sample and 20 respondents for the region B sample. (This difference in magnitude is related to the different sizes of the samples; unfortunately Facets software version 3.71.4 does not make it possible to use an asterisk to represent the same number of respondents in the two variable maps). For both variable maps, a dot represents fewer respondents than an asterisk. The third column displays the scenario facets, ordered from the highest levels of integrity a respondent needs to consider it as problematic at the top to the lowest levels at the bottom. Each number represents one of the scenarios, from vignette 1 to vignette 11. The fourth column shows the items facet-ordered with the highest level of integrity needed to see it as problematic at the top and the lowest at the bottom. The last column displays the 5-point Likert scale that respondents used for every item on every situation. The horizontal broken lines in this column represents the scale category threshold. The threshold is defined by the point at which the likelihood of one respondent choosing the next higher category on the rating scale is equal to the likelihood of this respondent choosing the next lower category. For example, in region A, respondents with scores between 0.3 and 1.9 were the most likely to give an answer of 4, while those with scores higher than 1.9 were the most likely to give an answer of 5. This means that respondents with levels of integrity higher than 1.9 will more probably answer 5 than 4 on the items.

**Table 5** Percentage of use, measure, and goodness-of-fit indices for the categories of the rating scale for regions A and B

Response labels*	Region A			Region B		
	% of use	Average measure	Outfit	% of use	Average measure	Outfit
1	10	−0.81	1.3	16	−0.72	1.2
2	10	−0.35	1.1	14	−0.27	0.9
3	17	0.16	0.9	18	0.19	0.8
4	26	0.93	0.9	13	0.63	0.8
5	36	1.92	1.0	39	1.26	1.2

\*All labels were normalized and, when necessary, inverted to reflect higher levels of integrity as the scale gets higher

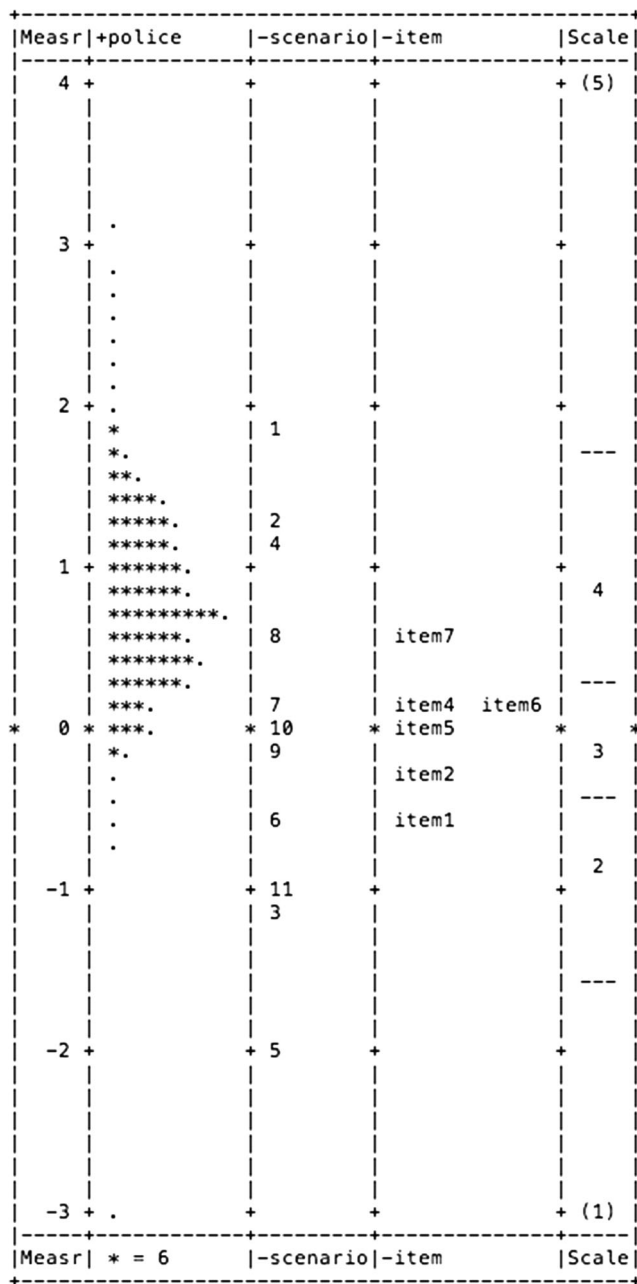


Fig. 1 Variable map of facets for region A

Ideally, there should be alignment between facets: that is, there should be no scenario or item for which there are no respondents at the same level of the scale. If this occurs, it could be explained by the inclusion of scenarios that are too “easy”—scenarios for which all respondents give the highest category of response for most of the items. The same is true for “too hard” situations, where all respondents give the lowest category of possible responses for most of the items. For both samples, vignettes 3, 5, 6, and 11 are associated with discrimination at the lowest level of the scale and are linked to a very small number of respondents. Those situations seem less useful for purposes of measurement because they

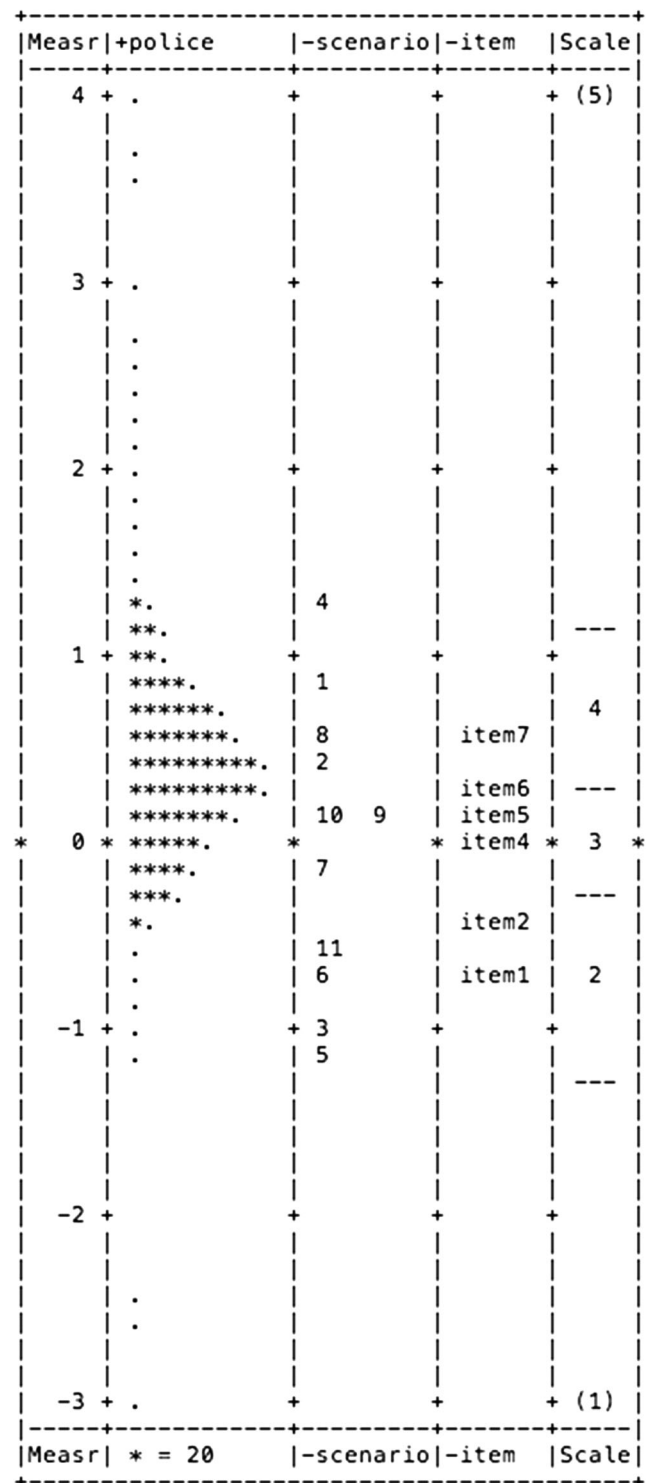


Fig. 2 Variable map of facets for region B

discriminate at scale levels for which we found almost no respondents. Items 1 and 2 show the same problem. It is worth noting that for both regions A and B, there was a percentage of respondents for whom we found no items associated with their level of integrity. This means that for this specific group of respondents, levels of integrity measured by the instrument

are higher than the item with the highest level of integrity (item 7). In these instances, social desirability may provide a possible explanation for such an answering pattern, a feature that can also be understood as yet one more quality of the statistical model we propose here. Identifying those respondents who are doing their best to “cheat” the questionnaire might be considered, in strict managerial terms, a rather interesting opportunity. Perhaps, future researches may involve controlling for social desirability in order to better assess how and to what extent this phenomenon might be associated with the tool’s global scores.

Finally, the separation index of the region A sample is 3.65 with a reliability of 0.93. For region B, the separation index is 3.46 with a reliability of 0.92. The separation index indicates how many measurement strata could be statistically distinguished on the instrument, indicating that it could help to distinguish different sub-groups based on their level of integrity. For regions A and B, we can distinguish between 3 and 4 different strata. Note that the notion of reliability differs from the classical approach: in the case of the MFRM, the reliability of the scale relies on the reliability of the separation index.

## Discussion

While the two samples differ considerably, both in terms of numbers of respondents ( $n = 455$  for region A;  $n = 1520$  for region B) and the moments the survey has been administered (in 2002 for region A and in 2006 for region B), it remains that the ranking of the 11 scenarios according to the measured levels of integrity is very similar for both samples. In other words, while it is likely that police culture in regions B and A differ significantly (if only, for instance, because one is in North America and the other in Continental Europe), the survey tool and the type of response analysis proposed here seem to be capable of comparing the respondents’ levels of integrity regardless of national professional police cultures. In other words, while it remains inevitable that any police organization will reflect the nationwide culture it polices, our work clearly reaffirm the existence of a specific police professional culture that transcend nations and continents. While the majority of respondents in both samples favored answers 4 and 5—thus choosing higher integrity levels—this survey tool and the type of response analysis we propose here still makes it possible to clearly position these same respondents along the integrity scale. This attribute is to be understood as yet another trait of robustness of the survey tool and the analysis model. At least one item of the survey, however, shows results that seemed to be far outside the general pattern: item three (“Would this behaviour be regarded as a violation of official policy in your agency?” 1 “Definitely not” to 5 “Definitely yes”). In this case, variance is very small as most respondents

chose 4 and 5; as mentioned earlier, this item was so problematic it had to be removed from the general model.

Two items on the questionnaire also indicate the same tendency for both samples—item 7 (colleagues propensity to report wrongful behavior) and item 1 (respondents own assessment of the seriousness of behavior).<sup>5</sup> In both samples, item 7 is associated with higher levels of integrity, while item 1 is associated with lower levels on our general integrity scale. In these cases, it takes a high level of integrity to think that colleagues might report the behavior depicted in the scenarios. And it takes a lower degree of integrity to consider the behavior as serious. We are faced here with two dimensions of the conceptualization of wrongful behavior by police officers: while a respondent might easily assess such behavior as wrong, it is quite another matter to express, in the same survey, a clear willingness to report the behavior, and even less evident to think that a colleague might also report it.

## Conclusion

Getting back to our introductory questions, we were able, first, to demonstrate that integrity is measurable and, second, that surveys are a practical way of obtaining such measurements. And, since the psychometric qualities of the 11-scenario questionnaire discussed here were used with two different samples from very different professional cultures, we are confident that the tool could be used for managerial purposes, provided that future users are prepared to move beyond looking at simple individual scores compared to averages or at standard deviation coefficients.

Our analysis revealed the survey tool’s capacity to distinguish and discriminate between respondents according to their level of integrity on the scale proposed by the MFRM model. The survey performs especially well for respondents reporting lower levels of integrity. This quality seems, in a managerial sense, ideal, since police organizations are not only looking for candidates showing high levels of integrity but also, and perhaps more importantly, looking to eliminate candidates who perform poorly on integrity tests.

This being said, work remains to be done and supplementary analysis might be conducted to identify interpretation norms (score distribution in a population of police respondents). In addition, supplementary analysis might reveal cutoff points in the scale, which would make it possible to identify with greater precision those respondents with a problematic level of integrity related to police work. Given that we are now in a position to use a very reliable instrument, as long as its

<sup>5</sup> Item 7: “Do you think MOST POLICE OFFICERS IN YOUR AGENCY would report a fellow police officer who engaged in this behaviour?”; answer scale goes from 1 “Definitely not” to 5 “Definitely yes”. Item 1: “How serious do YOU consider this behaviour to be?”; answer scale goes from 1 “Not at all serious” to 5 “Very serious”.

output is processed through the kind of statistical model we present in these pages, the next task will be to assess integrity in our own data base (and perhaps other data bases from elsewhere around the world, thus starting yet another true comparative effort), this time testing for potential differences between strata: male versus female police officers, patrol officers versus investigators and managers, and so on. This should provide material for future publications. By doing so, we clearly expect to be in a better position to shed new lights on the practical utility of the Klockars 11 vignette questionnaire to reveal the integrity levels of police officers.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no competing interests.

**Ethical Approval** None required (this has been a research conducted under administrative requests, as specified in the article).

**Informed Consent** Informed consent was obtained from all participants included in the study.

## References

- Alain M (2003) An exploratory study of quebec's police officers attitudes toward ethical dilemmas. In Klockars C. ed. *The contours of police integrity*, Thousand Oakes, Sage Publication
- Alain M (2004) Une mesure de la propension des policiers québécois à dénoncer des comportements dérogatoires, éléments de culture policière et cultures organisationnelles. *Déviance et Société*, 28(1): 2–32
- Bittner E (1970) *Functions of the police in modern society*. US Dept of health, education, and welfare, center for studies of crime and delinquency, Washington D.C. 122p
- Bond TG, Fox CM (2001) *Applying the Rasch model: fundamental measurement in the human sciences*. Lawrence Erlbaum Assoc, Mahwah
- Brown B, Reed Benedict W (2002) *Perceptions of the police: past findings, methodological issues, conceptual issues and policy implications*. *Policing: An Int J of Police Strategies & Management* 25(3): 543–580
- Caldero M (1997) *Value consistency within the police: the lack of gap*. Paper presented at the (March) 1997 annual meetings of the academy of criminal justice sciences, Louisville, Kentucky
- Carrer F (2004). *Insécurité urbaine et politiques d'interventions: la réalité italienne*. *Déviance et Société* 4:463-485
- Carrer F, Alain M (2011) *L'analisi sul campo, i dati della ricerca sulla Polizia local toscana*, in (sous la direction de F. Carrer et M. A. Seniga), *L'Etica della Polizia :Teoria e Pratica*, Maggiore Editore, Santarcangelo di Romagna pp. 133–166
- Carrer F, Alain M (2012) *Spécificités et constantes de la culture professionnelle policière: une mise en relief de l'opinion de policiers de la région toscane*. *Criminologie* 45(1):153–180
- Chappell A, Piquero A (2004) *Applying social learning theory to police misconduct*. *Dev Beh* 25(2):89–108
- Engel RS, Worden RE (2003) *Police officers' attitudes, behavior, and supervisory influences: an analysis of problem solving*. *Criminology* 41(1):131–166
- Engelhard G (2013) *Invariant measurement: using Rasch models in the social, behavioral, and health sciences*. Routledge, New York
- Fitzgerald M, Hough M, Joseph I, Qureshi T (2002) *Policing for London*. Willan, Cullompton
- Fogelson R (1977) *Big-city police*. Harvard University Press, Cambridge
- Gilligan C (1982) *In a different voice*. Harvard University Press, Cambridge
- Greene JR, Bergman WT, McLaughlin EJ (1994) *Implementing community policing: cultural and structural change in police organizations*. In: Rosenbaum DP (ed) *The challenge of community policing: testing the promises*. Sage, Thousand Oaks, pp 92–109
- Haller MH (1971) *Civic reformers and police leadership: Chicago, 1905–1935*. In: Hahn MH (ed) *Police in urban society*. Sage Publications, Oakland, pp 49–50
- Haller MH (1996) *Historical roots of police behaviour: Chicago, 1890–1925*. In: Brandl SG, Barlow DE (eds) *Classics in policing*. Anderson publishing co, Cincinnati, pp 7–25
- Hickman MJ, Piquero AR, Greene J (2004) *Police integrity: exploring the utility of a risk factor model*. In: Hickman, Piquero AR, Greene J (eds) *Police integrity and ethics*. Wadsworth/Thompson Learning, Belmont, pp 67–83
- Hickman MJ, Piquero AR, Powell ZA, Greene J (2016) *Expanding the measurement of police integrity*. *Policing: An Int J of Police Strategies & Management* 39(2):246–267
- Ivković SK (2015) *Studying police integrity*. In *Measuring police integrity across the world*. Springer, New York, NY, pp. 1–36
- Ivkovic SK, Haberfeld MR (eds) (2015a) *A comparative perspective on police integrity, measuring police integrity across the world*. Springer-Verlag, New York, pp 329–368
- Ivkovic SK, Haberfeld MR (eds) (2015b) *Measuring police integrity across the world: Studies from established democracies and countries in transition*. Springer, New York
- Ivkovic SK, Klockars CB (1995) *Cross-cultural study of police corruption: perceptions of offense seriousness; police perceptions of disciplinary fairness and code of silence*. Bureau of Justice Assistance, United States
- Ivkovic SK, Haberfeld M, Peacock R (2012) *Rainless west: the integrity survey's role in agency accountability*. *Pol Quart* 16(2):148–176
- Ivkovic SK, Haberfeld M, Kang W, Peacock R, Sauerman A (2016) *A multi-contry comparative study of the perceived police disciplinary environments*. *Policing: An Int J of Police Strategies & Management* 39(2):338–353
- Kadish SH (1962) *Legal norms and discretion in the police and sentencing processes*. *Harvard Law Review* 75(5):904–931
- Karkkainen BC (2005) *The police power revisited: phantom incorporation and the roots of the takings muddle*. *Minn L Rev* 90:826
- Kingshott B, Prinsloo J (2004) *Universality of the "police canteen culture"*. *Acta Criminologica* 17(1):1–16
- Kleinig J (1996) *The ethics of policing*. Cambridge University Press
- Klockars CB, Ivkovic SK (1999) *The measurement of police delinquency*. In: Laufer W, Adler F (eds) *The criminology of criminal law*. *Advances in criminological theory*, vol 8. Transaction Publishers, New Brunswick, pp 87–106
- Klockars CB, Ivkovic SK, Harver WE, Haberfeld MR (1997) *Measurement of police integrity, final report*. National Institute of Justice, United States
- Klockars CB, Ivkovic SK, Haberfeld MR (2003) *The contours of police integrity*. Sage, Thousand Oakes
- Klockars CB, Ivković SK, Haberfeld MR (2007) *Enhancing police integrity*. Springer, New York, pp 251–270
- More HW, Miller LS (2014) *Effective police supervision*. Routledge
- Neyroud P (2003) *Policing and ethics*. In: Newburn T (ed) *Handbook of policing*. Cullompton, Willan
- Neyroud P, Beckley A (2001) *Policing, ethics and human rights*. Willan, Cullompton



- Noddings N (1984) *Caring: a feminine approach to ethics and moral education*. University of California Press, Berkeley
- Punch M (2003) Rotten orchards: “pertilence”, police misconduct and system failure. *Polic Soc* 13(2):171–196
- Punch M, Huberts L, Lamboo T (2000) Perceptions on Integrity of Dutch Police Officers in Comparative Perspective. IIPE conference on ethics in the new Millenium: Ottawa
- Rokeach M, Miller M, Snyder J (1971) The value gap between police and policed. *J Soc Issues* 27(2):155–171
- Skolnick J (2002) Corruption and the blue code of silence. *Police Pract Res* 3(1):7–19
- Stoebuck WB (1980) Police power, takings, and due process. *Wash Lee L Rev* 37:1057
- Terrill W (2001) *Police coercion: application of the force continuum*. LFB Scholarly Publishing LLC, New York
- Torres G (1996) Taking and giving: police power, public value, and private right. *Envtl L* 26:1
- Van Hulst M (2013) Storytelling at the police station, the canteen culture revisited. *Br J Criminol* 53:624–642
- Waddington PAJ (1999) Police (canteen) sub-culture, an appreciation. *Br J Criminol* 39(2):287–309
- Walker S (1976) The urban police in American history: a review of literature. *J Police Sci Adm* IV(Sept):252–260
- Walker S, Katz CM (2012) *Police in America*. McGraw-Hill
- Weisburd D, Greenspan R, Hamilton EE, Bryant KA, Williams H (2001) *The abuse of police authority: a National Study of Police Officer’s Attitude*. Police foundation, Washington
- Westmarland L (2005) Police ethics and integrity: breaking the blue code of silence. *Polic Soc* 15(2):145–165