Measuring Public Perceptions of the Police

Article *in* Policing An International Journal of Police Strategies and Management · November 2010

DOI: 10.1108/13639511011085097

2 authors:

Edward R Maguire
Arizona State University
92 PUBLICATIONS **2,321** CITATIONS

SEE PROFILE

Devon Johnson
George Mason University
25 PUBLICATIONS **935** CITATIONS

SEE PROFILE

Measuring public perceptions of the police

Edward R. Maguire
Department of Justice, Law and Society, American University, Washington, DC, USA, and
Devon Johnson
Administration of Justice Department, George Mason University, Manassas, Virginia, USA

Received 18 June 2009
Revised 21 May 2010
Accepted 2 June 2010

Abstract

Purpose – The purpose of this paper is to test Mastrofski's six-dimensional conceptualization of perceived service quality of the police.

Design/methodology/approach – The paper draws on data collected from a mail survey of residents in a suburban Virginia (USA) community and uses confirmatory factor analysis to test the six-dimensional model of service quality.

Findings – The six-dimensional model does not fit the data. Instead a one-factor model fits the data, suggesting that public perceptions of police service quality are a one-dimensional construct.

Research implications – Further research is necessary on two fronts. First, research should seek to determine whether the findings are consistent across different community contexts and types of samples. Second, research on related perceptual constructs, including procedural justice and legitimacy, should explore measurement properties in more detail.

Originality/value – This is one of few studies to examine the construct validity of public perception measures in police research.

Keywords Police, Attitudes to the police, Measurement, United States of America

Paper type Research paper

1. Introduction

Policing is a service industry. Those who request assistance from police are its clients, as are those who are involuntarily subjected to police authority. These clients have varying opinions about the quality of the service they receive from police. Tapping into clients' perspectives of police is vital for understanding a fundamental element of the relationship between citizens and the state – the tenor and character with which formal social control and regulatory authority is imposed and perceived. While

The authors are grateful to Megan Gantley and Brittany Davenport for their assistance in preparing the manuscript and to participants in the 2002-2003 Administration of Justice Honors Seminar at George Mason University for their assistance in designing the survey instrument and collecting/cleaning the data used in this paper. They include: Roberta Stiftar, Kathryn Simmons, M. Christine Ibar, Brian Alsbaugh, and Shelane Bryant. Thanks also to the police officers in Orangetown (a pseudonym) who participated in the development of the survey instrument, to the police administrators who allowed us to carry out the research, and to Stephen Mastrofski and Cameron McIntosh for providing comments on an earlier draft.

Policing: An International Journal of Police Strategies & Management
Vol. 33 No. 4, 2010
pp. 703-730

1363-951X
DOI 10.1108/13639511011085097

applying the customer service perspective to the police has some limitations, it also offers a number of benefits[1].

The research outlined in this paper sought to specify and test a multidimensional conceptual model of police service quality. Based on the large body of research on service quality, most of which has developed in industries other than policing, Mastrofski (1999) articulated six dimensions of service quality in policing:

(1) attentiveness;
(2) reliability;
(3) responsiveness;
(4) competence;
(5) manners; and
(6) fairness.

We developed a survey instrument based on Mastrofski's conceptualization of police service quality and administered it to citizens in one community who had recent police contact. We then used the data to test measurement models consistent with Mastrofski's conceptualization. Our findings suggest that the six dimensions of perceived service quality are not empirically distinguishable. In this sample, we can measure overall perceptions of service quality, but we cannot parse it into empirically distinct dimensions. These results have clear implications for several domains of related research on public perceptions of the police, including the growing and influential body of research on procedural justice and the legitimacy of police and other authority figures.

2. Background

Scholarship on public perceptions of police is rooted in four overlapping traditions of theory and research. In this section we provide a brief review of each tradition and its application to policing. The first is the rapidly expanding body of social science research on service quality (Cronin and Taylor, 1992; Parasuraman *et al.*, 1985). Much of this research has been done in the private sector and published in marketing and business journals. Researchers have developed and tested standardized instruments for measuring service quality across the service industries (Parasuraman *et al.*, 1988). A handful of studies have applied the service quality framework to policing (Donnelly *et al.*, 2006; Reisig and Chandek, 2001; Webb, 1998). Although the customer service perspective is appealing, the police are unique in a number of ways among service industries. Perhaps the most important differences are their ability to use state-sanctioned force, their near total monopoly over their specific service sector, and their capacity to deliver their "services" involuntarily to clients (for example, by arresting or using force against them). Thus, while useful in some ways, the service quality perspective also has limitations in its application to policing.

A second intellectual tradition that serves as a foundation for research on public perceptions of police focuses on justice. Justice theorists outline at least three specific kinds of justice – procedural, interactional, and distributive – that are related both conceptually and practically to public perceptions of police (Greenberg, 1987)[2]. Procedural justice refers to the perceived fairness of the processes leading to an

outcome (e.g. Thibaut and Walker, 1975; Lind and Tyler, 1988). Interactional justice refers to the perceived fairness and quality of the interpersonal interaction between a decision maker and the individuals affected by the decision (e.g. Bies, 2005; Bies and Moag, 1986; Brockner and Greenberg, 1990)[3]. Distributive justice refers to the perceived fairness of an outcome (e.g. Brockner and Siegel, 1996; McFarlin and Sweeney, 1992). Some research examines perceptions of justice as a dependent variable (Frey, 1997), while other research examines the effects of these perceptions on other dependent variables (Liljander, 1999; Nel *et al.*, 2000). Research on perceptions of justice has taken place in many settings where authority figures interact with those subordinate to their authority, including families (Gold *et al.*, 1984), workplaces (McFarlin and Sweeney, 1992), courts (Casper *et al.*, 1988), and prisons (Henderson *et al.*, 2010). Since police are a fundamental component of the administration of justice system, it is important to ask whether those with whom they interact perceive them as just. A recent flurry of research in policing has examined the antecedents and effects of procedural, interactional, and distributive justice (e.g. Paternoster *et al.*, 1997; Reisig *et al.*, 2007; Reisig and Lloyd, 2009; Sunshine and Tyler, 2003; Tyler, 1990; Tyler and Huo, 2002). This body of research has generated intriguing findings about the effects of these different types of justice on compliance with the law and legal authorities.

A third tradition of public perception research examines the global notion of citizen satisfaction with government, public agencies, and public services (e.g. Kelly and Swindell, 2002; Stipak, 1977, 1979). This wide-ranging body of research is rooted in multiple disciplines, including political science, public administration, and economics. It is concerned primarily with global or diffuse citizen attitudes such as "satisfaction with" or "confidence in" government agencies or institutions. Research on citizen satisfaction tends to rely on different assumptions and designs than research on service quality and justice because it does not assume that a citizen has experienced a specific type of service or an encounter with a particular agency. Thus while research on service quality and justice is often able to gauge client perceptions of a particular transaction with a service provider or authority figure, citizen satisfaction research focuses on more global or diffuse perceptions about the institution (e.g. Brandl *et al.*, 1994; Frank *et al.*, 2005). Although only a small proportion of citizens has direct face-to-face contact with a police officer each year (Langan *et al.*, 2001), citizens still have opinions about the quality of policing in their community (or state or nation), and hundreds of public opinion polls have attempted to measure these opinions (Gallagher *et al.*, 2001). Many studies have tested models that explain citizen satisfaction with police (e.g. Brandl *et al.*, 1994; Reisig and Parks, 2000; Worrall, 1999). Because public opinion polls gauge attitudes among citizens who may not have been direct service recipients, this body of research is usually distinct from research on service quality and justice.

Finally, a fourth tradition of public perception research focuses on the concept of legitimacy. Legitimacy is "a quality possessed by an authority, a law, or an institution that leads others to feel obligated to obey its decisions and directives voluntarily" (Tyler and Huo, 2002, p. 102). The notion of voluntary compliance is the defining characteristic of legitimacy. If the majority of people chose not to comply voluntarily with the law or legal authorities, formal social control institutions would become overwhelmed. According to Zelditch (2006, p. 325): "if a regime is legitimate, even the

disaffected, no matter how much they dislike the regime, tend, for a time at least, to willingly comply with it . . . ". Institutions that are not perceived as legitimate are more likely to encourage resentment, hostility, and opposition among citizens and clients. Philosophers have speculated about the legitimacy of government and other social institutions since antiquity, and social scientists from numerous disciplines – including political science, sociology, social psychology, and others – have sought to measure, explain, and otherwise understand legitimacy. The police have endured challenges to their legitimacy since their inception, and people have widely varying opinions on the extent to which they view the police as legitimate and worthy of compliance. A recent summary of research on fairness and effectiveness in policing emphasized the importance of police legitimacy in the eyes of the public, pointing out that it can "be created or undermined by their behavior" (Skogan and Frydl, 2004, p. 297). Researchers have measured legitimacy in a number of ways, typically asking people about their perceived obligation to obey the law, cynicism about the law or legal authorities, or their degree of "support, allegiance, institutional trust, and confidence" (Skogan and Frydl, 2004, p. 299). Public perceptions of legitimacy overlap in both concept and measurement with all three of the previous research traditions.

2.1 Combining the perspectives

The four research traditions outlined above are sometimes separate and independent and sometimes combined. For instance, several studies outside of policing have examined the effects of justice on perceptions of service quality (Carr, 2007; Liljander, 1999; Nel *et al.*, 2000; Smith *et al.*, 1999). Similarly, some research examines the influence of service quality on overall satisfaction ratings with an organization or agency (e.g. Andaleeb, 2001; Tian-Cole *et al.*, 2002). These are just some of the many ways in which these four bodies of scholarship overlap. One ongoing complication that we will discuss in more detail later is that the same indicators are frequently used in these different bodies of research to measure different concepts (e.g. Colquitt, 2001; Reisig *et al.*, 2007).

Within the policing arena, there is significantly less overlap between these bodies of scholarship. For instance, while dozens of studies in marketing and business journals have examined the links between procedural justice and service quality, these relationships have not yet been explored in policing. Nonetheless, some research in policing has examined the links between procedural justice and legitimacy (e.g. Reisig *et al.*, 2007; Sunshine and Tyler, 2003; Tyler and Huo, 2002), and other research has focused on the relationships between general and specific attitudes toward the police (e.g. Brandl *et al.*, 1994; Reisig and Parks, 2000; Skogan, 2006). While many studies have sought to measure public perceptions of police, very few of these studies have paid serious attention to understanding the structure of these attitudes (Scaglion and Condon, 1980; Sullivan *et al.*, 1987; Webb and Marshall, 1995). Little is known about the dimensionality of public perceptions of police, about the relationships between the salient dimensions, and about their antecedents and effects.

2.2 Developing a measure of perceived police service quality

Drawing on Parasuraman *et al.*'s (1985, 1988, 1991) conceptualization and measurement of service quality in the private sector, Mastrofski (1999) identified six characteristics that "illuminate the service aspects of policing," including

attentiveness, reliability, responsiveness, competence, manners, and fairness. According to Mastrofski, citizens want the police to be available and accessible (attentiveness). In addition, citizens want police service to be predictable (reliability), client-centered (responsiveness), and effective at getting the job done (competence). Finally, citizens want the police to treat them with respect (manners) and use fair procedures (fairness). Mastrofski's conceptualization is one of the most carefully articulated statements in the literature about the structure of public perceptions of the police, therefore testing its validity is important.

We test the validity of Mastrofski's six-dimensional model as part of a larger concern with understanding the nature of public perceptions of police. We had three goals in mind when beginning this project. First, we wanted to translate Mastrofski's conceptualization of police service quality into survey items that sought explicitly to measure the dimensions he articulated. Second, we wanted to specify a measurement model containing multiple indicators of each dimension and then test the conceptual model. Third, if the empirical model behaved in accordance with the conceptual model, we intended to test a series of propositions about the effects of service quality on overall satisfaction with police. As we will demonstrate, however, problems with the measurement model made it inappropriate to proceed to the third step. The methods, results and implications of our research are discussed below.

3. Data and methods

The data for our study come from a 2003 survey of citizens in a suburban Virginia community, conducted in partnership with a local police agency. Given the nature of the data, our results cannot be generalized to larger populations. Fortunately, that is not our main concern here. Rather, our primary goal was to conduct an empirical test of a conceptual model of perceived police service quality. Thus, we were primarily concerned with issues of conceptualization and measurement.

3.1 Selecting the sample

An extensive body of research exists on public attitudes toward the police, ranging from academic studies, to local surveys conducted by police agencies, to national public opinion polls. Each of these methodologies has its strengths and weaknesses. The general population has opinions about the police that are important to gauge, but understanding the factors that shape those opinions is difficult because only a small proportion of people has contact with the police in any given year (Langan *et al.*, 2001; Smith *et al.*, 1999b). For that reason, many agencies have turned to contact surveys, which are administered only to those people who have had direct contact with the police within a certain time frame (Bondurant, 1991; Langan *et al.*, 2001; Mastrofski, 1981; Mestre, 1992). Contact surveys provide useful information from people well positioned to render judgment about the police, but these people often constitute only a small segment of the community.

For this study, we surveyed citizens who experienced recent contact with the police. Unlike typical contact surveys that ask respondents about a specific encounter with a service provider, our interest here was to measure more global attitudes about the service quality provided by the police agency. We viewed it as likely that some service recipients might have experienced multiple contacts with the agency, and we did not

want to discard impressions formed during previous contacts. Because we chose to survey individuals who had contact with the police, but to measure their general impressions beyond that specific contact, our instrument represented an unique hybrid of two common research approaches.

We had initially hoped to survey a broad base of service recipients, including victims, witnesses, arrestees, ticketed drivers, and others who interacted with the police. However, the police department's preference was to limit the population to three categories of people: those involved in motor vehicle collisions (including those who received citations), those who filed police incident reports, and those listed as victims on offense reports. At the request of the department, we excluded witnesses and those who had received a "simple" citation (outside the context of a motor vehicle collision) due to the brevity of the contact. The police department's information technology (IT) coordinator extracted a pseudo-random sample of 500 people residing within the two zip codes covered by the agency, whose contact with the police fell into one of these three categories and occurred in the previous 12 months[4]. Table I lists the nature of the contact for the 374 citizens for whom such information was available[5]. It illustrates the diverse mix of police-citizen contact types represented among the citizens selected for our sample.

3.2 Developing the questionnaire

When designing the instrument, we focused our efforts on developing quantitative measures of the six dimensions of service quality outlined by Mastrofski (1999): attentiveness, reliability, responsiveness, competence, manners, and fairness. We struggled with the conceptual meaning of attentiveness. On its face, it seemed to overlap with responsiveness, but as defined by Mastrofski (1999), it appeared to

Table I.
Nature of contact for people in the sample

Case type	Number
Motor vehicle accident	130
Larceny	68
Vandalism	44
Assault	39
Suspicious person	14
Burglary	11
Animal case (dog/cat bites)	7
Natural death	7
Property report	7
Fraud/forgery	7
Intimidation	6
Telephone violation	6
Auto theft	5
Mental health case	5
Missing person	5
Offenses against family/child	2
Open window or door	2
Drunken/disorderly behavior	2
Robbery	2
Other	5

represent the concept of accessibility more than attentiveness. Therefore we chose to develop a measure of accessibility rather than attentiveness. We were mindful that constructing valid and reliable measures of the six dimensions required us to include at least three, and preferably four, questions measuring each dimension. The items used to measure each dimension were developed by a working group consisting of the first author, five undergraduate honors students, and several officers from the police agency under study. The final instrument contained 34 questions: 27 on service quality and satisfaction, two on prior contact with the police, and five on demographics.

3.3 Administering the survey

We mailed the initial batch of surveys in April 2003. Per the recommendations of Dillman (2000), we personalized the mailing in several ways to improve response rates, including handwriting the principal investigator's name and address on the outgoing survey packet and the incoming envelope, and using postage stamps rather than metered mail. The research team tracked the incoming survey responses as soon as they started to arrive, and a reminder postcard was sent to non-respondents once there was a noticeable decline in the incoming responses. A second mailing of the complete survey packet was sent to the remaining non-respondents two weeks later. In total, 138 completed surveys were returned. After adjusting the denominator for bad addresses and out of scope respondents (e.g. minors and those who had died), the final response rate was 32.1 percent.

While this response rate is far from ideal, it is consistent with other mail survey response rates. For example, a meta-analysis of 64 mail surveys of the general public found an average response rate of 35 percent (Green *et al.*, 1998)[6]. The sample size alone is not necessarily problematic, given numerous statistical advances for testing multivariate models with small samples (Hoyle, 1999). The larger concern is the potential for nonresponse bias, in other words, if the respondents are systematically different than non-respondents. However, a low response rate does not necessarily imply nonresponse bias. For example, a recent meta-analysis of 59 studies containing estimates of nonresponse bias concluded that "nonresponse differences are largely similar across the range of nonresponse rates found" (Groves and Peytcheva, 2008, p. 175). Unfortunately, we have no way of estimating nonresponse bias in this study, so we do not know to what extent it represents a limitation.

3.4 Profile of survey respondents

As shown in Table II, the majority of survey respondents were white, married, college-educated, and female. Of the respondents 84 percent were white, and 58.5 percent were female. Asians constituted 4.5 percent of the respondents, while African Americans and Hispanics made up 3.8 percent each. A total 27 percent had some college or an Associate's degree, 27 percent held a Bachelor's degree, and 30.8 percent had a Master's or advanced degree. The vast majority of respondents (87.1 percent) were in the 35-65 age range. The fewest respondents, only 6.1 percent, came from the 25-34 year age range. Only 26 percent of the respondents reported that they had received some form of traffic citation in the previous two years. We did not explicitly sample those who had received citations, but these cases may have resulted from our decision to sample from people involved in collisions. Less than one percent (0.8 percent) of respondents had been arrested during the same time period.

Variable	Categories	Percent
Race	White	84.2
	Black	3.8
	Asian	4.5
	Native American	1.5
	Hispanic/Latino	3.8
	Other	2.3
Age	18-24	6.8
	25-34	6.1
	35-44	28.8
	45-54	22.7
	55-64	14.4
	65+	21.2
Sex	Male	41.5
	Female	58.5
Education	No High School/some High School	3.8
	High School Diploma	10.0
	Some College/Associate's Degree	27.7
	Bachelor's Degree	27.7
	Master's Degree or Advanced Degree (MD, JD, PhD, etc.)	30.8
Marital status	Never married	16.9
	Separated/divorced	11.5
	Widowed	5.4
	Married	66.2
Cited within the past two years?	No	74.0
	Yes	26.0
Arrested within the past two years?	No	99.2
	Yes	0.8

Table II.
Descriptive statistics

4. Findings

4.1 Descriptive findings

Table III contains descriptive statistics for each police service quality item, including the percentage of respondents selecting each response option and overall mean scores. Potential scores for each service quality item ranged from 1 (strongly disagree) to 6 (strongly agree). Mean scores for each item were computed from these ordinal scores, with higher mean scores representing perceptions of greater service quality.

4.2 Measuring service quality

Our initial strategy in this study was to develop, test, and refine a measurement model that specified six dimensions of service quality. Once these measures were calculated and refined, we planned to examine their effects, together with numerous control variables, on overall satisfaction with the police. Our initial analyses, however, suggested that for this population, the six dimensions of service quality were not empirically distinguishable. As we will discuss later, we view this serendipitous finding as fundamentally important for several lines of research on perceptions of service providers and authority figures, including the growing literature on procedural justice and legitimacy.

Table III.
Survey findings

Variable	Question	Dimension	Strongly disagree (%)	Disagree (%)	Slightly disagree (%)	Slightly agree (%)	Agree (%)	Strongly agree (%)	Mean score
V1	The Chief of the City of Orangetown Police Department is willing to speak with citizens who have concerns about the police	Accessibility	1.0	5.9	2.9	15.7	51.0	23.5	4.80
V3	City of Orangetown police dispatchers answer telephone calls in a timely manner	Accessibility	0.8	0	2.5	13.1	54.9	28.7	5.07
V12	City of Orangetown police officers, when handling calls for service, advise citizens of a telephone number where they can be reached should the citizen need follow-up assistance	Accessibility	1.7	0.8	2.5	12.5	42.5	40.0	5.13
V25	It is easy to contact the police	Accessibility	0	0.8	2.4	4.9	50.4	41.5	5.29
V6	City of Orangetown police officers appear to be knowledgeable about police procedures	Competence	1.7	0.8	2.5	9.1	56.2	29.8	5.07
V8	City of Orangetown police officers resolve problems effectively	Competence	0.8	5.0	5.9	16.8	42.0	29.4	4.82
V16	City of Orangetown police officers are knowledgeable about resources available in the community	Competence	1.8	0.9	2.7	11.5	57.5	25.7	4.99
V20	City of Orangetown police officers are well trained	Competence	0.09	1.7	3.4	9.4	53.8	30.8	5.06
V10	City of Orangetown police officers remain neutral and fair	Fairness	3.5	7.1	4.4	9.7	49.6	25.7	4.72
V13	Bias-based policing is not a problem in the City of Orangetown Police Department	Fairness	2.8	5.6	6.5	9.3	53.7	22.2	4.72

(*continued*)

Table III.

Variable	Question	Dimension	Strongly disagree (%)	Disagree (%)	Slightly disagree (%)	Slightly agree (%)	Agree (%)	Strongly agree (%)	Mean score
V15	City of Orangetown police officers use fair and impartial decision making when resolving disputes in the community	Fairness	1.0	6.7	6.7	9.5	52.4	23.8	4.77
V2	City of Orangetown police officers treat people equally	Fairness	2.6	8.8	7.0	4.4	55.3	21.9	4.67
V24	City of Orangetown police officers use fair and impartial decision making when issuing citations	Fairness	3.6%	6.3%	5.4%	6.3%	54.1%	24.3%	4.74
V9	City of Orangetown police officers address citizens in a respectful manner and appropriate tone	Manners	1.6	5.7	3.3	8.1	44.7	36.6	4.98
V4	City of Orangetown police officers take into consideration the feelings of citizens with whom they have had contact	Manners	2.4	4.1	4.1	12.2	48.8	28.5	4.86
V17	City of Orangetown police officers pay attention and listen to what citizens say to them	Manners	3.2	4.8	3.2	12.0	48.8	28.0	4.82
V18	City of Orangetown police officers readily identify themselves to citizens with whom they have contact by providing their name and contact information	Manners	0	0	3.3	8.9	49.6	38.2	5.23
V5	City of Orangetown police officers are consistent in the services they provide to citizens	Reliability	2.6	4.3	2.6	12.2	51.3	27.0	4.86

(*continued*)

Variable	Question	Dimension	Strongly disagree (%)	Disagree (%)	Slightly disagree (%)	Slightly agree (%)	Agree (%)	Strongly agree (%)	Mean score
V19	City of Orangetown police officers attempt to maintain a safe environment for citizens	Reliability	0	0	3.3	13.8	48.0	35.0	5.15
V7	City of Orangetown police officers take the appropriate action when a citizen is in need of services	Reliability	0.8	1.7	5.0	5.8	54.5	32.2	5.08
V21	City of Orangetown police officers follow through on commitments they make to citizens.	Reliability	2.6	2.6	2.6	16.7	51.8	23.7	4.83
V22	City of Orangetown police officers attend to the safety and welfare of citizens.	Responsiveness	0.8	0.8	1.7	10.8	56.7	29.2	5.09
V23	City of Orangetown police dispatchers are responsive to citizens who request services	Responsiveness	0.8	1.6	1.6	8.2	54.1	33.6	5.14
V14	City of Orangetown police officers keep citizens informed and follow up with them when necessary	Responsiveness	2.6	2.6	11.3	11.3	49.6	22.6	4.70
V11	City of Orangetown police officers meet or exceed the expectations of citizens in responding to requests for assistance	Responsiveness	0.8	4.1	6.6	11.6	44.6	32.2	4.92
V26	City of Orangetown police officers take action when problems are brought to their attention	Responsiveness	1.6	2.5	3.3	12.3	50.0	30.3	4.98

Table III.

Ideally we would have multiple independent samples with sufficient sample size to enable us to specify and test an initial model on one sample and then continue to calibrate and test more refined models successively on fresh samples. This type of testing and calibration process would ensure that changes in model specification at each successive step are not the result of sampling error or trivial variation across samples. Unfortunately, we only have one sample available and it is small relative to the number of variables. Because of the small sample size, we rely on several modeling approaches, some of which make greater demands of the data than others and some of which are more appropriate for small samples. The goal is to determine whether the findings are robust across a series of different statistical methods.

The first step in the analysis was to form additive indices measuring each of the six dimensions. This approach is not ideal from a measurement perspective, but given the small sample size, it is an easy way to examine the measurement properties of the six constructs. All of the survey items on service quality were measured using six ordinal categories ranging from one ("strongly disagree") to six ("strongly agree"), thus higher scores correspond with perceptions of greater police service quality. Table IV contains descriptive statistics resulting from this effort. Two conclusions are evident from this initial analysis. First, as evidenced by the alpha coefficients on the bottom row, the indices are internally consistent. Within each dimension, respondents who viewed service quality as high on one item tended to view it as high on the other items as well. Second, the correlations between the six indices are very high, ranging from 0.73 to 0.93, with a mean correlation of 0.86. This is the first piece of evidence to signal a potential problem with discriminant validity. It suggests that the dimensions of service quality articulated by Mastrofski (1999) may not be empirically distinguishable in this sample.

The second step in the analysis was to estimate a confirmatory factor analysis model with latent variables representing each of the six dimensions of service quality. We chose the robust mean and variance adjusted weighted least squares (WLS) estimator available in the software package Mplus (Muthén and Muthén, 1998-2007). Monte Carlo simulations have found that the robust WLS estimator performs well in models with ordinal outcomes, including those with small samples and skewed distributions (Flora and Curran, 2004; Muthén *et al.*, 1997)[7]. The CFA model we specified is complex relative to the sample size, therefore we also used a bootstrapping procedure to estimate test statistics and standard errors[8]. Simulation studies have found bootstrapping to be a useful procedure for estimating parameters in confirmatory factor models with ordinal data and small sample sizes (Zhang and Browne, 2006). We encountered estimation problems in the analyses both with and without bootstrapping. We traced the source of these problems to the latent variable covariance matrix (Ψ) which was not positive definite in both cases, most likely because of multivariate linear dependency among the latent variables[9]. The correlation estimates between the latent variables (some of which are out-of-range due to the nonpositive definite Ψ matrix) range from 0.90 to 1.03 with a mean of 0.98 in the CFA without bootstrapping; in the bootstrapped CFA, the correlations ranged from 0.83 to 1.0 with a mean of 0.94. These results are consistent with the findings from the first step of our analysis using additive indices, except in this case the correlations between factors are even larger.

Scale	Manners	Accessibility	Competence	Fairness	Responsiveness	Reliability	Overall
Manners	1.0						
Accessibility	0.733	1.0					
Competence	0.900	0.797	1.0				
Fairness	0.900	0.758	0.856	1.0			
Responsiveness	0.908	0.784	0.931	0.843	1.0		
Reliability	0.907	0.836	0.895	0.875	0.922	1.0	
Overall	0.861	0.689	0.863	0.867	0.865	0.883	1.0
α	0.863	0.756	0.917	0.960	0.885	0.867	*0.977*

Table IV.
Pearson correlations and alpha coefficients

The large correlations that we have obtained so far between the separate dimensions of service quality signal a problem with discriminant validity. Separate dimensions of a phenomenon may be correlated with one another, but when the correlations are too high, the natural question is whether the dimensions are really distinct from one another. According to Brown (2006, p. 166), "in applied research, a factor correlation that equals or exceeds 0.85 is often used as the cutoff criterion for problematic discriminant validity." A typical strategy for dealing with highly correlated factors with poor discriminant validity is to combine them into a single factor. Taken together, the findings so far suggest that a six-factor model is not appropriate for these 26 items. Since the correlations between the factors are uniformly high, we examined the fit of a one-factor model instead.

We began by estimating a one-factor CFA model using the same estimator as in previous models, first without and then with bootstrapping. The initial one-factor CFA model fit the data reasonably well, though fit statistics and other indicators (such as modification indices and the analysis of residuals) were useful for highlighting some localized areas of strain in the model[10]. We treated this initial one-factor model as a baseline model and explored options for modifying it in subsequent analyses. The modification indices revealed that freeing some of the error correlations would improve the fit of the model. Based on the modification indices and substantive interpretability, we freed two of the error correlations[11]. The revised model fit the data slightly better than the original model[12]. The final step in testing the one-factor model of service quality was to estimate the model using bootstrapping. The findings from the bootstrapped estimates mirror those from the non-bootstrapped estimates. The overall model fit the data well and 24 of the 26 items had statistically significant loadings on the factor, with standardized loadings ranging from 0.575 to 0.992 and averaging 0.867[13]. The standardized factor loadings from this analysis are shown in Table V.

Although this study presented some analytical challenges due to its small sample size, our results were consistent across multiple types of analysis ranging from simple to complex. A basic approach using additive indices to represent the six dimensions of service quality showed that they were highly correlated (mean correlation = 0.86). A confirmatory factor analysis (CFA) using a robust estimator that has been shown to perform well with small sample sizes yielded similar findings (mean correlation = 0.98). Finally, a bootstrapping procedure to estimate the same model obtained similar findings (mean correlation = 0.94). Based on these consistent findings across methods, we concluded by testing and refining a one-factor model of perceived service quality that achieved a reasonably good model fit.

Taken together, these various results suggest that for this sample, using this instrument, respondents were unable to distinguish between different dimensions of service quality. Thus we conclude that for this sample, the six-dimensional model of police service quality has poor discriminant validity. In this case, police service quality appears to have only one underlying dimension, not six. It appears that respondents had a general notion of service quality, but that it could not be parsed into the finer dimensions we attempted to measure here. The separate measures of service quality in this study appear to measure a single concept – presumably overall service quality or some other similar perceptual dimension.

Var	Question	λ
V1	The Chief of the City of Orangetown Police Department is willing to speak with citizens who have concerns about the police	0.792
V3	City of Orangetown police dispatchers answer telephone calls in a timely manner	0.672
V12	City of Orangetown police officers, when handling calls for service, advise citizens of a telephone number where they can be reached should the citizen need follow-up assistance	0.756
V25	It is easy to contact the police	0.692
V6	City of Orangetown police officers appear to be knowledgeable about police procedures	0.898
V8	City of Orangetown police officers resolve problems effectively	0.950
V16	City of Orangetown police officers are knowledgeable about resources available in the community	0.906
V20	City of Orangetown police officers are well trained	0.907
V10	City of Orangetown police officers remain neutral and fair	0.992
V13	Bias-based policing is not a problem in the City of Orangetown Police Department	0.967
V15	City of Orangetown police officers use fair and impartial decision making when resolving disputes in the community	0.967
V2	City of Orangetown police officers treat people equally	0.947
V24	City of Orangetown police officers use fair and impartial decision making when issuing citations	0.917
V9	City of Orangetown police officers address citizens in a respectful manner and appropriate tone	0.901
V4	City of Orangetown police officers take into consideration the feelings of citizens with whom they have had contact	0.947
V17	City of Orangetown police officers pay attention and listen to what citizens say to them	0.936
V18	City of Orangetown police officers readily identify themselves to citizens with whom they have contact by providing their name and contact information	0.575
V5	City of Orangetown police officers are consistent in the services they provide to citizens	0.946
V19	City of Orangetown police officers attempt to maintain a safe environment for citizens	0.723
V7	City of Orangetown police officers take the appropriate action when a citizen is in need of services	0.962
V21	City of Orangetown police officers follow through on commitments they make to citizens	0.848
V22	City of Orangetown police officers attend to the safety and welfare of citizens	0.880
V23	City of Orangetown police dispatchers are responsive to citizens who request services	0.820
V14	City of Orangetown police officers keep citizens informed and follow up with them when necessary	0.866
V11	City of Orangetown police officers meet or exceed the expectations of citizens in responding to requests for assistance	0.903
V26	City of Orangetown police officers take action when problems are brought to their attention	0.866

Table V.
Factor loadings

5. Discussion

We can only speculate as to why the six dimensions we hoped to examine were not empirically distinguishable from one another, but three explanations seem most plausible. First, the community in which we carried out this study is a small, prosperous, homogeneous community with low crime and little disorder, and the police agency receives little exposure in the media compared with others in the region. Therefore, the range of public opinions about the police is probably narrower than in more heterogeneous or more troubled communities. It is possible that the survey instrument we developed in this study may produce better, more detailed measures of service quality in busier, more heterogeneous communities with more social problems than the city we surveyed.

Second, the depth and level of detail inherent in citizens' opinions may depend on the quantity or intensity of their contacts with the police. The police-citizen contacts experienced by people in our sample appear, in some cases, to have been fleeting or of low-intensity. Although all of the respondents were selected from a police database containing individuals who had experienced a recent contact with the police, some of the respondents wrote that they had insufficient experience to answer our detailed questions[14]. Our instrument may work better with respondents whose contacts with the police were frequent or intense enough to enable them to form more fine-grained opinions about police service quality. This interpretation of our findings is consistent with findings from research in a very different regulatory context – the relationships between nursing home regulators and the nursing home industry. Makkai and Braithwaite (1996, p. 96) reported that "Australian nursing home regulation is a domain where corporate satisfaction with the regulatory process is mostly high … concerns about impartiality, ethicality, and correctability are rarely intense. These … facets might loom as more important in regulatory settings with greater heterogeneity of client satisfaction." This interpretation is also consistent with a body of marketing research suggesting that clients' perceptions of service encounters are influenced by the duration and affective content of these encounters, as well as the spatial proximity of the service provider and recipient (Price *et al.*, 1995). Clients may experience shallow attitudes toward encounters that are fleeting, emotionally neutral, and distant. Their attitude structures may become significantly more complex when experiencing encounters that are extended, emotionally charged, or intimate.

Third, our hybrid approach of sampling recent contacts but asking broader questions about the police that transcend a specific contact may have led us away from the very kind of specificity we were trying to measure. Although we were unable to measure the six individual dimensions of service quality using this sample, it would be useful to test whether these dimensions might still be meaningful and measurable in different contexts, or using different sampling or research strategies. Perhaps when asked about the nature of a specific contact, respondents may draw on different cognitive referents when formulating responses to survey items. Little is known about how sampling choices influence the measurement of public perceptions of police, though some research has started to disentangle the relationship between global and specific attitudes toward police (Brandl *et al.*, 1994).

If these suspicions about our findings are correct, they suggest that theories attempting to account for public perceptions of police need to focus more intently on

the contexts in which police and citizens come into contact with one another. It may be the case that citizens' attitudes toward the police do not just vary in magnitude across communities – the very structure of the attitudes themselves may differ. While a small body of research has examined the structure of public attitudes about police, we are not familiar with any research that has examined community differences in the measurement of these attitudes (Scaglion and Condon, 1980; Sullivan *et al.*, 1987; Webb and Marshall, 1995). Perhaps in stable, peaceful communities (like the research site in this study) where the majority of police-citizen contacts are fleeting, not very serious, and of low-intensity, citizens do not develop the kinds of specific, nuanced attitudes toward the police that we sought to measure in this study. Again, if this is true, researchers interested in measuring citizens' perceptions of police might need to tailor their methods to the population they are interested in studying. Citizen satisfaction surveys administered to the general population (who may or may not have had contact with the police) would need to pose only a handful of general survey questions. Surveys administered to citizens who have had limited, fleeting, or not-very-recent contact with the police could be a bit longer and contain more detailed questions than general community surveys. Long, detailed survey instruments would be used to study the perceptions of those whose contact with the police was frequent, intense, and/or very recent. These citizens would be expected to have the most detailed, fully formed, and perhaps multidimensional opinions about the police. These conclusions are speculative and further research is needed to test their validity.

The findings of our study also raise larger questions about conceptual and empirical ambiguity in the largely separate bodies of research on public satisfaction with police, police service quality, procedural justice, and legitimacy. To what extent do these concepts overlap, and is it possible to distinguish one from the others, both conceptually and empirically? Our review of the research in the four traditions suggests a significant degree of conceptual overlap. For instance, citizen satisfaction constructs are sometimes used to measure legitimacy. Similarly, fairness is an important component of both service quality and procedural justice. Indeed, many of the items used in this study to measure fairness are similar to items used by Tyler and his colleagues to measure procedural justice (Tyler, 1990; Tyler and Huo, 2002; Sunshine and Tyler, 2003). Moreover, the results of our study call into question the findings of research that fails to account for the measurement properties – particularly the dimensionality and discriminant validity – of constructs that measure public perceptions of the police.

5.1 Implications for research on procedural justice and legitimacy

Although the findings reported in the present research focus specifically on perceptions of service quality, they have significant implications for the more general body of research on public perceptions of police, particularly those studies that pay insufficient attention to the dimensionality of their measures. We offer two recent examples from scholarship on procedural justice and legitimacy to illustrate this point. Research on procedural justice and legitimacy has a long history in social psychology (Leventhal, 1980; Thibaut and Walker, 1975; Tyler and Lind, 1992), but has begun to expand rapidly into other disciplines, including criminology and criminal justice. Procedural justice scholarship is now an important part of the literature on public

perceptions of police (Skogan and Frydl, 2004). In the two example studies, Tyler and Huo (2002) and Sunshine and Tyler (2003), perceptions of police drawn from a single survey were used to construct measures intended to reflect distinct and theoretically meaningful concepts like procedural justice, motive-based trust, and legitimacy. Moreover, both studies examined the relationships between these measures in multivariate models and drew inferences about cause-and-effect.

First, Tyler and Huo (2002) examined both the antecedents and the consequences of procedural justice using data from telephone surveys of residents in Oakland and Los Angeles (hereafter referred to as the California study) as well as Chicago. The authors presented a variety of analyses but for our purposes we will just concentrate on two of them. In one set of analyses using the Chicago data, procedural justice was measured using an index constructed from eight survey questions that asked respondents about their personal experience with the courts or the police. These items were summed into an additive index that comprised one of two indicators (along with motive-based trust) of a latent variable measuring "social motives." In that analysis, perceived legitimacy, also measured as a latent variable with two summative indicators[15], was endogenous to procedural justice. The authors concluded that social motives (including procedural justice) had a "distinct and significant" influence on legitimacy (Tyler and Huo, 2002, p. 80). The items used to measure procedural justice and legitimacy overlapped in content. For example, some of the survey questions used to construct the legitimacy indices asked about the extent to which the police "treat everyone equally" or "favor some people over others" (Tyler and Huo, 2002, p. 79), while the procedural justice measure was constructed from a series of survey questions that asked whether the respondent was treated fairly, politely, properly, and honestly. This overlap, when considered in the light of the service quality findings presented above, suggests the possibility that the items Tyler and Huo used to measure different concepts located on different sides of a regression equation might be measuring the same underlying concept. Although procedural justice and legitimacy may have different conceptual meanings (just as the dimensions of service quality we examined earlier appeared conceptually distinct), it is not clear whether these constructs can be distinguished empirically given the overlap in the items used to measure them.

In a separate analysis using the California data, Tyler and Huo (2002) treated procedural justice as endogenous to several other variables, including additive indices measuring quality of treatment and quality of decision-making. The authors concluded that both indices had independent effects on people's perceptions of the procedural justice they experienced (Tyler and Huo, 2002, p. 84). The quality of decision-making index was constructed from three survey questions asking whether the officer was honest, "made decisions based on the facts," and "treated me the same as he/she would treat anyone else in the same situation" (Tyler and Huo, 2002, p. 83). The quality of treatment index was constructed from three survey items that focused on whether the officer treated the respondent politely, with concern for his/her rights, and with dignity and respect. There is some inconsistency in Tyler's work about the relationship between procedural justice, quality of decision-making (QD), and quality of treatment (QT). In some studies, QD and QT are treated as antecedents of procedural justice (Tyler and Huo, 2002), while in others they are treated as defining characteristics (Blader and Tyler, 2003; Sunshine and Tyler, 2003)[16].

More recently, Sunshine and Tyler (2003) examined the influence of perceptions of procedural justice on perceptions of police legitimacy using data from two telephone surveys of New York City residents. Procedural justice was treated as a latent variable comprised of three additive indices measuring quality of treatment, quality of decision making, and procedural fairness using measures similar to those contained in Tyler and Huo (2002). Legitimacy was also treated as a latent variable comprised of three additive indices measuring trust in the police, affective feelings about the police, and "perceived obligation to obey the directives of a legal authority" (Sunshine and Tyler, 2003, p. 539). Procedural justice had a strong effect on legitimacy, with a standardized regression coefficient of 0.74. Legitimacy, in turn, affected measures of cooperation with police, support for police (which they term empowerment), and compliance with the law. The authors concluded that "the influence of procedural justice generally flowed through legitimacy" (Sunshine and Tyler, 2003, p. 531).

The same overlap in the items used to measure procedural justice and legitimacy in Tyler and Huo (2002) is evident in the Sunshine and Tyler study as well. For instance, one of the survey questions used to measure legitimacy was "people's basic rights are well protected by the police in your neighborhood" (p. 540). A similar survey question used to measure procedural justice asked about the extent to which police "respect people's rights" (p. 543). Another legitimacy question was "the police in your neighborhood are generally honest" (p. 540), while a similar procedural justice question asked about the extent to which the police "give honest explanations for their actions" (p. 543). Thus, while there are important conceptual differences between some of the indicators of procedural justice and legitimacy, for other items there appears to be conceptual overlap.

We draw on these two studies as examples of a larger tendency in the procedural justice and legitimacy literature to pay insufficient attention to the conceptualization and the measurement properties of theoretically meaningful constructs measuring public perceptions of police. The authors make a clear case that the concepts are separate and distinct from one another, but they do not carry out any of the necessary analytical work necessary to confirm that the measures are empirically distinct. This is a vital next step in the progression of the research, especially since researchers have not done a good job so far of demonstrating that constructs used on different sides of a regression equation are not tapping the same underlying concepts. Without using the proper measurement tools to confirm that constructs built using similar indicators are empirically distinct, it is impossible to know whether the findings are a methodological artifact. These measurement problems are not simply academic; they go to the root of the conclusions that we can meaningfully draw from these influential studies.

Recall that the present study sought to measure six separate conceptual dimensions of perceived police service quality. We used a single survey instrument to measure these concepts and discovered that the resulting measures could not be empirically distinguished from one another. Instead of measuring six separate aspects of police service quality that were conceptually distinct in our minds, we ended up only being able to measure overall service quality. Had we chosen to ignore certain crucial steps in exploring the measurement properties of our constructs, we could have easily made some crucial scientific mistakes. Consider two very real possibilities.

First, suppose that we wanted to explore the independent effects of these six dimensions of service quality on a single dependent variable like overall satisfaction with police (an analysis that is common in both service quality and procedural justice research and that we had in fact had hoped to conduct). If we had failed to examine the correlations between these dimensions, or if we had failed to examine collinearity diagnostics, we would have missed an important finding: that the six dimensions are not independent of one another, and therefore their separate effects on overall satisfaction cannot be estimated[17]. Collinearity is not mentioned and no collinearity diagnostics are reported in either Tyler and Huo (2002) or Sunshine and Tyler (2003). More generally, this problem is endemic in research on procedural justice and legitimacy (Reisig *et al.*, 2007).

Second, suppose we wanted to test a theory about the causal relationships between these (or other) dimensions drawn from the same survey, so that some ended up in a causal model as exogenous variables and others as endogenous variables. Given that the six dimensions were measuring the same underlying concept, were drawn from the same survey instrument, and had the same available response options, we would likely have found large regression coefficients and high levels of explained variance[18]. In such an instance, we might then have concluded mistakenly that the evidence supports the theory. Reisig *et al.*, (2007, pp. 1010-1011) urge researchers to use caution in these instances: "... the observed relationship between an independent variable (e.g. procedural justice) and a dependent variable (e.g. legitimacy) will be artificially high if the two scales feature redundant information."

These problems are common in research on procedural justice and legitimacy because there are no standardized measures, the concepts underlying the measures are described or specified in unstable ways across studies, and the same (or similar) indicators are used to measure different concepts. For instance, Colquitt (2001, p. 387) concludes that these "dimensionality and measurement issues create theoretical and practical problems. An inability to separate purportedly distinct constructs at a measurement level leads to confusion regarding the nomological network of those constructs." Similarly, in their analysis of procedural justice, Reisig *et al.* (2007, p. 1008) lamented that "gauging the psychometric properties of existing process-based variables is frustrated by the fact that composite measures have differed from one study to the next"[19].

One way to avoid these problems is to draw on separate data sources to measure exogenous and endogenous variables. In survey research, unfortunately, that is difficult to do. Another approach is to use caution in translating concepts into tangible measures. At a minimum, using caution would mean paying serious attention to the conceptual substance and the construct validity of the measures to ensure that they are both conceptually and empirically distinct from one another. Inattention to dimensionality and discriminant validity is endemic in most research on public perceptions of police.

Research on citizen or client perceptions of police and other authority figures is conceptually and operationally messy. The concepts overlap with one another and they are translated into measures that overlap as well. The dimensionality of citizen perceptions is unknown and the science of understanding these dimensions is evolving in a way that can best be described as chaotic. Given the theoretical vitality and practical utility of these concepts and this body of research, beginning to build an orderly and cumulative body of research is essential.

6. Conclusion

This study specified and tested a conceptual model of perceived service quality in police agencies. Our primary intellectual foundation was the large body of research on service quality, justice, citizen satisfaction with government, and legitimacy. We tested and rejected a measurement model based on six dimensions of service quality articulated by Mastrofski (1999): accessibility, reliability, responsiveness, competence, manners, and fairness. Our findings suggest that respondents in this study were unable to distinguish between the six dimensions of service quality. Rather, responses to the 26 service-quality items suggested a one-dimensional pattern – a single dimension that we interpret as overall perception of service quality. This study demonstrates the need to pay careful attention to conceptualization and measurement in assessing service quality, procedural justice, satisfaction, legitimacy, and other related perceptual constructs.

Notes

1. As Tyler and Huo (2002, p. 43) argue, "legal authorities can utilize a customer service perspective in one important respect: they can handle problems in ways that show a concern for how people are treated and that reflect a good-faith effort to deal with people's needs."
2. These are not the only kinds of justice discussed by justice theorists. Other major forms include restorative and retributive justice.
3. Some scholars treat interactional justice as a unique concept separate from procedural justice (Bies and Moag, 1986; Colquitt, 2001; Makkai and Braithwaite, 1996), while others treat it as a component of procedural justice (Tyler and Bies, 1990; Tyler and Hou, 2002).
4. The sample was not selected using traditional random sampling methods. We are reasonably confident that the methods used by the IT coordinator to select cases approximate a random sample. However, because we were unable to ascertain the total population of cases in each category, we do not know whether the proportions of cases selected within each category are equivalent. For instance, we do not know whether our sample over-represents people in collisions but under-represents crime victims. Furthermore, we do not know how large our sample is relative to the population of cases eligible for inclusion. Thus, we cannot generalize our findings to a larger population.
5. Unfortunately, due to database problems at the police department, we were unable to recover information on the nature of the contact for 126 of the 500 people selected for the sample.
6. While telephone surveys generally produce higher response rates, they are more expensive and time-consuming. Thus, public agencies like police departments typically rely on mail surveys to gather community feedback.
7. According to Flora and Curran (2004), robust WLS leads to "slightly biased test statistics and standard errors when large models are estimated with small samples. This inflation of the test statistic increases Type I error rates for the chi square goodness-of-fit test, thereby causing researchers to reject correctly specified models more often than expected. It may be that researchers can supplement the chi-square goodness-of-fit test with other fit indices often computed in applications of SEM to retain a model that would otherwise be rejected on the basis of the chi-square goodness-of-fit test alone." They recommend the comparative fit index (CFI) and the root mean square error of approximation (RMSEA).
8. Monte Carlo simulations have cast considerable doubt on commonly used rules-of-thumb about the number of parameters to be estimated or the number of indicators per factor relative to the sample size. Although a large sample size is always preferable, simulation

research has concluded that CFA performs well with small samples when a larger number of indicators per factor is used (opposite what common rules-of-thumb suggest). Moreover, CFA performs better with small samples when "saturation" (the magnitudes of the factor loadings) is uniformly high rather than low or mixed (Marsh and Hau, 1999). Although this study has a small sample size, the number of indicators per factor is high and the loadings are uniformly high. These are the conditions under which CFA is expected to perform well in spite of small samples. For the bootstrapping analysis, we attempted 5,000 replications, each one drawn on a random subset of cases. The averages of the parameter estimates across the replications serve as the point estimates, and the variability across all replications is used to compute standard errors. We are grateful to Cameron McIntosh for recommending the bootstrapping approach

9. A nonpositive definite matrix causes computational problems during the estimation of CFA models (Brown, 2006). Often this results from a singular matrix in which "one or more rows or columns in the matrix are linearly dependent on other rows or columns" (Brown, 2006, p. 187). Singular matrices have no inverse, thus computations requiring the inverse of a matrix cannot be completed. Typically when this happens, it can be traced back to one or more of three explanations: a negative variance or residual variance for a latent variable, a correlation between two latent variables that equals 1.0 or more; or a linear dependency among three or more latent variables (Muthén and Muthén, 1998-2007). We do not see any evidence of a negative variance or residual variance, but we do find evidence of high correlations between the latent variables.

10. Two fit statistics for this initial model were inflated ($\chi^2 = 126.95$, df = 35, $p < 0.0001$; RMSEA = 0.146), but the others suggested a good fit (CFI = 0.975; TLI = 0.993; WRMR = 1.046. Findings from simulation research suggest that the WRMR should be "approximately" less than one (Yu, 2002). All of the factor loadings are statistically significantly different from zero.

11. Three modification indices had a value of 3.84 or greater (the critical value of chi-square with one degree of freedom). Error correlations are ordinarily fixed at zero on the premise that the latent variable should account for all the nonrandom variation between items. Freeing the error correlations represents an implicit acknowledgement that this premise may not hold. The first pair of items highlighted in the modification indices was q18 ("Orangetown police officers readily identify themselves to citizens with whom they have contact by providing their name and contact information.") and q19 ("Orangetown police officers attempt to maintain a safe environment for citizens."). We see no clear substantive connection between these two items, so we chose not to alter the model based on this first modification index. The second item pair was q2 ("Orangetown police officers treat people equally") and q13 ("Bias-based policing is not a problem in the Orangetown Police Department"). In this case there is a clear substantive connection between the two items and not all of the relationship between them can be explained by the factor. Freeing the residual correlation between them ($r = 0.60$) acknowledges the existence of a trivial factor related to equality and bias in policing. The third item pair consisted of q14 ("Orangetown police officers keep citizens informed and follow-up with them when necessary.") and q21 ("Orangetown police officers follow through on commitments they make to citizens"). Once again there is a clear substantive connection between these two items having to do with following up with citizens after an initial contact. Freeing the residual correlation ($r = 0.54$) between them acknowledges this connection.

12. Two fit statistics were inflated ($\chi^2 = 120.53$, df = 35, $p < 0.0001$; RMSEA = 0.140), but the others all suggested a good fit (CFI = 0.976; TLI = 0.993; WRMR = 1.010).

13. The only available test of model fit for the bootstrapped model is the weighted root mean residual (WRMR) which should be approximately less than one (Yu, 2002). The WRMR for

our model was 0.809 which suggests a good fit. The two nonsignificant loadings presented evidence of having encountered estimation problems, so we do not interpret them as casting substantive doubt on the model. The point estimates for these two loadings were both large ($\lambda_{q7} = 0.962$; $\lambda_{q10} = 0.992$), but the standard errors were seriously inflated. The standard error of the estimate for item q7 was more than 12 times the size of the next largest standard error. The standard error for q10 was more than 260 times the size of the next largest standard error (excluding item q7). With ordinal data, sometimes these kinds of estimation problems can result from highly correlated sets of skewed items with some near-zero cells in the joint distributions with other variables.

14. The survey contained a comments section in which many of the respondents provided feedback about the police agency in question. However, seventeen respondents provided comments about the survey itself, and all echoed a similar theme: that the respondent had experienced insufficient contact with the police to provide accurate responses to all of the survey questions. Some of these respondents recommended adding a "don't know" or "not applicable" category to the response options for the service quality questions. These comments raise important issues for those interested in measuring police service quality using citizen surveys, and reflect a longstanding discussion among survey researchers about the use of a neutral response category (Schuman and Presser, 1981; Bishop, 1987). Respondents sometimes find it easier and less cognitively taxing to choose a neutral category than to spend time reflecting on whether they might have an opinion in one direction or the other. Yet, omitting this category sometimes means that respondents who truly have no opinion must either choose a response option that does not apply to them or choose not to answer the question at all. Response rates for the 26 service quality items ranged from 75-92 percent with a mean of 87 percent and a median of 89 percent. To determine whether missing data might have influenced the findings, we re-estimated the models using full-information maximum likelihood, which uses all of the observed data in generating parameter estimates. The results were similar and the conclusions we drew from these analyses were the same.

15. The two indicators were obligation to obey rules and institutional trust. The obligation index contained six items and the institutional trust index contained 19.

16. Some of the confusion over the role of QD and QT is evident in Tyler and Huo (2002): "the relational model articulated by Tyler and Lind (1992, p. 52) identifies three interrelated elements that *define* procedural fairness: motive-based trust, the quality of decision making procedures (neutrality), and the quality of treatment (status recognition)" (emphasis added). Later on the same page they state: "As noted, Tyler and Lind (1992) treated trust as one of three *antecedents* of procedural justice, while Tyler and Blader (2000) treated motive-based trust as an aspect of quality of treatment, one of two *antecedents* of procedural justice" (emphasis added). It is unclear from this discussion of Tyler and Lind (1992) whether QD and QT were defining characteristics of procedural justice (as in the first quote), or antecedents (as in the second). Tyler and Huo (2002) are clear in pointing out that QD and QT are antecedents of procedural justice in their study. On the other hand, Blader and Tyler (2003) and Sunshine and Tyler (2003) both treat QD and QT as indicators of a latent procedural justice variable. Implicit in this specification is the idea that procedural justice causes QD and QT.

17. Although the Gauss-Markov assumptions are not violated by a "nearly linear" relationship among the independent variables, high correlations among them inflate standard errors for regression coefficients and make it difficult to generate "precise estimates of the unique effects of independent variables" (Berry, 1993, p. 27).

18. These issues are referred to by social scientists as "common method" or "monomethod" biases and can result from a number of factors. Podsakoff *et al.* (2003) list four main

categories of common method biases, each one containing multiple types of more specific biases: common rater effects, item characteristic effects, item context effects, and measurement context effects. While there is not sufficient space here to describe them all in detail, it is notable that all of them are present in most of the research on procedural justice and service quality. Data for the independent and dependent variables in these studies are typically drawn from the same person (a common rater), the survey items are constructed similarly (similar item characteristics), and all items are typically contained in the same survey instrument (identical item and measurement contexts). The best approach for dealing with common method bias is to measure the independent and dependent variables using different data sources, different instruments, or different measurement occasions. These approaches are often not feasible in survey research on citizen perceptions of police, yet statistical remedies are also available (Podsakoff *et al.*, 2003). Paying proper attention to measurement and dimensionality issues – particularly to ensure that the independent and dependent variables do not measure the same underlying concept – can go a long way toward minimizing common method bias. For a recent example of research on both service quality and procedural justice that attends carefully to the possibility of common method variance, see Carr (2007).

19. Most studies do not address dimensionality; they simply specify a dimensional scheme and "confirm" it using coefficient alpha. Reisig *et al.* (2007) outline the severe pitfalls in using this approach. Existing research specifies procedural justice as having between one and four dimensions (e.g., Blader and Tyler, 2003; Erdogan *et al.*, 2001; Kim and Mauborgne, 1991; Makkai and Braithwaite, 1996; Reisig *et al.*, 2007; Sunshine and Tyler, 2003; Tyler, 1990; Tyler and Huo, 2002). Moreover, the findings of the studies reporting multiple factors are very different from one another (for instance, see Colquitt, 2001; Blader and Tyler, 2003).

References

Andaleeb, S.S. (2001), "Service quality perceptions and patient satisfaction: a study of hospitals in a developing country", *Social Science & Medicine*, Vol. 52 No. 9, pp. 1359-70.

Berry, W.D. (1993), *Understanding Regression Assumptions*, Sage, Beverly Hills, CA.

Bies, R.J. (2005), "Are procedural justice and interactional justice conceptually distinct?", in Greenberg, J. and Colquitt, J. (Eds), *Handbook of Organizational Justice*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 85-112.

Bies, R.J. and Moag, J.F. (1986), "Interactional justice: communication criteria of fairness", in Lewicki, R.J., Sheppard, B.H. and Bazerman, M.H. (Eds), *Research on Negotiations in Organizations*, JAI Press, Greenwich, CT, pp. 43-55.

Bishop, G.F. (1987), "Experiments with the middle response alternative in survey questions", *Public Opinion Quarterly*, Vol. 51 No. 2, pp. 220-32.

Blader, S.L. and Tyler, T.R. (2003), "A four component model of procedural justice: defining the meaning of a fair process", *Personality and Social Psychology Bulletin*, Vol. 29 No. 6, pp. 747-58.

Bondurant, E. (1991), "Citizen response questionnaire: a valuable evaluation tool", *The Police Chief*, pp. 74-6, November.

Brandl, S., Frank, J., Worden, R. and Bynum, T. (1994), "Global and specific attitudes toward the police: disentangling the relationship", *Justice Quarterly*, Vol. 11 No. 1, pp. 119-34.

Brockner, J. and Greenberg, J. (1990), "The impact of layoffs on survivors: an organizational justice perspective", in Caroll, J.S. (Ed.), *Applied Sociology Psychology and Organizational Settings*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Brockner, J. and Siegel, P. (1996), "Understanding the interaction between procedural and distributive justice: the role of trust", in Kramer, R.M. and Tyler, T.R. (Eds), *Trust in Organizations Frontiers of Theory and Research*, Sage, Beverly Hills, CA.

Brown, T.A. (2006), *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York, NY.

Carr, C.L. (2007), "The FAIRServ model: consumer reactions to services based on a multidimensional evaluation of service fairness", *Decision Sciences*, Vol. 38 No. 1, pp. 107-30.

Casper, J.D., Tyler, T. and Fisher, B. (1988), "Procedural justice in felony cases", *Law & Society Review*, Vol. 22 No. 3, pp. 483-507.

Colquitt, J.A. (2001), "On the dimensionality of organizational justice: a construct validation of a measure", *Journal of Applied Psychology*, Vol. 86 No. 3, pp. 386-400.

Cronin, J.J. and Taylor, S.A. (1992), "Measuring service quality: a reexamination and extension", *The Journal of Marketing*, Vol. 56 No. 3, pp. 55-68.

Dillman, D. (2000), *Mail and Internet Surveys: The Tailored Design Method*, Wiley & Sons, New York, NY.

Donnelly, M., Kerr, N.J., Rimmer, R. and Shiu, E.M. (2006), "Assessing the quality of police services using SERVQUAL", *Policing: An International Journal of Police Strategies and Management*, Vol. 29 No. 1, pp. 92-105.

Erdogan, B., Kraimer, M.L. and Liden, R.C. (2001), "Procedural justice as a two-dimensional construct: an examination in the performance appraisal context", *Journal of Applied Behavioral Science*, Vol. 37 No. 2, pp. 205-22.

Flora, D.B. and Curran, P.J. (2004), "An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data", *Psychological Methods*, Vol. 9 No. 4, pp. 466-91.

Frank, J., Smith, B. and Novak, K. (2005), "The basis of citizens' attitudes toward the police", *Police Quarterly*, Vol. 8 No. 2, pp. 206-28.

Frey, F. (1997), "The management of justice through accounts: constructing acceptable justifications", unpublished doctoral dissertation, Virginia Polytechnic Institute, Blacksburg.

Gallagher, C., Maguire, E., Mastrofski, S. and Reisig, M. (2001), "The public image of the police", Final report to the International Association of Chiefs of Police, available at: www.theiacp.org/PoliceServices/ExecutiveServices/ProfessionalAssistance/ThePublicImageofthePolice/tabid/198/Default.aspx (accessed 3 May 2010).

Gold, L., Darley, J.M., Hilton, J.L. and Zanna, M.P. (1984), "Children's perceptions of procedural justice", *Child Development*, Vol. 55 No. 5, pp. 1752-9.

Green, K.E., Boser, J.A. and Hutchinson, S.R. (1998), "Response-rate differences and response-enhancement effects by population type", *Psychological Reports*, Vol. 83, pp. 336-8.

Greenberg, J. (1987), "A taxonomy of organizational justice theories", *Academy of Management Review*, Vol. 12 No. 1, pp. 9-22.

Groves, R.M. and Peytcheva, E. (2008), "The impact of nonresponse rates on nonresponse bias", *Public Opinion Quarterly*, Vol. 72 No. 2, pp. 167-89.

Henderson, H.M., Wells, W., Maguire, E.R. and Gray, J. (2010), "Evaluating the measurement properties of procedural justice in a correctional setting", *Criminal Justice and Behavior*, Vol. 37 No. 4, pp. 384-99.

Hoyle, R.H. (Ed.) (1999), *Statistical Strategies for Small Sample Research*, Sage, Thousand Oaks, CA.

Kelly, J.M. and Swindell, D. (2002), "A multiple-indicator approach to municipal service evaluation: correlating performance measurement and citizen satisfaction across jurisdictions", *Public Administration Review*, Vol. 62 No. 5, pp. 610-21.

Kim, W.C. and Mauborgne, R.A. (1991), "Implementing global strategies: the role of procedural justice", *Strategic Management Journal*, Vol. 12, S1, pp. 125-43.

Langan, P., Greenfeld, L., Smith, S., Durose, M. and Levin, D. (2001), *Contacts Between Police and the Public: Findings from the 1999 National Survey*, Bureau of Justice Statistics, Washington, DC.

Leventhal, G.S. (1980) in Gergen, K.J., Greenberg, M.S. and Willis, R.H. (Eds), *Social Exchange: Advances in Theory and Research*, Plenum Press, New York, NY.

Liljander, V. (1999), "Consumer satisfaction with complaint handling following a dissatisfactory experience with car repair", in Dubois, B., Lowrey, T., Shrum, L.J. and Vanhuel, M. (Eds), *European Advances in Consumer Research*, Association for Consumer Research, Provo, UT.

Lind, E.A. and Tyler, T.R. (1988), *The Social Psychology of Procedural Justice*, Plenum Press, New York, NY.

Makkai, T. and Braithwaite, J. (1996), "Procedural justice and regulatory compliance", *Law and Human Behavior*, Vol. 20 No. 1, pp. 83-98.

Marsh, H.W. and Hau, K.T. (1999), "Confirmatory factor analysis: strategies for small sample sizes", in Hoyle, R.H. (Ed.), *Statistical Strategies for Small Sample Research*, Sage, Thousand Oaks, CA, pp. 251-84.

Mastrofski, S. (1981), "Surveying clients to assess police performance: focusing on the police-citizen encounter", *Evaluation Review*, Vol. 5 No. 3, pp. 397-408.

Mastrofski, S. (1999), "Policing for people", paper presented at the third Ideas in American Policing lecture, November 1998, Police Foundation, Washington, DC, available at: www.policefoundation.org/pdf/Mastrofski.pdf (accessed 3 May 2010).

McFarlin, D.B. and Sweeney, P.D. (1992), "Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes", *The Academy of Management Journal*, Vol. 35 No. 3, pp. 626-37.

Mestre, J. (1992), "Community feedback program: twelve years later", *Law and Order*, Vol. 40 No. 10, pp. 57-60.

Muthén, B.O., du Toit, S.H.C. and Spisic, D. (1997), "Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes", unpublished manuscript.

Muthén, L.K. and Muthén, B.O. (1998-2007), *Mplus User's Guide*, 5th ed., Muthén and Muthén, Los Angeles, CA.

Nel, D., Athron, T., Pitt, L. and Ewing, M. (2000), "Customer evaluations of service complaint experiences in the public sector", *Journal of Nonprofit and Public Sector Marketing*, Vol. 7 No. 3, pp. 3-30.

Parasuraman, A., Zeithaml, V. and Berry, L. (1985), "A conceptual model of service quality and its implications for future research", *Journal of Marketing*, Vol. 49 No. 4, pp. 41-50.

Parasuraman, A., Zeithaml, V. and Berry, L. (1988), "SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, Vol. 64 No. 1, pp. 12-40.

Parasuraman, A., Zeithaml, V. and Berry, L. (1991), "Refinement and reassessment of the SERVQUAL scale", *Journal of Retailing*, Vol. 67 No. 4, pp. 420-50.

Paternoster, R., Bachman, R., Brame, R. and Sherman, L. (1997), "Do fair procedures matter? The effect of procedural justice on spouse assault", *Law and Society Review*, Vol. 31 No. 1, pp. 163-204.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.L. and Podsakoff, N.P. (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 88 No. 5, pp. 879-903.

Price, L.L., Arnould, E.J. and Tierney, P. (1995), "Going to extremes: managing service encounters and assessing provider performance", *Journal of Marketing*, Vol. 59 No. 2, pp. 83-97.

Reisig, M.D. and Chandek, M. (2001), "The effects of expectancy disconfirmation on outcome satisfaction in police-citizen encounters", *Policing: An International Journal of Police Strategies and Management*, Vol. 24 No. 1, pp. 88-99.

Reisig, M.D. and Lloyd, C. (2009), "Procedural justice, police legitimacy, and helping the police fight crime", *Police Quarterly*, Vol. 12 No. 1, pp. 42-62.

Reisig, M.D. and Parks, R.B. (2000), "Experience, quality of life, and neighborhood context: a hierarchical analysis of satisfaction with police", *Justice Quarterly*, Vol. 17 No. 3, pp. 607-30.

Reisig, M.D., Bratton, J. and Gertz, M.G. (2007), "The construct validity and refinement of process-based policing measures", *Criminal Justice and Behavior*, Vol. 34 No. 8, pp. 1005-28.

Scaglion, R. and Condon, R.G. (1980), "The structure of black and white attitudes toward police", *Human Organization*, Vol. 39 No. 3, pp. 280-3.

Schuman, H. and Presser, S. (1981), *Questions and Answers in Attitude Surveys*, Academic Press, New York, NY.

Skogan, W.G. (2006), "Asymmetry in the impact of encounters with police", *Police & Society*, Vol. 6, pp. 99-126.

Skogan, W. and Frydl, K. (Eds) (2004), *Fairness and Effectiveness in Policing: The Evidence*, National Academies Press, Washington, DC.

Smith, A.K., Bolton, R.N. and Wagner, J. (1999a), "A model of customer satisfaction with service encounters involving failure and recovery", *Journal of Marketing Research*, Vol. 36 No. 3, pp. 356-72.

Smith, S.K., Steadman, G.W., Minton, T.D. and Townsend, M. (1999b), *Criminal Victimization and Perceptions of Community Safety in 12 Cities, 1998*, Bureau of Justice Statistics, Washington, DC.

Stipak, B. (1977), "Attitudes and belief systems concerning urban services", *The Public Opinion Quarterly*, Vol. 41 No. 1, pp. 41-55.

Stipak, B. (1979), "Citizen satisfaction with urban services: potential misuse as a performance indicator", *Public Administration Review*, Vol. 39 No. 1, pp. 46-52.

Sullivan, P.S., Dunham, R.G. and Alpert, G.P. (1987), "Attitude structures of different ethnic and age groups concerning police", *The Journal of Criminal Law & Criminology*, Vol. 78 No. 1, pp. 177-96.

Sunshine, J. and Tyler, T. (2003), "The role of procedural justice and legitimacy in shaping public support for policing", *Law and Society Review*, Vol. 37 No. 3, pp. 513-48.

Thibaut, J. and Walker, L. (1975), *Procedural Justice*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Tian-Cole, S., Crompton, J.L. and Willson, V.L. (2002), "An empirical investigation of the relationships between service quality, satisfaction and behavioral intentions among visitors to a wildlife refuge", *Journal of Leisure Research*, Vol. 34 No. 1, pp. 1-24.

Tyler, T. (1990), *Why People Obey the Law*, Yale University Press, New Haven, CT.

Tyler, T. and Bies, R. (1990), "Beyond formal procedures: the interpersonal context of procedural justice", in Carroll, J. (Ed.), *Applied Social Psychology and Organizational Settings*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Tyler, T. and Blader, S. (2000), *Cooperation in Groups: Procedural Justice, Social Identity, and Behavioral Engagement*, Psychology Press, London.

Tyler, T. and Huo, Y. (2002), *Trust in the Law, Encouraging Public Cooperation with the Police and Courts*, Russell Sage Foundation, New York, NY.

Tyler, T. and Lind, E.A. (1992), "A relational model of authority in groups", in Zanna, M.P. (Ed.), *Advances in Experimental Social Psychology*, Academic Press, San Diego, CA, Vol. 25, pp. 115-91.

Webb, D.A. (1998), "Segmenting police customers on the basis of their service quality expectations", *The Services Industries Journal*, Vol. 18 No. 1, pp. 72-100.

Webb, V.J. and Marshall, C.E. (1995), "The relative importance of race and ethnicity on citizen attitudes toward the police", *American Journal of Police*, Vol. 14 No. 2, pp. 45-66.

Worrall, J. (1999), "Public perceptions of police efficacy and image: the 'fuzziness' of support for the police", *American Journal of Criminal Justice*, Vol. 24 No. 1, pp. 47-66.

Yu, C.Y. (2002), "Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes", unpublished doctoral dissertation, University of California, Los Angeles, CA.

Zelditch, M. Jr (2006), "Legitimacy theory", in Burke, P.J. (Ed.), *Contemporary Social Psychological Theories*, Stanford University Press, Palo Alto, CA, pp. 324-52.

Zhang, G. and Browne, M.W. (2006), "Bootstrap fit testing, confidence intervals, and standard error estimation in the factor analysis of polychoric correlation matrices", *Behaviormetrika*, Vol. 33 No. 1, pp. 61-74.

About the authors

Edward R. Maguire is Associate Professor and Chair in the Department of Justice, Law and Society at American University. He received his PhD in Criminal Justice from the State University of New York at Albany in 1997. He has held previous positions at George Mason University, the University of Nebraska, the US Department of Justice, and the United Nations. Professor Maguire's professional interests cover a wide range, but most of his work focuses on three topics: police organizations, violent crime, and social science measurement. Edward R. Maguire is the corresponding author and can be contacted at: maguire@american.edu

Devon Johnson is an Assistant Professor in the Administration of Justice Department at George Mason University. She received her PhD in Sociology from UCLA. She has held previous positions at the W.E.B. Du Bois Institute for African and African American Research at Harvard University and the Center for the Study of Urban Poverty at UCLA. Her current research examines public perceptions of crime and the criminal justice system, police-citizen relations, popular support for punitive crime policies, and public opinion toward domestic counter-terrorism policies and practices. Professor Johnson has received awards for her research from the American Association for Public Opinion Research and the Law and Society Division of the Society for the Study of Social Problems.