Criminal Justice and Behavior

Evaluating the Measurement Properties of Procedural Justice in a Correctional Setting

Howard Henderson, William Wells, Edward R. Maguire and Jameel Gray

Additional services and information for *Criminal Justice and Behavior* can be found at:

Email Alerts: http://cjb.sagepub.com/cgi/alerts

Subscriptions: http://cjb.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://cjb.sagepub.com/cgi/content/refs/37/4/384

EVALUATING THE MEASUREMENT PROPERTIES OF PROCEDURAL JUSTICE IN A CORRECTIONAL SETTING

HOWARD HENDERSON
WILLIAM WELLS
Sam Houston State University

EDWARD R. MAGUIRE
American University

JAMEEL GRAY
Aurora University

Research and theory on the effects of fair procedures has gained popularity over the past decade. This is understandable given the inherent appeal of these ideas and the supporting evidence. Research suggests that authorities are able to secure compliance from subordinates when they use fair procedures and when they are viewed as legitimate. Unfortunately, empirical studies of procedural justice and legitimacy are hampered by weak measures of key theoretical constructs. The purpose of this study is to examine the measurement properties of procedural justice in a sample of inmates. Results show that a one-factor model of procedural justice fits the data well, though the authors find evidence of a method effect. Results also demonstrate important differences between the use of a summated procedural justice scale and a scale derived from a factor analysis. These findings illustrate the importance of paying careful attention to construct validity in measures of procedural justice.

Keywords: procedural justice; measurement validity; corrections

The idea that authority figures can generate greater levels of compliance by using fair procedures and treating people with dignity and respect is compelling. Policies aimed at bringing about voluntary compliance and decision acceptance through the use of fair procedures seem more appealing and efficient than policies built around threats of sanctions to induce compliance on unwilling citizens (Tyler & Huo, 2002). Research on procedural justice demonstrates that when people are treated fairly by authority figures, they are more likely to obey the law and authorities. Elements of procedural justice that are believed to enhance voluntary compliance can be found in some contemporary approaches to crime reduction, including restorative justice (Tyler, Sherman, Strang, Barnes, & Woods, 2007).

Empirical research has provided a great deal of support for the hypothesized effects of fair procedures on positive outcomes, most important of which appears to be the perception that authorities are legitimate and deserving of voluntary compliance (Hicks & Lawrence, 2004; Paternoster, Bachman, Brame, & Sherman, 1997; Sunshine & Tyler, 2003; Tyler, 1990; Tyler & Huo, 2002; Tyler et al., 2007; Tyler & Wakslak, 2004). The hypothesis states that when authorities make decisions fairly, based on facts rather than personal preferences, and treat citizens with dignity and respect, then people will have a sense of greater obligation to comply with rules and will trust authorities. The ultimate result is that when citizens

view legal authorities, including the law, courts, and police, as being legitimate, they will be more likely to comply voluntarily with rules and accept decisions. Research has estimated the relationships between fair procedures and perceptions of legitimacy and supports the predicted relationships.

Unfortunately this body of research has neglected a fundamental aspect of the research process by failing to devote sufficient attention to the measurement properties of key theoretical constructs (Maguire, Johnson, & Kuhns, 2008; Reisig, Bratton, & Gertz, 2007). In their comprehensive review of the literature, Maguire and colleagues (2008) point to several measurement problems in existing procedural justice scholarship: (a) inconsistent use of subscales and indicators to reflect procedural justice and legitimacy, (b) inattention to the construct validity of key concepts, (c) failure to utilize statistical methods that are appropriate for ordinal indicators, and (d) the inclusion of similar indicators on both sides of regression equations. Given the vitality and popularity of this scholarship, as well as its implications for policy and practice, it is necessary to examine the measurement properties of these core concepts. In this study we offer new evidence about the measurement properties of survey items commonly used to measure procedural justice. This study also contributes to the literature by examining procedural justice in a correctional setting and by using more appropriate statistical methods.

THE MEANING OF PROCEDURAL JUSTICE

The notion of procedural justice captures the idea of fundamental fairness in the processes of dispute resolution and in the allocation of rewards and punishments (Leventhal, 1980; Lind & Tyler, 1988; Thibaut & Walker, 1975). One of the most noted applications of procedural justice is that associated with the actions of criminal justice authorities, including police and members of the courtroom workgroup. In the context of police decision behavior, procedural justice has been defined as "fairness of the processes through which the police make decisions and exercise authority" (Sunshine & Tyler, 2003, p. 514). Procedural justice is often conceptualized as having multiple dimensions, including consistency, participation opportunity, neutrality, treatment with respect and dignity, and trust in decision makers (Leventhal, 1980; Thibaut & Walker, 1975; Tyler, 2004).

Fair treatment matters to people because it communicates important information about their social status and inherent worth. Tyler (1990) explains that people are concerned about their standing in society. Treatment by authority figures sends messages to people about their social status, both in society at large and in particular social groups.

> Procedures that allow them [citizens] to present evidence on their own behalf affirm status, because they allow people to feel that they are taking part in their social group. Similarly, the willingness of the authority to listen to them and consider their arguments is a recognition of their social standing. (Tyler, 1990, p. 176; see also Lind & Earley, 1992)

The use of fair procedures signals that citizens are valued and protected members of society who are afforded the kinds of treatment everyone is due; procedures that deny citizen input, do not afford dignified and respectful treatment, and suggest that decisions will be based primarily on personal opinions signal to citizens that they are less valued group members and are not worthy of fair treatment.

When police use fair procedures, several important outcomes are expected: greater satisfaction with police, more positive ratings of officers, greater willingness to accept police

decisions, and stronger perceptions that police are legitimate authorities who should be supported and obeyed (Tyler, 1990; Tyler & Huo, 2002). Research has demonstrated that the perception of procedural justice has a significant influence on the acceptance of law and of decisions made by legal authorities. For instance, individual assessments of procedural justice have been found to increase the likelihood that an individual will accept random drug tests (Konovsky & Cropanzano, 1991), reduce the chances of violent recidivism (Paternoster et al., 1997), improve evaluations of court judges who preside over traffic court cases (Tyler, 1984), and increase perceptions that police are legitimate authorities deserving of voluntary obedience (Sunshine & Tyler, 2003; Tyler, 1990). Overall, the literature demonstrates that when citizens perceive that they have been treated fairly, they are more likely to respond positively to authorities and the institutions that employ them, even when the outcomes they receive are less favorable.

Although researchers have primarily focused on studying procedural justice in the context of policing and courts, few have systematically examined the concepts and realities of procedural justice in the correctional setting. The reality is that the correctional environment, with its concentrated set of authority figures interacting regularly in a bounded space with people who have already evidenced a disregard for the law, represents an ideal stage for testing propositions about procedural justice and legitimacy. Procedural injustices in a prison can lead to more than dissatisfaction among the inmate population; they can undermine a sense of order and result in disturbances, violence, and rebellion (Butler & Drake, 2007; DiIulio, 1987; Reisig & Mesko, 2009; Sparks & Bottoms, 1995; Useem & Kimball, 1989; Useem & Reisig, 1999; Woolf, 1991). As Butler and Drake (2007) argue, "Disrespectful treatment within the prison context can have especially detrimental results, because feeling devalued may already be a common daily experience for inmates" (p. 120). In one of the first studies to test predictions about the effects of procedural justice in a correctional setting, Reisig and Mesko (2009) examined Slovenian inmate misconduct. Their results were as predicted: Inmate perceptions of procedural fairness were negatively correlated with self-reported and official indicators of misconduct. Clearly, much is to be gained from increasing our understanding of procedural justice in U.S. prisons.

MEASURING PROCEDURAL JUSTICE

The theoretical and policy importance of procedural justice seems clear, but existing research has failed to adequately assess the measurement properties of key constructs. Writing in 1980, Leventhal reported, "The criteria that define the rules of fair procedures can only be guessed at this time, because there have been few studies of procedural fairness" (p. 39). Since these observations, few measurement advancements have been made (see Maguire et al., 2008; Reisig et al., 2007; Tyler, 1988). The existing literature pays little attention to the reliability and validity of measures of perceived procedural justice, relying on an inconsistent pool of indicators and haphazard specification of the underlying structure of these perceptions. Evidence suggests it is necessary to refine the way procedural justice is measured by social scientists (Maguire et al., 2008; Reisig et al., 2007).

Tyler (1988, p. 107) contends that Thibaut and Walker (1975) and Leventhal (1980) identify four important dimensions of fair procedures: "consistency, decision quality, bias suppression, and representation." The more recent body of literature (Reisig et al., 2007; Tyler & Huo, 2002; Tyler & Smith, 1997) treats procedural justice as two-dimensional, consisting of quality of decision making and quality of treatment. Quality decision making is apparent

when authorities base their decisions on pertinent facts rather than individual preferences and biases, and quality treatment exists when authorities treat people with dignity and respect. Researchers have used a variety of scales and items to create summated measures of procedural justice when estimating its effects. Despite results that show acceptable reliabilities (alpha > .80), there is no established standard for measuring procedural justice and little available evidence to evaluate measurement quality (Reisig et al., 2007).

An examination of three published articles that report on studies that have measured procedural justice illustrates the problem (Sunshine & Tyler, 2003; Tyler, 2005; Tyler & Wakslak, 2004). Sunshine and Tyler (2003) reported the results of two studies, one of which measured procedural fairness with a 20-item summated index and the second with an 11-item summated index. In addition to the different numbers of items, the items themselves differed across the scales. The smaller scale with 11 items included some items that were not included in the 20-item index. Tyler and Wakslak (2004) replicated some items Sunshine and Tyler (2003) used to measure procedural justice. However, Tyler and Wakslak (2004) treated some items as indicators of trust in the *motives* of authorities, a concept that Tyler and Huo (2002) and Tyler (2003) characterize as being theoretically distinct from procedurally fair treatment. Finally, Tyler (2005, p. 329) measured the fairness of police procedures with 5 distinct scales comprising 17 items. The scales included quality of decision making, quality of treatment, racial profiling, racial harassment, and participation opportunity. Such inconsistencies across studies make it difficult to evaluate findings, comprehend the meaning of key concepts like procedural justice and motive-based trust, and draw summary conclusions. Researchers have started to tackle this limitation and assess the measurement properties of key constructs.

In one of the few attempts to assess the measurement properties of procedural justice, distributive justice, and legitimacy, Reisig et al. (2007) find support for a modified procedural justice scale using a set of core items that have been used in existing studies. Reisig et al. examined a set of 10 items to measure procedural justice: 5 each to indicate quality of treatment and quality of decision making. The survey items they used to measure procedural justice are listed in Table 1 and represent items that are routinely used to measure the construct. Interitem correlations showed that items used to measure procedural justice were correlated, to an important extent, with items used to measure other concepts. In other words, "the discriminant validity of these scales was questionable" (p. 1016). Factor analysis showed procedural justice was best measured with 8 of the 10 items because 2 items ("police make decisions based on personal opinion" and "police don't listen to all of the citizens involved before deciding what to do") loaded stronger on a different factor: distributive justice. Interestingly, these 2 items were reverse coded, suggesting a possible methods effect. In addition, 2 items that should have theoretically loaded on distributive justice loaded on procedural justice instead. Reisig et al. dropped these 4 problematic items from their analysis. Cronbach's alpha for the revised, 8-item procedural justice scale equaled .90. Overall, the results suggest that scale quality was improved by purging weak items: The revised scales performed better than the original scales (i.e., improved discriminant validity and reduced collinearity). In terms of predictive validity, Reisig et al. found that the original and revised procedural justice scales were correlated with legitimacy and not substantially correlated with predictor variables measuring other concepts like distributive justice.

One limitation of this measurement approach is that procedural justice was measured using a summated scale, which assumes the items are measured without error and are of equal

TABLE 1: Items Used in Previous Research to Assess the Measurement Properties of Procedural Justice

Study	*Items*
Reisig, Bratton, and Gertz (2007)	*Procedural Justice—Quality of Treatment*
	Police treat citizens with respect
	Police take the time to listen to people
	Police treat people fairly
	Police respect citizens' rights
	Police are courteous to people they come into contact with
	Procedural Justice—Quality of Decision Making
	Police make decisions based upon the facts
	Police explain their decisions to the people they deal with
	Police make decisions based on their own personal feelings
	Police make decisions to handle problems fairly
	Police don't listen to all citizens involved before deciding what to do
Maguire, Johnson, and Kuhns (2008)	*Procedural Justice*
	The police know how to carry out their official duties properly
	The police are neutral and fair when dealing with citizens
	The police address citizens in a respectful manner and appropriate tone
	The police are responsive to the needs of citizens
	The police show care and concern for the welfare of the citizens they deal with
	The police treat everyone equally
	The police make decisions based on facts, not their personal biases or opinions
	The police consider the views of people involved before making their decisions
	The police clearly explain the reasons for their actions

importance. Leventhal (1980) and Tyler (1988) suggest that individuals may give differential importance to aspects of treatment when assessing fairness. The factor analysis by Reisig and his colleagues (2007) showed that individual items made unequal contributions to the latent factor (see Reisig et al., 2007, Table 3, p. 1017), suggesting that the use of factor scores rather than a summated scale would have improved measurement precision. Maguire et al. (2008) report that this problem has also hampered efforts to measure legitimacy.

Two other articles by Reisig and his colleagues explore the measurement of procedural justice outside the United States. In a study of attitudes toward the police among Jamaican adolescents, Reisig and Lloyd (2009) factor analyzed 6 procedural justice items and found that they all loaded on a single factor with good measurement properties. In their study of attitudes toward correctional authorities in Slovenia, Reisig and Mesko (2009) factor analyzed 6 procedural justice items and found that they also loaded on a single factor with good measurement properties. Taken together, the results of studies carried out by Reisig and his colleagues in three very different settings suggest that procedural justice is a one-dimensional construct. At the same time, their results also raise an intriguing question. The two studies that used 6 items to measure procedural justice found that it was one-dimensional. The one study that used 10 items found evidence of what is sometimes referred to as a "nuisance factor," in this case an uninterpretable factor on which only 2 items loaded strongly. Would a larger item pool that taps into a wider domain of procedural justice items result in different inferences about the dimensionality of procedural justice?

Maguire and his colleagues (2008) focused specifically on assessing the measurement properties of legitimacy, but their findings hold important implications for the measurement of procedural justice. In addition to assessing measurement models for legitimacy, they also examined nine indicators of procedural justice that are routinely used in procedural justice scholarship (see Table 1). Results show their measurement models could not effectively discriminate between procedural justice and legitimacy. They first estimated a confirmatory factor model with indicators of procedural justice and legitimacy that were constrained to follow theoretically prescribed loadings. This initial confirmatory model did not fit the data, so they estimated two exploratory factor models. The first exploratory model contained 15 items and generated a three-factor solution that failed to clearly identify a procedural justice component. A second exploratory model utilized an expanded pool of 26 items. Again, the model did not identify a distinct procedural justice factor; procedural justice items loaded together with legitimacy items. A final confirmatory model produced similar results: Items that are commonly used to measure procedural justice did not load onto a distinct procedural justice factor. Rather, these items loaded onto factors with items commonly used to measure legitimacy. This finding is problematic, given that the two constructs are theoretically distinct.

Procedural justice has been measured in dozens of studies, but its underlying structure or "dimensionality" has not been explored in detail. The modal approach in the research is to specify procedural justice as having one or more dimensions, then to form additive indices measuring each of those dimensions, and finally to report alpha coefficients to demonstrate that the indices are reliable. Measurement problems are endemic in the literature (Maguire & Johnson, 2009; Reisig et al., 2007). Across the wide range of procedural justice studies conducted over the past two decades, the number of dimensions typically ascribed to procedural justice ranges from one to three. Without empirical evidence about the underlying structure of procedural justice, it is difficult to know with confidence what this concept actually entails.

The purpose of the current study is to assess the measurement properties of a set of items used in previous research to measure procedural justice. The analysis makes a unique contribution by examining perceptions of procedural justice within a correctional setting. Published studies have examined citizens' perceptions of courts, laws, and police, while ignoring corrections (see Reisig & Mesko, 2009, for an exception), an arena within which legitimacy and procedural justice are highly salient.

METHOD

PARTICIPANTS

Paper-and-pencil survey data were collected from male inmates in a large work release center in Chicago, Illinois. The work release center is classified as the lowest security risk category for facilities operated by the Illinois Department of Corrections. Qualification for work release status is determined with a community correctional center male classification instrument. The assessment instrument is used to create a classification system for work release eligibility. Eligible inmates are then placed at one of the nine Adult Transitional Work Release centers located throughout the state when space becomes available. Inmates

are supervised in program activities structured around employment; vocational training classes; and several different individual and group counseling classes that consist of substance abuse, drug education (in addition to the actual drug treatment), mental health counseling, anger management, family reintegration, and parenting classes. The treatment programs were supervised by licensed drug and alcohol treatment counselors as well as part-time psychiatrist/psychologists who were responsible for conducting the mental heath assessments and screenings. Approximately 45% of the total population of inmates were serviced by the treatment staff in some form or another.

At the time of the study, 321 adult male inmates were residents in the work release program. Participation was solicited during group meetings held at the work release center every Monday and Tuesday between October and December 2006. Some residents did not attend group meetings because they worked irregular schedules, used unsupervised leave during the week, or attended other mandatory programming. Inmates who were not solicited during group meetings to participate in the study were approached individually.

During individual and group meetings with potential participants, the researcher explained the nature of the survey and how confidentiality would be maintained. Inmates who declined to participate in the study were allowed to leave the meeting. A researcher read the cover letter, study description, disclaimer, and survey instrument aloud for the participants. In an attempt to alleviate concerns that surveys would be viewed by staff members, and to enhance validity, inmates were instructed to place their completed surveys in sealed envelopes and were informed that all surveys would be mailed to a university in the southern region of the state for data entry.

Out of the total inmate population ($n = 321$), 249 inmates (78%) were asked to complete the survey. We were unable to request participation from 22% of inmates ($n = 72$) despite several attempts to contact them. These potential respondents had been paroled from the center ($n = 39$), had their work release status revoked ($n = 26$), or had physically escaped from the center ($n = 7$). Of the 249 inmates we approached to request participation in the study, 213 (85.5%) consented to participate. Characteristics of the final sample are presented in Table 2.

MEASURING PROCEDURAL JUSTICE

The item pool for the survey instrument was drawn from existing research on the measurement of four key concepts of interest in our research: procedural justice, legitimacy, motive-based trust, and distributive justice. Because so many survey items have been used to measure these concepts in previous research, it was not feasible to replicate all existing items. We decided to use the items that have been most frequently used. Very little of this research has been conducted in correctional settings, so it was necessary to adapt the items to reference correctional authority figures rather than police, judges, employers, and other authorities.

Procedural justice, the principal concept of interest in the present study, was measured using 12 items that reflect overall assessments of procedural fairness, the quality of decision making, and the quality of treatment (see Table 3). Many theorists and researchers have treated procedural justice as a two- or three-dimensional concept. A common approach by those viewing it as a two-dimensional concept is to differentiate between quality of decision making and quality of treatment (Reisig et al., 2007; Reisig & Lloyd, 2009;

TABLE 2:   Sample Characteristics

Variable	
Race	
African American	135 (65.2%)
White, non-Hispanic	41 (19.8%)
Hispanic	24 (11.6%)
Other	7 (3.4%)
Highest level of education	
Less than high school/GED	61 (29.3%)
High school/GED	82 (39.4%)
Some college classes, no degree	48 (23.1%)
College degree	17 (8.1%)
Age	Mean = 31.92, SD = 9.57
Offense	
Drug	131 (62.1%)
Property	59 (28%)
Violent	16 (7.6%)
Other	5 (2.3%)
Days in custody through study completion (December 2006)[a]	Mean = 627.88, SD = 787.39
Days in work release through study completion	Mean = 401.02, SD = 265.55

a. Information on custody dates and admission to the work release center were collected from official records maintained at the center.

Reisig & Mesko, 2009; Tyler, 2003; Tyler & Huo, 2002). Sunshine and Tyler (2003) treat procedural justice as a three-dimensional concept consisting of fairness of decision making, quality of treatment, and overall fairness. In their analysis, Casper, Tyler, and Fischer (1988) treat overall perceptions of fairness as a dependent variable, rather than as a dimension of procedural justice. Other three-dimensional schemes are also posed in the literature (Tyler, 2005; Tyler & Wakslak, 2004). We treat respondents' answers on these 12 survey items as indicators of one or more latent variables representing perceptions of procedural justice. The indicators are ordinal categorical variables with four categories: *never*, *seldom*, *sometimes*, and *almost always*. Higher scores on each item reflect a greater amount of procedural justice.

We examined the latent structure of the survey items using two approaches. First, we specified a three-factor confirmatory factory analysis (CFA) model of procedural justice consistent with model specifications used in previous research. We then modified the initial model based on diagnostics intended to reveal sources of misfit. Simulation research suggests that this model-building strategy is only effective if the initial model is valid. A more effective approach when the initial CFA model fits the data poorly is to use exploratory factor analysis (EFA) to examine the underlying structure of the indicators. EFA imposes no structure on the data and is useful for assessing the dimensionality of a set of items and detecting items that discriminate poorly (such as those that do not load on any factors or that load on multiple factors). Whereas CFA is often preferred over EFA, CFA can be misleading when the initial model fits poorly. According to Brown (2006), modification indices used in CFA to respecify the model

> are often useful for determining the particular sources of strain in the solution. However, these statistics are most apt to be helpful when the solution contains minor misspecifications. When the initial model is grossly misspecified, specification searches are not nearly as likely to be successful. (p. 159)

TABLE 3:   Items Used to Measure Procedural Justice

Item	Responses			
	Almost Always	Sometimes	Seldom	Never
Overall assessment of fairness				
Q6a. How often do correctional officers make decisions about how to handle problems in fair ways?	29 (13.6%)	106 (49.8%)	61 (28.6%)	17 (8.0%)
Q6b. How often do correctional officers treat inmates fairly?	13 (6.2%)	113 (53.6%)	67 (31.8%)	18 (8.5%)
Quality of decision making				
Q6c. How often do correctional officers treat inmates with dignity and respect?	11 (5.3%)	92 (44.0%)	73 (34.9%)	33 (15.8%)
Q6d. How often do correctional officers accurately understand and apply the rules?	34 (16.0%)	111 (52.4%)	51 (24.1%)	16 (7.5%)
Q6e. How often do correctional officers make decisions based on facts not their personal biases or opinions?	21 (9.9%)	98 (46.2%)	67 (31.6%)	26 (12.3%)
Q6f. How often do correctional officers try to get the facts in a situation before deciding how to act?	20 (9.4%)	88 (41.5%)	76 (35.8%)	28 (13.2%)
Q6g. How often do correctional officers give honest explanations for their actions?	13 (6.1%)	67 (31.6%)	85 (40.1%)	47 (22.2%)
Q6h. How often do correctional officers apply rules consistently to different people?	48 (22.9%)	94 (44.8%)	49 (23.3%)	19 (9.0%)
Q6i. How often do correctional officers treat everyone equally?	16 (7.5%)	71 (33.3%)	74 (34.7%)	52 (24.4%)
Quality of treatment				
Q6j. How often do correctional officers respect inmates' rights?	10 (4.7%)	65 (30.5%)	90 (42.3%)	48 (22.5%)
Q6k. How often do correctional officers give inmates the chance to express their views before making decisions?	12 (5.7%)	74 (34.9%)	81 (38.2%)	45 (21.2%)
Q6l. How often do correctional officers treat inmates politely?	9 (4.2%)	83 (39.2%)	89 (42.0%)	31 (14.6%)

Moreover, EFA serves as a useful check on findings derived from CFA. Using both is informative; if the two approaches lead to the same model specification, the researcher can be more confident that the findings are robust.

MODEL ESTIMATION

Our analysis is based on the underlying assumption that the ordinal survey responses are categorical approximations of underlying continuous random variables. Although the indicators are categorical, the latent variables are treated as continuous. Many of the methods commonly used in normal theory CFA and EFA with continuous indicators need to be adapted for use with ordinal indicators. Instead of using a covariance matrix as input, a polychoric correlation matrix is used for polytomous data (Brown, 2006). Both types of correlations treat the observed categorical variable *y* as a crudely categorized approximation of an underlying continuous latent response variable, *y**. According to Brown (2006),

> The underlying y* variables are related to observed categorical variables by threshold parameters (τ). In the case of a binary indicator (y = 0 or 1), the threshold is the point on y* where

y = 1 if the threshold is exceeded (and where y = 0 if the threshold is not exceeded). Polytomous items have more than one threshold parameter . . . the number of thresholds is equal to the number of categories minus one. (p. 390)

Although thresholds are an important part of the factor models used in this study, we are not concerned with interpreting them for substantive purposes.

Because the outcomes (indicators) are ordinal, we used a robust mean and variance adjusted weighted least squares (WLS) estimator available in the commercial structural equation modeling software package Mplus (Muthén & Muthén, 1998-2007). Monte Carlo simulations have found that the robust WLS estimator performs well in models with categorical outcomes, including those with skewed distributions and small samples (Flora & Curran 2004; Muthén, du Toit, & Spisic, 1997).

RESULTS

We began by specifying a CFA model of procedural justice consisting of three factors: quality of decision making, quality of treatment, and overall fairness (see Sunshine & Tyler, 2003). Fit statistics suggest that the model fit the data reasonably well, though with clear room for improvement (χ^2 = 77.6, df = 29, p < .0000; confirmatory fit index [CFI] = .947; Tucker Lewis index [TLI] = .978; root mean square error of approximation [RMSEA] = .089; weighted root mean square residual [WRMR] = .804). All of the factor loadings are positive and statistically significant. However, the correlations between the three latent variables (listed in Table 4) ranged from .893 to .949, indicating a clear discriminant validity problem. According to Brown (2006), "In applied research, a factor correlation that equals or exceeds .85 is often used as the cutoff criterion for problematic discriminant validity" (p. 166). These high correlations suggest that a one-factor model is the more appropriate specification. We also specified a two-factor model using only 10 of the items to measure quality of treatment and quality of decisions; the 2 items that measure overall perceptions of fair procedures were excluded. This is one of the most common specifications in research on procedural justice. We found that the correlation between the two factors was .946. This high correlation suggests that a one-factor solution is most appropriate.

Next we specified a CFA model of procedural justice consisting of just one factor. Fit statistics suggest that the model fits the data reasonably well, though fit can still be improved (χ^2 = 81.4, df = 30, p < .0000; CFI = .944; TLI = .978; RMSEA = .090; WRMR = .837). All of the factor loadings are positive and statistically significant. The one-factor model appears to be the most viable solution, though later we will explore some options for modifying it slightly.

To check our conclusion about the dimensionality of this set of items, we used EFA, relying on the same WLS estimator we used in the CFA models. Only one factor had an eigenvalue greater than one, so according to the Kaiser-Guttman criterion, we should retain only one factor (Guttman, 1954; Kaiser, 1960).

Moreover, the multifactor solutions were full of cross-loadings and were generally uninterpretable. The one-factor EFA solution is statistically equivalent to the one-factor CFA model we tested earlier, and its fit is identical. The results of both our CFA and EFA analyses are consistent with the inference that inmates' perceptions of procedural justice, as measured using this set of indicators, are one-dimensional.

TABLE 4:   Correlation Matrix Between Three Dimensions of Procedural Justice

	Overall Fairness	*Quality of Decision Making*	*Quality of Treatment*
Overall fairness	1.0		
Quality of decision making	.893	1.0	
Quality of treatment	.937	.949	1.0

Although we are confident in our inferences about the dimensionality of this set of 12 items, the model fit is still not optimal. Our results so far suggest that the partial misfit is not due to items whose relationship with procedural justice is weak or to the influence of additional latent variables. The most obvious alternative explanation for the partial misfit is a "method effect" (also called a "method artifact"). Detecting a method effect requires us to look more closely at the model's measurement error terms (also referred to as uniquenesses).

In the standard specification of a one-factor CFA model, the error terms for each item are assumed to be uncorrelated, because the relationship between the items is supposed to be due entirely to the influence of the latent variable. Because the standard CFA model specifies a zero correlation between the measurement error terms, a nonzero correlation represents a source of misfit. We reestimated the one-factor CFA model using an alternative parameterization that allows for the possibility of correlated measurement errors. Our inspection of the modification indices confirmed the presence of correlated errors. We observed no obvious substantive relationship between the items with correlated errors, but several of the items were located adjacent to one another on the survey instrument. We investigated the possibility of a method effect due to respondents answering adjacent items on the survey in similar ways. This type of method effect is similar to the notion of autocorrelation in temporal and spatial research, where observations in proximity to one another are more similar to one another than other pairs of nonproximate observations.

Our results indicate that 4 of the 11 correlations between error terms associated with adjacent items are significantly different from zero (Items 6a and 6b, 6d and 6e, 6e and 6f, and 6f and 6g in Table 5). Items 6a and 6b appeared adjacent to one another on the survey instrument, and both items concern making decisions fairly and treating inmates fairly. For these 2 items, respondents were presented the words *fair* and *fairly*. The other 4 items (6d, 6e, 6f, and 6g) do not refer to fair treatment directly but ask more specific questions about correctional officer decision making. These items ask about the application of rules, making decisions based on facts, and giving honest explanations for actions. These 4 items appear to measure important aspects about how correctional officers make decisions and explain actions. The 2 items that follow (6h and 6i) are not closely correlated with the previous 4 likely because they measure consistent treatment across different inmates, not methods for making decisions. In these instances there may be some "redundancy effect" whereby respondents respond in similar ways to items they feel they have already answered (Bradburn, 1983; Luzzo, 1993).

The revised model fits the data very well, considerably better than the previous specification without correlated errors (χ^2 = 51.6, *df* = 26, *p* < .0020; CFI = .972; TLI = .987; RMSEA = .068; WRMR = .630). This model represents our final one-factor model of perceived procedural justice. Item-level statistics—including standardized factor loadings (λ), standard errors (*SE*), and explained variances (R^2)—for the final model are presented in Table 5.

TABLE 5:   Final Confirmatory Factor Analysis (CFA) Model of Procedural Justice

Variable	λ	SE	R^2
Q6a. How often do correctional officers make decisions about how to handle problems in fair ways?	.577	.048	.333
Q6b. How often do correctional officers treat inmates fairly?	.843	.029	.711
Q6c. How often do correctional officers treat inmates with dignity and respect?	.830	.028	.689
Q6d. How often do correctional officers accurately understand and apply the rules?	.664	.043	.441
Q6e. How often do correctional officers make decisions based on facts not their personal biases or opinions?	.642	.043	.412
Q6f. How often do correctional officers try to get the facts in a situation before deciding how to act?	.650	.050	.422
Q6g. How often do correctional officers give honest explanations for their actions?	.636	.046	.405
Q6h. How often do correctional officers apply rules consistently to different people?	.352	.062	.124
Q6i. How often do correctional officers treat everyone equally?	.603	.047	.364
Q6j. How often do correctional officers respect inmates' rights?	.767	.033	.589
Q6k. How often do correctional officers give inmates the chance to express their views before making decisions?	.634	.047	.401
Q6l. How often do correctional officers treat inmates politely?	.723	.039	.522

To illustrate the utility of paying careful attention to measurement, we contrast our CFA-based measure of procedural justice with an additive index computed from the same items. Recall that using additive indices to measure procedural justice is the modal approach in the literature. Additive indices are sometimes warranted; the extent to which they are appropriate for a given problem depends on how well the items meet two key assumptions. First, additive indices assume that all items have an equivalent relationship (as evidenced by equal factor loadings) with the latent variable. Second, they also assume zero measurement error (and therefore, by default, no correlated errors). The evidence here does not support either assumption. The factor loadings listed in Table 5 are not equivalent, ranging from .352 to .830. The assumption of zero measurement error is clearly incorrect, with R^2 values for the indicators ranging from .124 to .711. Moreover, our analysis found that these measurement errors were correlated, possibly due to a method effect. When we regress our CFA-based measure of procedural justice on the additive index measure, we obtain a standardized regression coefficient of .448 and an R^2 of .201. Thus, the additive index measure of procedural justice explains only 20.1% of the variance in our CFA-based measure *of the same concept*.

Another way of thinking about the performance of our CFA measure relative to an additive index is to examine the relationships between these measures and other theoretically relevant variables. For instance, consider the relationship between age and perceptions of procedural justice. Rebellion against authority is a frequent theme among teens and young adults. As people age, they mature and develop more balanced perspectives on authority. Therefore, we might expect to find a positive relationship between age and perceptions of procedural justice. To test this assertion, we regressed (separately) our two measures of procedural justice—one derived from CFA and the other an additive index—on the age of the inmate. Age exerts a nonsignificant effect ($\beta = -.125$) on the additive measure of procedural justice; the sign of the coefficient is negative, and the explained variance is only .016. Based on this analysis, we would conclude that age is unrelated to perceptions of procedural justice in this sample of inmates. However, when we use the CFA-based measure

of procedural justice, the results are very different. Now age exerts a significant positive effect (β = .358) and explains 12.8% of the variance in perceptions of procedural justice. We provide this brief example to illustrate the point that paying attention to measurement is not an esoteric or irrelevant concern. It is a vital part of theory testing that is paid short shrift in most procedural justice scholarship.

DISCUSSION

The implications of procedural justice are appealing, and empirical tests have, thus far, offered consistent support. The underlying ideas contend that authorities can increase the likelihood of gaining cooperation, support, and compliance from the community by using procedures that signal that citizens are valued members of society. Procedures that send these messages are believed to entail quality treatment, quality decision making, and citizen participation opportunities. The existing body of research has supported the hypothesized relationships, but it has not paid sufficient attention to the measurement properties of key concepts, such as procedural justice and legitimacy. The body of existing research demonstrates that inconsistent sets of items and dimensions are often used to measure these concepts. The purpose of the analysis reported here was to provide new evidence about the measurement of procedural justice in a correctional setting.

Results demonstrate that a two- or three-dimensional model of procedural justice, similar to the typical description of procedural justice, is not supported. The data appear to fit a one-dimensional model. Two pieces of evidence suggest that researchers should be cautious when using summated scales to measure procedural justice in the future. First, the factor analytic results suggest that individual survey items do not make equal contributions to scales, and individual item error terms are correlated. Both of these findings suggest a summated index is inappropriate. Second, the summated procedural justice index and the factor analytic scale, which both seek to measure the same concept, are not closely correlated (r^2 = .201). Similarly, the relationships between age and the two measures of procedural justice show substantial differences. The relationship between age and the factor analytic measure of procedural justice is positive and statistically significant; the relationship when using the summated index measure is negative and not statistically significant. The results from previous research that measures procedural justice with a summated scale should be viewed with caution.

The findings reported here and elsewhere (Maguire et al., 2008) suggest it is important to more completely examine the measurement properties of procedural justice and legitimacy. This may proceed by utilizing data collection methods such as surveys that attempt to measure multiple concepts discussed in the procedural justice literature, including legitimacy, motive-based trust, and distributive and procedural justice. These concepts are nearly always measured with survey methods. It may be the case there are methods effects operating such that the survey itself is influencing response patterns that generate results showing the observed relationships (Campbell & Fiske, 1959; Luzzo, 1993).

Given the possibility that methods effects may be influencing substantive relationships, future investigations into the measurement properties of key concepts in procedural justice scholarship can make advancements by estimating a multitrait multimethod matrix (MTMM) (Campbell & Fiske, 1959). This approach recognizes the possibility that using a single method

to measure independent and dependent variables may generate correlations that are due to the measurement approach itself, not simply because the latent constructs are closely related. For instance, the analysis reported here uncovered meaningful correlations between some items that appeared adjacent to one another on the survey. These correlated items were worded in similar ways, raising the possibility of a "redundancy effect" (Bradburn, 1983). MTMM estimates discriminant and convergent validity through the measurement of theoretically distinct and similar concepts with multiple measurement methods. The patterns of correlations revealed in the MTMM are a reflection of the strength of validity.

MTMM seems applicable for assessing the measurement properties of distinct concepts, such as distributive and procedural justice, motive-based trust, and legitimacy. For MTMM, these concepts should be measured using multiple methods, one of which would be commonly used survey items. For instance, procedural justice could be measured using (a) paper-and-pencil surveys that contain traditionally used Likert scale responses and (b) interviews using a series of open-ended items that are analyzed with concept mapping techniques (Jackson & Trochim, 2002; Trochim, 1989). Concept mapping is a method used to analyze responses to open-ended survey items. Respondents participate in the text-coding process by sorting statements believed to fit together into categories. Multidimensional scaling and cluster analysis are then used to identify thematic clusters (Jackson, Mannix, Peterson, & Trochim, 2002; Jackson & Trochim, 2002). This approach seems applicable because it "can lead to theorizing about scale subdimensions or uncover theoretical areas that need more investigation. Concept mapping can also be used to generate items and identify dimensions in the process of scale development" (Jackson & Trochim, 2002, p. 332). Another possibility is to use a single survey that contains different question methods. In this instance, the mode of administration is the same but by the question format represents distinct methods. A recent MTMM utilized different question formats on certified public accountant (CPA) examinations to measure similar and distinct concepts (Pitoniak, Sireci, & Luecht, 2002).

In addition to assessing measurement properties, future research can make advances by giving substantive attention to the potential effects of fair procedures and legitimacy on order in correctional environments. Given the unique importance of achieving and maintaining order in these settings, procedural justice theory holds special appeal to those followers of the Parsonian solution to challenges presented to the established order, such as exists in the correctional setting. Parsons (1951) provided a challenge to the Hobbesian solution to managing order through the complete domination of force by contending that order is best achieved by the union of shared norms and values. In this case, by treating inmates with a sense of respect, by way of fundamental fairness, order within the correctional setting may be more amenable to a consensus between the state and the incarcerated. Meaningful research in the area of procedural justice and legitimacy must remain cognizant of the reality of confinement and the perceptions of the incarcerated human psyche, as well as the ability of both to influence the overall perception of the enactment of justice.

REFERENCES

Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research*. New York: Academic Press.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Butler, M., & Drake, D. H. (2007). Reconsidering respect: Its role in Her Majesty's Prison Service. *Howard Journal of Criminal Justice*, *46*, 115-127.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Casper, J. D., Tyler, T., & Fischer, B. (1988). Procedural justice in felony cases. *Law & Society Review*, *3*, 438-508.

DuIulio, J. (1987). *Governing prisons*. New York: Free Press.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*, 149-161.

Hicks, A. J., & Lawrence, J. A. (2004). Procedural safeguards for young offenders: Views of legal professionals and adolescents. *Australian and New Zealand Journal of Criminology*, *37*, 401-418.

Jackson, K. M., Mannix, E. A., Peterson, R. S., & Trochim, W. (2002, June). *A multi-faceted approach to process conflict*. Paper presented at the annual conference of the International Association for Conflict Management, Salt Lake City, UT.

Jackson, K. M., & Trochim, W. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods*, *5*, 307-336.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. *Journal of Applied Psychology*, *76*, 698-707.

Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum.

Lind, E. A., & Earley, P. C. (1992). Procedural justice and culture. *International Journal of Psychology*, *27*, 227-242.

Lind, E. A., & Tyler, T. R. (1988) *The social psychology of procedural justice*. New York: Plenum.

Luzzo, D. A. (1993). A multi-trait, multi-method analysis of three career path development measures. *Career Development Quarterly*, *41*, 367-345.

Maguire, E. R., & Johnson, D. (2009). *Measuring public perceptions of the police: Dimensionality and discriminant validity*. Unpublished manuscript.

Maguire, E. R., Johnson, D., & Kuhns, J. (2008). *The meaning and measurement of legitimacy in procedural justice scholarship*. Unpublished manuscript.

Muthén, B. O., du Toit, S. H. C., & 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., & Muthén, B. O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

Parsons, T. (1951). *The social system*. London: Routledge & Kegan Paul.

Paternoster, R., Bachman, R., Brame, R., & Sherman, L. W. (1997). Do fair procedures matter? The effects of procedural justice on spouse assault. *Law & Society Review*, *31*, 163-204.

Pitoniak, M. J., Sireci, S. G., & Luecht, R. M. (2002). A multitrait-multimethod validity investigation of scores from a professional licensure examination. *Educational and Psychological Measurement*, *62*, 498-516.

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

Reisig, M. D., & Lloyd, C. (2009). Procedural justice, police legitimacy, and helping the police fight crime. *Police Quarterly*, *12*, 42-62.

Reisig, M. D., & Mesko, G. (2009). Procedural justice, legitimacy, and prisoner misconduct. *Psychology*, *Crime*, *and the Law*, *15*, 41-59.

Sparks, J. R., & Bottoms, A. E. (1995). Legitimacy and order in prisons. *British Journal of Sociology*, *46*, 45-62.

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

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. New York: John Wiley.

Trochim, W. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, *12*, 1-16.

Tyler, T. R. (1984). The role of perceived injustice in defendants' evaluations of their courtroom experience. *Law and Society Review*, *18*, 51-74.

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

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

Tyler, T. R. (2003). Procedural justice, legitimacy, and the effective rule of law. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 30, pp. 431-505). Chicago: University of Chicago Press.

Tyler, T. R. (2004). Enhancing police legitimacy. *The Annals of the American Academy of Political and Social Science*, *593*, 84-99.

Tyler, T. R. (2005). Policing in Black and White: Ethnic group differences in trust and confidence in the police. *Police Quarterly*, *8*, 322-342.

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

Tyler, T. R., Sherman, L., Strang, H., Barnes, G. C., & Woods, D. (2007). Reintegrative shaming, procedural justice, and recidivism: The engagement of offenders' psychological mechanisms in the Canberra RISE drinking-and-driving experiment. *Law and Society Review*, *41*, 553-586.

Tyler, T. R., & Smith, H. (1997). Social justice and social movements. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 595-629). New York: McGraw-Hill.

Tyler, T. R., & Wakslak, C. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority. *Criminology*, *42*, 253-281.

Useem, B., & Kimball, P. (1989). *States of siege: U.S. prison riots 1971-1986*. New York: Oxford University Press.

Useem, B., & Reisig, M. (1999). Collective action in prisons: Protests, disturbances, and riots. *Criminology*, *37*(4), 735-759.

Woolf, L. J. (1991). *Woolf report: Prison disturbances April 1990*. London: Her Majesty's Stationery Office.

Howard Henderson is an assistant professor in the College of Criminal Justice at Sam Houston State University. He received his PhD in criminal justice from Sam Houston State University in 2006. His research interests are focused on offender risk assessment instruments and corrections at the community and institutional levels.

William Wells is an associate professor in the College of Criminal Justice at Sam Houston State University. He received his PhD from the University of Nebraska at Omaha in 1999. His research interests include police innovations, gun-related violence, and criminal justice responses to gun-related problems.

Edward R. Maguire is an associate professor of justice, law, and society in the School of Public Affairs at American University. He received his PhD in criminal justice from the State University of New York at Albany in 1997. His professional interests cover a wide range, but most of his work focuses on three topics: police organizations, violent crime, and social science measurement.

Jameel Gray is a college instructor in the Criminal Justice Department at Aurora University. He received his MA in administration of justice from Southern Illinois University–Carbondale. He is also a program coordinator with the federally funded Project Safe Neighborhoods program aimed at reducing gun and gang violence in the city of Chicago.