

The Structure of Citizen Perceptions of Crime and Disorder: New Insights from a Caribbean Community

Edward R. Maguire¹ · Todd Armstrong² · Devon Johnson³

Published online: 27 May 2016
© Springer Science+Business Media New York 2016

Abstract

Objectives This study builds on existing research from US cities on the construct and discriminant validity of perceptual measures of crime and disorder. It seeks to determine whether citizens distinguish between crime and disorder.

Methods This study draws on quantitative and qualitative data from a high-crime community in Trinidad and Tobago, a small-island developing nation in the eastern Caribbean. Analysis of the quantitative data relies on exploratory and confirmatory factor analysis methods designed for use with categorical observed variables and continuous latent variables.

Results In contrast to previous research, we find that citizens do distinguish between physical disorder and general crime, but there is a perceptual overlap for some drug-related offenses and types of social disorder.

Conclusions This study raises questions about the external validity of research on the relationship between perceptions of crime and disorder conducted in the US, and contributes to ongoing discussions and debates about the meaning of disorder. The findings suggest the need for theory and research to explain how context shapes not only the magnitudes of these perceptions, but also their structures. The results also demonstrate the benefits of mixed-methods research approaches in this area of study.

Keywords Disorder · Incivilities · Crime · Measurement · Discriminant validity · Caribbean

✉ Edward R. Maguire
edmaguire@asu.edu

¹ School of Criminology and Criminal Justice, Arizona State University, 411 N. Central Ave, Suite 600, Phoenix, AZ 85004, USA

² Sam Houston State University, Huntsville, TX, USA

³ George Mason University, Fairfax, VA, USA

Introduction

The relationship between crime and disorder is a fundamental yet unresolved question in criminology. Broken windows theory, first explicated by Wilson and Kelling (1982), suggests that disorder leads to crime. The causal relationship specified by the theory implied an attractive public policy solution for addressing crime: by adopting strategies intended to reduce disorder, communities could reduce crime. Criminologists and other social theorists were quick to scrutinize both the theory and the public policies flowing from it. Some have tested the theoretical relationship between disorder/incivilities and crime, with many finding little evidence of a statistical relationship between measures of these phenomena (Bogges and Maskaly 2014; Brown et al. 2004; Doran and Lees 2005; Gault and Silver 2008; Markowitz et al. 2001; Perkins et al. 1992; Perkins and Taylor 1996; Sampson and Raudenbush 1999; Skogan 1990; Taylor 2001; Xu et al. 2005; Yang 2010). Others have debated whether the implementation of policing strategies aimed at reducing disorder—variously labeled order maintenance, quality of life, zero tolerance, or broken windows policing—was responsible for reducing crime (Braga et al. 2015; Kubrin et al. 2010; Rosenfeld et al. 2007; Weisburd et al. 2015). Still others have criticized either the theory or the policies based on it from more normative grounds, focusing on its likelihood to undermine people's civil rights (Harcourt 2001).

Although the balance of the evidence suggests there may not be a causal link between disorder and crime as posited by Wilson and Kelling, scholars have questioned the validity of the measurement of crime and disorder commonly used in this line of research.¹ Much of the research on the relationship between crime and disorder has relied on citizen survey data to form measures of perceived disorder (for an exception see Sampson and Raudenbush 1999).² Taylor (1999) argues that when disorder is measured using citizen survey data, researchers should demonstrate that citizen perceptions of disorder are distinct from their perceptions of crime. If these perceptions are not distinct, then part of the association between perceived disorder and crime found in empirical tests of this relationship can be attributed to the inability of citizens to discriminate between these two phenomena. To the extent that measurement is a concern, the conclusions drawn from past research about the causal relationship between disorder and crime may be inaccurate.

Previous assessments of the discriminant validity of perceived disorder and crime have yielded mixed evidence. Existing research has relied on samples drawn from within the United States and has not yet focused on communities with a pronounced crime problem (Armstrong and Katz 2010; Gau and Pratt 2008; Gau and Pratt 2010; Ross and Mirowsky 1999; Worrall 2006). In an effort to expand the scope and external validity of this body of research, the present study examines the construct and discriminant validity of perceived crime and disorder, using survey data from a high crime community located in Trinidad and Tobago.

¹ Though beyond the scope of this study, the definition, conceptualization, and measurement of “disorder” in criminological research remains unresolved (Johnson et al. in press; Kubrin et al. 2010; Ross and Mirowsky 1999). Scholars do not use a consistent label (i.e. disorder, incivilities, neighborhood problems), and there is not a clearly agreed upon definition or consistently applied framework for measuring disorder (Skogan 2015).

² Measures of disorder created from citizen surveys are based on subjective perceptions of disorder rather than objective conditions in the community. Perceptual measures of disorder are most appropriate for tests of broken windows theory, as residents' perceptions of, and responses to, neighborhood conditions are a central aspect of the theory (Gau and Pratt 2010). Alternatively, the use of objective measures of disorder may be better matched to other research questions.

Background

In the relationship between disorder and crime proposed by Wilson and Kelling (1982), perceived disorder is thought to lead to crime when it increases fear of crime among residents and citizen withdrawal from public spaces. Citizen withdrawal then leads to reduced informal social control, which in turn emboldens offenders. Perceiving less informal social control, offenders engage in more crime, leading to additional withdrawal from public spaces and an increase in serious crime. Although Wilson and Kelling focused on robbery as the final outcome, other crimes, like prostitution and drug sales, are postulated to generate effects akin to those of disorder (including citizen withdrawal from public areas and reduced informal social control). Less serious crime is part of the explanation for more serious crime.

Given this specification, it is not surprising that less serious forms of crime, such as prostitution and drug use/sales, are often included in measures of disorder (e.g. Perkins and Taylor 1996; Perkins et al. 1993; Skogan 1990), as overlap between disorder and crime is inherent in broken windows theory. However, the inclusion of less serious forms of crime in measures of disorder lends weight to concerns about the discriminant validity of measures of perceived crime and disorder.

Discriminant validity, popularized through Campbell and Fiske's (1959) multitrait-multimethod matrix technique for assessing construct validity, requires that measures of different concepts computed using a single measurement method be empirically distinguishable from each other. If there is substantial overlap between perceptions of disorder and perceptions of crime, these perceptual measures may have weak discriminant validity. Concerns regarding the discriminant validity of perceptual measures of disorder and crime may be exacerbated when both concepts are measured using the same survey instrument, as the shared method can increase the strength of the estimated relationship between them (Campbell and Fiske 1959). Thus, using the same survey instrument to measure perceptions of crime and disorder may artificially inflate the estimated relationship between the two constructs, making it more difficult to distinguish empirically between them.

The discriminant validity of perceptual measures of disorder and crime may also be influenced by context. Wilson and Kelling (1982: 32) described a causal link between disorder and crime in neighborhoods where "property is abandoned, weeds go up...fights occur...an inebriate slumps to the sidewalk and is allowed to sleep it off." Clearly this passage does not describe much of suburban or middle-class America. Similarly, areas where the disorder/crime link is most salient are described as being 'vulnerable to criminal invasion' (Wilson and Kelling 1982: 32). The description of the link between disorder and crime offered by Wilson and Kelling suggests that the two phenomena are more likely to emerge as separable and functionally distinct in areas with elevated levels of one or both, where increases in disorder lead to increases in serious crime. And indeed, research provides support for this proposition. Based on data from a large rural and semi-urban district in Washington State, Gau and Pratt (2010) found that residents who live in communities with low levels of disorder are not able to differentiate disorder from crime, whereas residents residing in more disorderly neighborhoods *do* make a distinction. They write: "as people saw more disorder in their neighborhood, their ability to discern more routine signs of disorder from actual instances of crime improved. People who did not consider disorder to be very problematic in their neighborhoods, on the other hand, tended to collapse the two conditions into a single one" (Gau and Pratt 2010: 763).

The argument that the discriminant validity of perceived measures of crime and disorder may vary with the extent to which these issues are salient in the lives of citizens is similar to one invoked in recent research on perceptions of police service quality and legitimacy in low and high crime communities. These studies suggest that certain perceptual constructs may only emerge as empirically separable when the issues they address are highly salient (Johnson et al. 2014; Maguire and Johnson 2010). These findings are potentially instructive for research on perceptions of crime and disorder. In communities where crime and disorder have low issue salience, residents may think of them together as part of one perceptual package. In neighborhoods where crime and disorder are high and signal threats to safety, residents may be “cognitively and emotionally on a heightened state of alert...and particularly sensitive to and attuned to those events that might indicate a risk of potential harm” (Innes 2005: 21). As a result, residents in these communities may have more nuanced perceptions of disorder and crime and may be more likely to differentiate between them.

The Discriminant Validity of Perceptual Measures of Crime and Disorder

A small body of research has examined the discriminant validity of measures of perceived crime and disorder using citizen survey data. Ross and Mirowsky (1999) tested the relationships between perceptions of crime and disorder, as well as perceived social and physical disorder using survey data from 2482 respondents in Illinois households. They found that perceptions of pure physical “decay” (e.g. abandoned buildings) were distinguishable from perceptions of “disorder” (e.g. people hanging out, drug use, alcohol use) though the correlation between the factors was strong and a few items cross-loaded on both factors (e.g. graffiti, noise, trash, vandalism). The authors speculated that the cross-loadings may have emerged because: “These cues are physical, but they indicate the presence of people. They indicate social disorder and physical decay” (Ross and Mirowsky 1999: 423). Ross and Mirowsky also found that perceived crime loaded strongly on the latent disorder factor, indicating that perceptions of crime and disorder were indistinguishable in their sample of Illinois residents.

Worrall (2006) tested the discriminant validity of perceptual measures of crime and disorder with survey data from approximately 14,000 respondents across 12 cities where police departments were practicing community policing. Exploratory and confirmatory factor analyses provided mixed results. Worrall (2006) tested the discriminant validity of incivilities measures relative to citizen perceptions of crime. He found that residents’ perceptions of perceived crime were not distinct from their perceptions of either physical or social incivilities. On the basis of these findings, Worrall (2006: 379) called for additional research on “whether people can separate perceptions of crime from perceptions of incivilities.”

Gau and Pratt (2008: 171) tested the relationship between citizen perceptions of crime and disorder with data from a sample of the “21 largest municipalities in the 20 counties in eastern Washington.” The results of both exploratory and confirmatory factor analyses showed that perceptions of crime and disorder were not empirically distinguishable. Gau and Pratt reported a strong, positive correlation ($r = .92$) between perceived crime and disorder, a finding that signaled weak discriminant validity. They concluded that “perceptions of crime and perceptions of disorder seem to constitute a single latent construct” (p. 179).

Armstrong and Katz (2010) tested the discriminant validity of perceptual measures of crime and disorder using survey data from a sample of 800 citizens residing in Mesa,

Arizona. An exploratory factor analysis found that citizen reports of vandalism loaded more strongly on a perceived crime factor than on a disorder factor. Similarly, citizen perceptions of assault loaded more strongly on a disorder factor than on a crime factor (measures of crime with high loadings on the crime factor included burglary, car theft, and robbery). The findings from confirmatory factor analyses provided equivocal evidence supporting discriminant validity. Two-factor models had a marginally better fit than one-factor models, but the overall fit of both models remained questionable. Based on their results, Armstrong and Katz (2010: 302) called for additional research “in an attempt to determine whether or not more conclusive evidence for the discriminant validity of perceptions of incivilities relative to perceptions of crime may be developed.”

Together, these studies show that the discriminant validity of perceptual measures of disorder remains unresolved. Consistent with Wilson and Kelling’s specification, each of these studies included citizen perceptions of less serious crime in measures of disorder. Exploring the fit of this specification in a wider variety of contexts may reveal additional insights about the relationships between perceptions of crime and disorder, and the implications of these relationships for the conceptual meaning of disorder. The findings from these studies may also be influenced by the nature of the samples. Armstrong and Katz (2010) relied on a sample drawn from Mesa, Arizona, a middle class community. Gau and Pratt (2008) conducted their analysis using data from municipalities in eastern Washington. They argued that broken windows theory was never intended to encompass only urban communities, thus their sample consists of a mix of rural, suburban and urban community residents. Although we do not disagree with their argument about the scope of broken windows theory, we do suspect that it may be more difficult for citizens to distinguish between disorder and crime if they live in communities where one or both phenomena have low base rates and are not very salient, as demonstrated in a later study by Gau and Pratt (2010).

To a certain extent, Worrall’s (2006) research addresses this concern by analyzing data from twelve US cities.³ While these cities include high crime areas, they also include low crime areas, as is evident in the descriptive findings. Of the 13,918 completed surveys, roughly 70 % of respondents indicated that they were not aware of crimes being committed in their neighborhood. However, the estimates Worrall presents are pooled, with factor structures and parameter estimates constrained to be equal across all twelve cities. Separate factor analyses for each city would likely have revealed different factor structures. Alternatively, even if the factor structures were the same, a multiple group analysis that allowed the factor loadings and correlations between factors to vary across cities may have revealed substantial variation that would be masked in a pooled analysis. Thus it remains possible that in distressed communities with high base rates of crime and disorder, perceptual measures of these phenomena may be more conceptually and measurably distinct.

The Present Study

In the current study, we build on analyses reported by Ross and Mirowsky (1999), Worrall (2006), Gau and Pratt (2008, 2010) and Armstrong and Katz (2010). We draw on quantitative survey data from a high-crime community in Trinidad and Tobago, a small island

³ The cities in the data used by Worrall (2006) were Chicago, IL; Kansas City, MO; Knoxville, TN; Los Angeles, CA; Madison, WI; New York, NY; San Diego, CA; Savannah, GA; Spokane, WA; Springfield, MA; Tucson, AZ; and Washington, DC.

developing nation in the eastern Caribbean, to examine the construct and discriminant validity of perceptual measures of crime and disorder. Carrying out a study in this setting is consistent with recent studies that underscore the importance of examining the generalizability of theories and research findings in different contexts, including developing nations (Johnson et al. 2014; Kochel et al. 2013; Reisig and Lloyd 2009; Tankebe 2009). By expanding the scope of this research beyond the US, the present study helps to broaden the external validity of this line of research.

Data and Methods

The Research Setting

The Republic of Trinidad and Tobago is a small, two-island developing nation in the eastern Caribbean, about seven miles northeast of Venezuela. With the discovery of oil in 1910, Trinidad became one of the most prosperous nations in the Caribbean. Trinidad and Tobago obtained its independence from Great Britain in 1962 though it remains a member of the Commonwealth of Nations and British influence is evident in many sectors. From 1999 to 2008, Trinidad and Tobago suffered a 480 % increase in homicides, from 93 in 1999, to 540 in 2008. Maguire et al. (2008) found that most of the increase was due to homicides by firearm and was associated with the spread of gang warfare, much of which was concentrated in the disadvantaged hillside communities surrounding the capital city of Port of Spain.

In response to the rising homicide rate, the government of Trinidad and Tobago launched numerous data collection initiatives to diagnose the nation's crime problem. One initiative was the IMPACT Study, a series of citizen surveys in Belmont, a community in East Port of Spain particularly affected by the rise in gang-related violence. Belmont is well-known for its problems with disorder and crime; these problems are particularly concentrated in the Gonzales section of Belmont, an area facing more extreme environmental conditions due to its squatter population. Indeed, portions of Gonzales are essentially a shantytown, where homes are built of makeshift materials and residents have limited access to water, electricity, and other utilities. In addition to these utility and infrastructure problems, residents in Belmont and Gonzales experience high levels of unemployment and underemployment, and crime and violence represent significant concerns for the community. As one rough indicator of the conditions under which the residents of this community live, 64.4 % of respondents to our wave 1 survey reported having heard gunshots in the last 30 days, including 86.3 % of those from Gonzales and 41.5 % of those from other areas in Belmont.

Data

We rely primarily on quantitative survey data on perceptions of crime and disorder drawn from waves 1 and 2 of the IMPACT Study. A local research firm conducted face-to-face interviews with 1200 randomly-selected residents (approximately 600 in each wave).⁴

⁴ For purposes of a quasi-experiment not relevant to the present analysis, the IMPACT study design involved partitioning Belmont residents into two groups for sampling purposes: those who live in Gonzales (denoted from this point forward as “Gonzales”), and those who live elsewhere in Belmont (denoted from this point forward as “Belmont”). In Belmont, sampling was proportional to the size of the population, using

Interviews were completed from June 18 to August 12, 2006 for wave 1 and from July 6 to August 28, 2007 for wave 2. The AAPOR Response Rate #1 was 79 % for wave 1 (81 % in Belmont, 76 % in Gonzales) and 84 % for wave 2 (86 % in Belmont, 81 % in Gonzales; American Association for Public Opinion Research 2008). Descriptive statistics for the two samples are shown in Table 1.

The IMPACT survey covered numerous topics, including community cohesion, fear of crime and victimization, perceived crime and neighborhood problems, and attitudes toward the police. The instrument was carefully constructed based on a review of the relevant literature and focus groups in the community of study. Many of the survey items were drawn from previous research and the questionnaire was reviewed by local professionals to ensure that its terminology was appropriate for Trinidadian language and culture, especially for use in communities with low literacy.⁵ The instrument was further refined after pre-testing with a small sample.

Our quantitative analyses are supplemented by qualitative data drawn from interviews and focus groups conducted between 2005 and 2008 with police officers, community leaders, and neighborhood residents. We use the qualitative data to contextualize and interpret the findings from the quantitative analyses. We conducted three focus groups with community residents in January 2008 to better understand their perceptions of disorder and other neighborhood conditions (cf. Kubrin et al. 2010). Participants were recruited by a local research firm using a convenience sample; two groups included Gonzales residents (N = 13 for the young adult group, N = 16 for the adult group), and one group included young adults residing in other areas of Belmont (N = 10). The focus groups were audiotaped and lasted about one and a half hours each. The audiotapes were transcribed by a Trinidadian native living in the US and the transcripts were coded and analyzed using NVivo for relevant themes.

Measurement Strategy

Our general measurement approach was to treat respondents' answers on 22 individual survey items as indicators of one or more continuous latent variables representing perceptions of disorder and crime. These items were grouped into two sections on the questionnaire, and the indicators were ordinal categorical variables with either three or five

Footnote 4 continued

community boundaries based on census files from the Trinidad and Tobago Central Statistical Office. The sampling boundaries for Gonzales were based on those identified by community residents because the boundaries of Gonzales are debated. The official boundaries from the Port of Spain Corporation and Central Statistical Office used for statistical purposes are smaller than the boundaries identified by community residents (see Pride in Gonzales Committee (2005) *Gonzales Community Profile, East Port of Spain, Trinidad*, "4.1.1 Population Size & Growth"). Gonzales was then split into eight zones (chosen to reflect smaller neighborhoods within the community), and the sample was drawn proportional to the population in each zone. In order to select respondents, GIS maps for each area were generated showing roads and housing. A start house was located and a sampling interval calculated so that interviewers canvassed every "nth" house from the start location. Once the household was identified, adult respondents within each household were selected using the "last birthday" method to ensure that the probability of selecting an individual within the household was the same for all eligible respondents. If selected participants were not at home at the time of the visit, interviewers made three call backs before the case was coded as a non-response using AAPOR final distribution code 2.25 (American Association for Public Opinion Research 2008).

⁵ English is the official language in Trinidad, so no language translation of survey items was necessary. However, colloquial terms differ across cultures, and we wanted to capture these in the survey. For example, when asking respondents about "truancy," interviewers may have also used the phrase "breaking biche," which is a common term for skipping school in Trinidad.

Table 1 Sample characteristics

	Wave 1 (N = 602, %)	Wave 2 (N = 598, %)
Gender		
Male	45.5	47.3
Female	54.5	52.7
Race/ethnicity		
African	68.1	68.8
East Indian	5.7	6.4
Afro/Indian	25.6	19.3
Other	0.7	5.5
Age		
18–24	19.9	19.7
25–34	14.8	22.4
35–44	16.7	23.1
45–54	17.6	16.7
55–64	14.8	9.2
65+	16.2	8.9
Educational attainment		
None	0.5	0.5
Primary	30.5	21.8
Junior secondary	3.0	3.2
Secondary	54.6	54.0
Technical/vocational	5.7	7.6
Tertiary/university	5.7	13.0
Employment status in the last 4 months		
Unemployed—no income	3.1	6.3
Unemployed—government income	3.6	2.2
Unemployed—family income	21.9	10.1
Employed part time/odd jobs	12.9	14.7
Employed full time	38.2	55.3
Retired—pension income	20.2	11.4
Marital status		
Single, never married	39.4	47.0
Living with someone, but not married	21.4	17.5
Married	23.7	24.4
Divorced/separated	7.0	5.9
Widowed	8.5	5.2
Called or approached the police for assistance in the last 4 months		
Yes	14.5	11.3
No	85.5	88.7
Years living in the community		
1–10 years	17.6	24.5
11–20 years	22.5	23.0
21–30 years	22.2	23.0
31+ years	37.7	29.5

Table 2 Indicators of disorder and crime

Variable	Range	Wave 1 mean	Wave 2 mean
q27: Trash and garbage on the sidewalks/street	1–3	2.26	2.11
q28: Graffiti on buildings and walls	1–3	1.40	1.15
q29: Vacant or abandoned houses/buildings	1–3	1.62	1.36
q30: Poor lighting	1–3	1.82	1.22
q31: Abandoned cars	1–3	1.28	1.22
q33: Empty or overgrown lots of land	1–3	1.78	1.51
q34: Groups of teenagers or adults hanging out	1–3	1.94	1.54
q35: People buying and selling drugs on the streets	1–3	2.20	1.72
q36: People drunk in public/on the street	1–3	1.68	1.32
q38: People smoking marijuana in public	1–3	2.25	1.74
q39: Loud or unruly neighbors	1–3	1.49	1.37
q49: Burglaries	1–5	2.98	1.99
q50: Robberies	1–5	3.02	1.92
q51: Assaults or violent disputes	1–5	2.80	1.82
q53: Sexual assault or rape	1–5	2.45	1.42
q54: Drug use/abuse	1–5	3.22	2.46
q55: Drug dealing/trafficking	1–5	3.21	2.49
q56: Kidnapping	1–5	2.12	1.26
q57: Homicide	1–5	2.75	1.70
q58: Gangs	1–5	3.27	2.20
q59: Gang-related crime	1–5	3.27	2.07
q60: Domestic violence/child abuse	1–5	2.02	1.54

categories. In the first series of questions respondents were asked: “Now I’m going to read a list of things that are problems in some neighborhoods. For each, please tell me if it is a big problem, somewhat of a problem, or not a problem in your neighborhood.” Response options for these items (q27–q39) were coded: 1 = not a problem, 2 = somewhat of a problem, and 3 = a big problem. In the second series of questions respondents were asked: “The next group of questions is also about problems that might affect your community. Please tell me how serious the following problems are in your community.” Response options for these items (q49–q60) were coded: 1 = not at all serious, 2 = not too serious, 3 = somewhat serious, 4 = very serious, and 5 = extremely serious. Means for the 22 disorder and crime items in wave 1 and wave 2 are presented in Table 2.

We adopted a four-step approach to examining the latent structure of the survey items. First, we estimated confirmatory factor analysis (CFA) models using the wave 1 data to test the fit of two popular specifications that are consistent with the way perceptions of disorder and crime are conceptualized and measured in the literature. This step was not focused on model-building or revision. Instead it was intended to test the construct validity of existing conceptualizations of perceived disorder and crime used routinely by scholars. The remaining steps would have been unnecessary if one of these models fit the data well, but that was not the case. Since the first step did not provide clear evidence about the underlying factor structure of the data, in the second step we used exploratory factor

analysis (EFA) to examine the dimensionality of the perceived crime and disorder items in the wave 1 data. This revealed a number of insights about the underlying structure of these items. In step three, we used the information from the EFA to specify and test a new CFA model using the wave 1 data. That model fit the data well, and it is that model on which we base our discussion of the findings. In step four, we confirmed the fit of the CFA model from step three using the wave 2 data. This step served as a check to ensure that our previous model-fitting efforts did not capitalize on statistical chance (MacCallum et al. 1992).⁶

Findings

Step 1

We estimated a confirmatory factor analysis (CFA) using wave 1 data to test the fit of two model specifications popular in the literature. Although scholars have conceptualized and measured perceived disorder and crime inconsistently, a close review of past studies suggests that two model specifications have garnered the most theoretical and empirical support: (1) a three-factor model comprised of physical disorder, social disorder, and crime; and (2) a two-factor model comprised of disorder and crime (Armstrong and Katz 2010; Gau and Pratt 2008; Perkins et al. 1992; Ross and Mirowsky 1999; Sampson and Raudenbush 1999; Skogan 1990; Worrall 2006). In both specifications, the latent variables were permitted to be correlated with one another given the expected non-zero relationships between disorder and crime.

Since all items are ordinal, we relied on a robust, mean and variance adjusted weighted least squares (WLS) estimator available in the structural equation modeling software Mplus (Muthén and Muthén 2010). Monte Carlo simulations have found that the robust WLS estimator performs well in models with categorical outcomes, including those with skewed distributions (Beauducel and Herzberg 2006; Flora and Curran 2004; Muthén et al. 1997; Rhemtulla et al. 2012).⁷ Goodness of fit for all models was evaluated using multiple measures, including Chi square (χ^2), the root mean square error of approximation

⁶ Some may question the wisdom of starting with a CFA and then reverting to EFA. We viewed it as important to give existing conceptualizations the opportunity to “succeed” in an initial CFA before embarking on our own EFA analysis that starts from scratch. This process may have been less defensible if we had only one data set, since it involves considerable revisions to the model based on the data. However, the presence of a second independent data set to be used only for model confirmation makes this approach significantly more defensible. A recent review of current methodological considerations in EFA and CFA concluded that “it is reasonable to follow up a poor-fitting CFA model with an EFA” (Schmitt 2011, p. 315).

⁷ Many of the procedures used in normal theory EFA and CFA with continuous indicators need to be adapted for use with categorical indicators. The EFA and CFA methods used in this study are able to accommodate both binary and ordinal indicators based on survey items with different response options (Muthén 1983, 1984). For both the EFA and CFA, we rely on a latent response variable formulation that treats each ordinal response y_i as a crudely categorized approximation of an underlying continuous variable y_i^* . According to Brown (2006, p. 390), “the underlying y_i^* variables are related to the 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_i^* 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.” Although factor analysis and item response theory (IRT) models are often viewed as alternative or competing methodologies for the analysis of binary or ordinal items, the CFA approach used here is equivalent to an IRT modeling approach. (e.g., Brown 2006; Kamata and Bauer 2008; Muthén and Asparouhov 2002; Takane and DeLeeuw 1987).

(RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). For EFA models we added the standardized root mean square residual (SRMR), and for CFA models we added the weighted root mean square residual (WRMR; Brown 2006; Browne and Cudeck 1993; Hu and Bentler 1999; Muthén and Muthén 2000; Schreiber et al. 2006; Yu 2002).⁸

We began by estimating the three-factor model using the wave 1 data with 22 items assigned to load on either physical disorder (q27–q33), social disorder (q34 and q36–q39), or crime (q35 and q49–q60). The model fit the data poorly according to three of the five fit measures ($\chi^2 = 2095.8$, $df = 206$, $p < .000$; RMSEA = .123; CFI = .965; TLI = .960; WRMR = 2.43), thus we chose to reject this specification.⁹ Next, we estimated the two-factor model with items assigned to load on either generalized disorder (combining the items used in the physical and social disorder dimensions in the previous analysis) or crime. Once again, the model fit the wave 1 data poorly according to three of the five fit statistics ($\chi^2 = 2276.9$, $df = 208$, $p < .000$; RMSEA = .129; CFI = .961; TLI = .957; WRMR = 2.56) and therefore we rejected this specification.

In both cases, we could have used information like standardized residuals or modification indices to adjust the poorly fitting CFA models. According to Brown (2006, p. 159), these types of information “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 mis-specifications. When the initial model is grossly mis-specified, specification searches are not nearly as likely to be successful....” Inspection of the modification indices and standardized residuals revealed no clear pattern useful for re-specifying the CFA models to achieve better fit to the data. In such instances, it is typically more efficient to work *forwards* from an exploratory factor analysis rather than work *backwards* from an incorrectly specified CFA model (Asparouhov and Muthén 2009; Brown 2006; Browne 2001).

Step 2

Next, we carried out an exploratory factor analysis (EFA) on the 22 items using the wave 1 data. Once again, since all the items are ordinal, we relied on the same robust weighted least squares (WLS) estimator (Muthén and Muthén 1998–2007). This investigation is

⁸ There are no “golden rules” of model fit. Simulation studies routinely cast doubt on rules of thumb that are considered sacrosanct among applied researchers (e.g., Chen et al. 2008; Gerbing and Anderson 1993; Hu and Bentler 1999; Marsh et al. 2004; Yu 2002). Evaluating model fit is a holistic exercise that involves weighing the strengths and weaknesses of various fit indices. Here our appraisals of model fit are informed by the following considerations. We examine the p value associated with the mean and variance adjusted χ^2 . A non-significant χ^2 ($p > .05$) is an indicator of good fit, though the test is often considered too strict, with minor deviations indicating poor model fit. However, Bowen and Guo (2012: 145) note that “it is widely considered acceptable to conclude that a model fits the data well even if the value [of χ^2] is statistically significant” if other fit indices suggest good fit. For RMSEA, Browne and Cudeck (1993) conclude that values of .06 to .08 constitute acceptable fit, while values of .01 to .06 constitute “close fit.” Hu and Bentler (1999) also treat a RMSEA value of .06 as the upper threshold for close fit. For CFI and TLI, Hu and Bentler (1999) suggest that values of .95 or greater indicate close fit. Experts have recommended different threshold values for SRMR; Muthén and Muthén (2000) suggest that SRMR should be below .05 or .06, while Brown (2006) and Hu and Bentler (1999) recommend a threshold of .08. For WRMR, simulation evidence suggests that values below 1 are indicative of good fit (Yu 2002). Significantly more simulation research is needed to test the performance of these fit indices under a variety of conditions, including alternative estimators, models with different levels of complexity, and items with different distributional characteristics.

⁹ Only the TLI and CFI suggest that the model fits the data. The values of the Chi square (χ^2), RMSEA, and WRMR measures all suggest a poor fit.

concerned with discriminant validity, thus the correlation(s) between factors (if more than one factor is identified) are the principal parameters of interest. For that reason, we chose an oblique rotation method (Geomin) that allowed the correlation(s) between factors to be freely estimated. We chose Geomin over other oblique rotation methods on the basis of simulation evidence which reports that it provides “the most promising rotation criterion when little is known about the true loading structure” (Asparouhov and Muthén 2009, p. 16).¹⁰

We drew on multiple criteria for determining the number of factors to retain, including checking the factor solutions for interpretability and examining all of the fit indices introduced earlier (Brown 2006; Browne and Cudeck 1993; Cattell 1966; Guttman 1954; Hu and Bentler 1999; Kaiser 1960; Muthén and Muthén 2000; Yu 2002). Simulation evidence reveals that parallel analysis is a useful method for determining the optimum number of factors to extract, thus we also carried out a parallel analysis (Garrido et al. 2016; Timmerman and Lorenzo-Seva 2011; Weng and Cheng 2005).¹¹ Based on these various considerations, we found that the three-factor solution fit the data best ($\chi^2 = 1001.4$, $df = 168$, $p < .000$; RMSEA = .091; CFI = .973; TLI = .963; SRMR = .039), though there was clear room for improvement in model fit. Factor loadings from the 3-factor model are shown in Table 3.

The results of the initial EFA point to some interesting findings. The three factors are clearly interpretable. All six of the items intended to measure perceptions of physical disorder loaded cleanly on Factor 1, although the loading for q27 (trash and garbage) barely met our cutoff of 1.401. All four of the items intended to measure perceptions of social disorder loaded cleanly on Factor 2, although the loading for q39 (loud or unruly neighbors) also fell just above the cutoff of 1.401. Eleven of the twelve items intended to measure perceptions of crime loaded strongly on Factor 3; q35 (people buying and selling drugs on the street) loaded instead on Factor 2. Two of the items with primary loadings on Factor 3 also had cross-loadings on factor 2, including q54 (drug use/abuse) and q55 (drug dealing/trafficking).¹²

We interpret Factor 1 as a physical disorder factor since it contains items intended to measure physical disarray in a neighborhood and is consistent with past theoretical and empirical models. One item most commonly used to measure physical disorder in previous research (q27: Trash and garbage in the streets) had a weak loading (just above .40) on

¹⁰ At least one previous study that used EFA to examine the discriminant validity of perceptual measures of disorder and crime used Varimax rotation, an orthogonal rotation method that fixes the correlations between factors at zero. Varimax rotation is the default setting in the factor analysis routines of some statistical software packages and therefore it may sometimes be selected without careful consideration of the appropriate rotation method for a given research question (Fabrigar et al. 1999). While orthogonal rotation methods are useful for some research questions, oblique rotations should be used when studying discriminant validity, since the correlations between factors are the principal parameters of interest in such studies (Brown 2006; Kline 2013). Rotation is not used in CFA, so this issue is not pertinent to the CFA portions of our analysis.

¹¹ Parallel analysis relies on random data “with the same number of observations and variables as the original data. The correlation matrix of the random data is used to compute eigenvalues. These eigenvalues are compared to the eigenvalues of the original data. The optimum number of factors is the number of the original data eigenvalues that are larger than the random data eigenvalues” (Muthén and Muthén 1998–2015: 611). Since parallel analysis is not available in Mplus when using weighted least squares estimation with categorical data, we used a robust maximum likelihood estimator for this analysis. We examined both mean eigenvalues and 95th percentile eigenvalues from the random data. Both methods were consistent in suggesting that three factors should be extracted.

¹² Items were designated as having cross-loadings if they loaded at .4 or higher on one factor and at more than half their primary loading value on one or more other factors.

Table 3 EFA factor loadings (Wave 1)

	Factor 1: Physical disorder	Factor 2: Drugs/social disorder	Factor 3: General crime	Cross- loading?
q27: Trash and garbage on the sidewalks/street	0.407*	0.095	0.036	
q28: Graffiti on buildings and walls	0.545*	0.224*	0.154*	
q29: Vacant or abandoned houses/buildings	0.623*	0.281*	−0.024	
q30: Poor lighting	0.469*	0.171*	−0.016	
q31: Abandoned cars	0.718*	0.150	0.024	
q33: Empty or overgrown lots of land	0.742*	−0.011	−0.068	
q34: Groups of teenagers or adults hanging out	0.029	0.788*	−0.091	
q35: People buying and selling drugs on the streets	−0.008	0.783*	0.013	
q36: People drunk in public/on the street	0.157*	0.598*	0.050	
q38: People smoking marijuana in public	0.014	0.881*	−0.008	
q39: Loud or unruly neighbors	0.164*	0.407*	0.156*	
q49: Burglaries	0.209*	0.064	0.690*	
q50: Robberies	0.189*	0.026	0.751*	
q51: Assaults or violent disputes	0.247*	0.007	0.731*	
q53: Sexual assault or rape	0.282*	−0.057	0.745*	
q54: Drug use/abuse	0.157*	0.401*	0.505*	✓
q55: Drug dealing/trafficking	−0.058	0.444*	0.598*	✓
q56: Kidnapping	0.183*	−0.048	0.779*	
q57: Homicide	0.027	−0.013	0.857*	
q58: Gangs	−0.016	0.158*	0.821*	
q59: Gang-related crime	−0.014	0.140*	0.815*	
q60: Domestic violence/child abuse	0.096*	−0.001	0.713*	

Factor loadings greater than |.40| are listed in bold, and factor loadings that are statistically significant at the .05 level are marked with an asterisk. Items were designated as having cross-loadings if they loaded at |.40| or higher on one factor and at more than half the value of their primary loading on one or more other factors

Factor 1.¹³ Our qualitative data suggest that this issue may have a different meaning for some residents than the other items intended to measure physical disorder. The majority of residents in our focus groups viewed the significant problem of trash and garbage in the street as the result of poor and inconsistent garbage removal services; they blamed the government for this problem, and did not see it as a sign that residents did not care about

¹³ In several of the exploratory and confirmatory factor analyses we estimated, the loading for this item fell below our threshold of .4. For instance, in Table 4 we report results from a multiple group CFA in which the loading for this item fell below .4 for residents of Gonzales but not for residents of Belmont.

the neighborhood. Focus group participants explained that trash pickup is irregular, garbage trucks are not able to access all areas of the community due to the hilly terrain and lack of passable roads, the number and size of trash bins in the community is inadequate, and the many stray dogs in the neighborhood spread trash around the streets. As one young person from the Belmont group stated: “The garbage people real inconsistent. It will have garbage just pile up, pile up....” Similarly, a young adult resident from Gonzales noted: “If they come and pick up the garbage when they supposed to, the dogs will not have a chance to scatter it.” A few focus group members also attributed the trash problem to neighborhood residents, as a comment made by an adult resident of Gonzales suggests: “Yeah, but we are the ones who put the garbage there. Some people, they see an abandoned vacant lot, they throw the garbage there. They just don’t care.” The variation in views expressed during the focus groups may help account for the low factor loading.

Our results differ from those of Ross and Mirowsky (1999) who found that abandoned buildings loaded cleanly on a physical decay factor (which they consider the “purest physical disorder” measure) while items such as graffiti and vandalism (which they note are “physical cues that indicate the presence of people”) cross-loaded on both their decay and their disorder factor. To explain the emergence of the distinct decay factor, Ross and Mirowsky posited that residents may perceive abandoned buildings to be a problem associated with absentee landlords and therefore may not link them to a lack of social control in the community. For residents in the Trinidadian community we studied, however, the presence of vacant buildings *does* signal a breakdown in neighborhood social control, and therefore it is understandable why this item loads with the other physical disorder items. Data from our qualitative interviews and focus groups indicated that vacant buildings (and sometimes abandoned cars) were a significant concern to residents because they provided places for gangs to hide weapons and drugs and to evade police. As an adult in the Gonzales focus group noted: “Right now the abandon house is very useful because...the gunmen and them, that’s their hideout.” Another adult from Gonzales agreed: “When you’re passing at night and you see an abandoned house, and there’s somebody right in there waiting for you to pass—people who do drugs....” Taken together, these results suggest that residents’ perceptions of what is disorderly may vary across different contexts (Johnson et al. in press).¹⁴ As Korbin (2008: 207) reminds us, “disorder” is a social construction, and researchers need to pay attention to “residents’ subjective meanings of disorder in their communities.”

Factor 2 includes items that have been used to measure social disorder in past studies, but also includes items that measure a specific crime type: drug crime (including drug use, drug abuse, and drug sales). Substantively, this perceptual overlap makes sense. Groups of teenagers and adults hanging out or “liming”¹⁵ might be indicative of social disorder, but it may also indicate the presence of an illegal drug market.¹⁶ Indeed, the items loading on Factor 2 tap into the visible “street life” in this community—young men, many of whom are in gangs and are selling drugs, hanging out on the streets, drinking alcohol, smoking

¹⁴ Perceptions of what is disorderly also vary across individuals within the same context (see, for example, Hinkle and Yang 2014; Wallace et al. 2015).

¹⁵ The term “liming” in Trinidad and Tobago is similar to the term “hanging out” in the US. “Lime” is also used as a noun to indicate a party or social event.

¹⁶ Focus group respondents noted that some of the individuals hanging out on the street may act as gatekeepers for entry into the community (particularly in gang areas). As one young person from Belmont noted “Well, one thing about Belmont, you could walk through and people will be liming. They will watch you and ask you, ‘You from Belmont?’ ‘Yeah.’ ‘Well, go on your way.’ But once you from outside, or they don’t know your face, well....”

marijuana, often serving as lookouts for gangs, and occasionally committing acts of violence. Our qualitative observations are also consistent with this interpretation. For instance, many times when we entered certain parts of the community while observing police officers on patrol, we saw lookouts scrambling to notify others that the police were in the area. Residents told us that these notifications were typically followed up by young men in the area rapidly disarming themselves and hiding their drugs and weapons until the police left the area. One community leader who observed this post-notification process take place described it as highly scripted and well-practiced. Thus, it is understandable why these items hang together in the minds of residents.

One item used in previous research to measure social disorder (q39: Loud or unruly neighbors) had a weak loading (just slightly above .40 on Factor 2. Our qualitative data suggest that residents in this community may not share a common perception that loud neighbors are atypical or problematic. For example, when we asked focus group members if loud or unruly neighbors were a problem in their neighborhood, one youth from Belmont noted: “That is Belmont...that is normal...every Saturday morning they playing Mavado¹⁷ loud...it’s not really a problem....”

We interpret Factor 3 as a general crime factor comprised of items that measure more serious crime types. Two items had strong loadings on Factor 3 but cross-loadings with Factor 2: Drug use/abuse (q54) and drug dealing/trafficking (q55). Given that many of the items from Factor 2 focus on drug-related issues, this cross-loading pattern is substantively meaningful. Residents in our focus groups commented on the presence of “sprangers” who are engaged in crime to feed their drug habit. One young person from the Belmont focus group said: “The problem I see is not necessarily the marijuana, but you see the people smoking cocaine, they tend to steal a lot and take your stuff. That’s the problem.” Another agreed: “Sprangers, they are a part of the neighborhood, you can’t have a neighborhood without sprangers. That is Belmont security and neighborhood watch...they all watch it. They working better than URP.¹⁸ At any hour you could get a computer from a spranger, cell phone, clothes....” As a result of the relationship between perceptions of drug use and crime, we retain these items for the next step in the analysis and account for their substantive overlap in the model. In moving to the next stage of analysis, we have 22 items measuring three conceptually meaningful perceptual constructs: physical disorder, drugs/social disorder, and general crime.

Step 3

Earlier, in step one, we tested two confirmatory factor analysis (CFA) models consistent with the literature on disorder and crime and found that they fit the data poorly. As a result, in step two we estimated an exploratory factor analysis (EFA) model and found that a solution containing three factors fit the data best. Now we return to the CFA modeling framework to test a three-factor model derived from the EFA in step two. The model specifies a physical disorder factor measured using six items (q27–q30, q31, and q33), a drugs/social disorder factor measured using five items (q34–q36, q38, q39), and a general crime factor measured using eleven items (q49–q51 and q53–q60). This model also includes two items (q54: Drug use/abuse, and q55: Drug dealing/trafficking) that cross-load on the drugs/social disorder and crime factors. Allowing these two items to load on both

¹⁷ Mavado is a reggae recording artist from Jamaica.

¹⁸ The URP is the Unemployment Relief Program, a government program that provides brief periods of employment for jobless citizens.

factors accounts for the substantively meaningful cross-loadings observed in the EFA findings from step two. This initial CFA model did not fit the data very well ($\chi^2 = 1191.5$, $df = 204$, $p < .000$; CFI = .982; TLI = .979; RMSEA = .090; WRMR = 1.74).

Next, to explore local areas of misfit between the model and the data, we examined the modification indices. One clear pattern is immediately apparent: two of the highest modification index (MI) values represent error correlations between pairs of adjacent survey items measuring related crime types: q49 (burglaries) and q50 (robberies) [MI = 376.4]; and q58 (gangs) and q59 (gang-related crime) [MI = 244.6]. In a standard CFA model, the fundamental assumption is that the influence of the factors accounts for the correlations between items; once the effects of the factors are partialled out, these correlations should not be significantly different from zero. Here the modification indices tell a different story. The correlations between these items are greater than zero even after controlling for the influence of the factors. Thus, freely estimating the error correlations between these three pairs of adjacent items (rather than fixing them at zero, which is the usual default) will significantly improve the fit of the model, a pattern also observed in previous research by Gau and Pratt (2008). On the basis of our qualitative data, we also free the error correlation for a pair of non-adjacent items: q29 (vacant or abandoned housing or buildings) and q33 (empty or overgrown lots of land) [MI = 37.1]. Local residents attribute meaning to both types of properties, envisioning them as dangerous and unsightly places where they do not want their children to play. Freeing this error correlation acknowledges the likelihood that residents attribute particular meaning to these types of locations beyond simply being physically disorderly.

It is not clear whether the correlations between adjacent items are due to a methodological artifact or a more substantively meaningful explanation. When survey items share semantic content (for instance, by both referencing gangs), a methodological artifact can result in which the correlation between items is due partially to the factor and partially to the shared content. This is similar to a “response set” in which respondents respond the same way to blocks of items in sections of a survey, except in this case, they may be responding the same way to adjacent or non-adjacent items with shared semantic content. At the same time, these correlations may be substantive, due to a more complex factor structure in which a series of primary factors accounts for most of the item-level correlations, but one or more minor factors also account for a portion of these correlations (e.g., Chen et al. 2006). For instance, a dominant crime factor may account for most of the correlations between items intended to measure crime, but a part of those correlations may also be due to minor factors for property crime (q49 and q50) and gangs (q58 and q59). Unfortunately, both scenarios result in equivalent models, thus without additional data we are unable to distinguish whether these item-level correlations are due to either methodological or substantive causes. Freeing the error correlations between these pairs of items accounts for either possibility. The revised model fits the data considerably better ($\chi^2 = 612.4$, $df = 201$, $p < .000$; CFI = .992; TLI = .991; RMSEA = .058; WRMR = 1.16).

Based on some of the differences between the two areas where we carried out our surveys (Gonzales, and portions of Belmont outside of Gonzales), we also tested the fit of a multiple-group CFA model that allowed some of the model parameters to vary by neighborhood.¹⁹ A Chi square difference test revealed that constraining the factor loadings

¹⁹ This analysis relied on the Delta parameterization in Mplus, which fixes scale factors at one and factor means at zero in the first group and frees these parameters to be estimated in the second group (Muthén and Muthén 2010). The multi-group model includes additional free parameters and is complex relative to the sample size, thus we were somewhat concerned about the stability of the parameter estimates. Our

Table 4 Standardized CFA factor loadings for multiple-group model (Wave 1)

Factors and items	Belmont	Gonzales
Physical disorder		
q27: Trash and garbage on the sidewalks/street	0.697	0.389
q28: Graffiti on buildings and walls	0.912	0.832
q29: Vacant or abandoned houses/buildings	0.768	0.825
q30: Poor lighting	0.536	0.591
q31: Abandoned cars	0.890	0.783
q33: Empty or overgrown lots of land	0.601	0.585
General crime		
q49: Burglaries	0.868	0.789
q50: Robberies	0.873	0.822
q51: Assaults or violent disputes	0.907	0.841
q53: Sexual assault or rape	0.893	0.866
q54: Drug use/abuse	0.548	0.671
q55: Drug dealing/trafficking	0.727	0.766
q56: Kidnapping	0.898	0.855
q57: Homicide	0.901	0.839
q58: Gangs	0.894	0.910
q59: Gang-related crime	0.903	0.900
q60: Domestic violence/child abuse	0.854	0.749
Drugs/social disorder		
q34: Groups of teenagers or adults hanging out	0.641	0.832
q35: People buying and selling drugs on the streets	0.930	0.958
q36: People drunk in public/on the street	0.803	0.732
q38: People smoking marijuana in public	0.907	0.850
q39: Loud or unruly neighbors	0.804	0.651
q54: Drug use/abuse (cross-loading)	0.455	0.276
q55: Drug dealing/trafficking (cross-loading)	0.275	0.176

and thresholds to be equal across groups worsened fit relative to a model in which factor loadings and thresholds were freely estimated in both groups ($\chi^2 = 277.24$, $df = 62$, $p < .000$). The multiple group model with freely estimated factor loadings and thresholds

Footnote 19 continued

inspection of the estimates from both the single-group and multiple-group models alleviated our concerns about this issue. The parameter estimates from the single-group model (for which we do not have concerns about model complexity relative to sample size) represent a compromise between the estimates from the multiple group model. For example, in the multiple-group model, the correlations between crime and drugs/social disorder are .658 in Belmont and .767 in Gonzales. In the single-group model, the correlation is .695. The multiple-group parameter estimates are similar to the single group estimates, suggesting that in spite of model complexity, the estimates are stable.

fit the data reasonably well ($\chi^2 = 960.5$, $df = 402$, $p < .000$; RMSEA = .068; CFI = .991; TLI = .990).²⁰ Table 4 lists the standardized factor loadings for Belmont and Gonzales from this multiple-group model.

Step 4

Because we relied on the wave 1 data to inform our stepwise model modification process (through the use of EFA results and modification indices), one natural question that arises is whether the final model presented in step three may have resulted from adjusting the model to fit “small idiosyncratic characteristics of the sample” (MacCallum et al. 1992: 501). Thus, to confirm the fit of the final CFA model tested in step three, we fit the same model to the wave 2 data, which were not used as a source of information during the model revision process. The model fit the data well without any modifications ($\chi^2 = 835.0$, $df = 400$, $p < .000$; RMSEA = .060; CFI = .986; TLI = .984), suggesting that the wave 1 findings were not sample-specific.

Since this study focuses largely on discriminant validity, the principal parameters of interest are the correlations between factors. Table 5 shows the correlations from wave 1 and wave 2 for Belmont and Gonzales. The correlations between perceptions of physical disorder and crime are strong and positive in both groups and both waves of data. These four correlations range from .552 to .667, with a mean of .629, well below the customary threshold ($r \geq .85$) for inferring discriminant validity problems (Brown 2006). The same pattern holds for the correlations between physical disorder and drugs/social disorder (range .580–.803; mean .713) and between general crime and drugs/social disorder (range .658–.767; mean .700). Although the correlations between these constructs are high, they are lower than those reported in most of the previous research. For instance, Gau and Pratt (2008) found a correlation of .92 between measures of crime and disorder and concluded that discriminant validity for measures of perceived crime and disorder is low. Thus, unlike much of the previous research from the United States, our evidence from two waves of data from residents in a high crime Caribbean community suggests that while perceptions of

Table 5 Correlations between CFA factors (Waves 1 and 2)

	Physical disorder	General crime	Drugs/social disorder
Wave 1 results			
Physical disorder	1.0	.663*	.787*
General crime	.634*	1.0	.767*
Drugs/social disorder	.803*	.658*	1.0
Wave 2 results			
Physical disorder	1.0	.667*	.580*
General crime	.552*	1.0	.689*
Drugs/social disorder	.680*	.684*	1.0

Results from Belmont are presented in the lower portion of each correlation matrix; results from Gonzales are presented in the upper portion. Correlations that are statistically significant at the .05 level are marked with an asterisk

²⁰ The Mplus technical support team advises against the use of WRMR in multiple-group models, therefore we only report WRMR for single-group models (Muthén 2013).

physical disorder, crime, and drugs/social disorder may overlap considerably, some aspects of these phenomena are empirically separable.

Discussion

Based on a debate in the criminological literature about the relationship between perceived disorder and crime, as well as concerns about the measurement of these key concepts, this study sought to examine the construct and discriminant validity of perceptual measures of both phenomena from a high crime community in Trinidad and Tobago. We began our analysis by testing model specifications consistent with two popular conceptualizations of disorder and crime: a three factor model containing perceptual measures of physical disorder, social disorder, and crime; and a two factor model containing perceptual measures of disorder and crime. Both models fit the data poorly, so we conducted exploratory factor analyses meant to illuminate the dimensionality of the items. Based on findings from the EFA, we then tested a series of confirmatory factor analysis models. In the end, a three factor model containing measures of physical disorder, drugs/social disorder, and general crime fit the data well across both waves of the survey. These three factors had strong, positive correlations with one another, but the correlations were below the threshold typically used to infer discriminant validity problems.²¹ We discuss the implications of these findings for previous research on the discriminant validity of perceived disorder and crime, and for ongoing conceptual and theoretical debates about the meaning of disorder in the sections below.

Scholars examining the discriminant validity of perceived crime and disorder have concluded that the relationship between perceived crime and disorder is inestimable because of discriminant validity problems. For example, Worrall (2006: 376) concluded from his study of twelve communities that “respondents were unable to distinguish between perceived crime and either physical or social incivilities.” Similarly, Gau and Pratt (2008: 179) found that “perceptions of crime and perceptions of disorder seem to constitute a single latent construct.” Further, Armstrong and Katz (2010) found that citizen perceptions of crime were not distinct from their perceptions of disorder.

Contrary to this research evidence from a mix of US communities, our study of a high-crime community in a developing nation identified empirically distinct measures of perceived disorder and crime, raising questions about the external validity of past research. Our confirmatory factor analysis produced three distinct factors: a physical disorder factor, a general crime factor, and a social disorder/drugs factor. In light of these results, we conclude that perceptions of *physical* disorder and general crime are empirically separable. However, the emergence of a third perceptual dimension, consisting of drug-related crime and social disorder, suggests that citizens do not always conceive of crime and *social* disorder as distinct phenomena.

Our measure of drugs/social disorder combines behaviors that are criminal and those that are non-criminal, but troublesome to residents. The emergence of this dimension challenges the notion that disorder and crime are completely separable phenomena in the minds of citizens; instead, it appears there is significant perceptual overlap in certain domains. Residents appear to view these behaviors as part of a single perceptual package, either because these actual behaviors tend to coexist in the community, and/or because residents perceive them as co-existing. This interpretation is consistent with St. Jean’s

²¹ The mean factor correlation across waves and communities was .68. The threshold typically used to infer a problem with discriminant validity is usually either .80 or .85.

(2007) findings from Chicago that the conditions responsible for generating social disorder also tend to attract the sale and use of illegal drugs. In a circular fashion, the drug trade may then amplify the level of social disorder in the community. Under such conditions, it is not surprising that residents would view social disorder and drug use as overlapping.

We can only speculate about why our findings differ from those of earlier studies. One obvious explanation is that the differences in results are attributable to important differences in the nature of the communities where the research was conducted. For example, our findings suggest the possibility that in communities where disorder and crime are highly salient—like the one we studied in Trinidad & Tobago—certain aspects of these two phenomena can be separated, both empirically and conceptually. Gau and Pratt's (2010) research on rural and semi-urban communities in the US is consistent with this perspective. At this stage, our interpretation of the potential role of salience is merely speculation since we do not have access to data from multiple communities where salience levels vary. If, as we hypothesize, salience does influence people's sensitivity or ability to distinguish between crime and disorder, citizens who live in communities with low or moderate base-rates of these phenomena may find it difficult or impossible to distinguish between them. This is the pattern observed by Gau and Pratt (2010) in their comparison of low- and high-disorder communities in Washington state. Our hypothesis and their results suggest the possibility of a threshold effect in which the characteristics of a community influence not only the *magnitudes* of people's perceptions, but also the *structures* of these perceptions. In terms of structure, these perceptions may be unidimensional in contexts where crime and disorder are not particularly salient issues in the daily lives of residents, but multidimensional in communities where residents live with a heightened sense of alert that enhances or focuses their perceptual sensitivity.

Our findings are sufficiently different from those of other discriminant validity studies that it seems prudent for scholars to begin exploring the structure of public perceptions of disorder and crime in communities with a variety of different characteristics. To test hypotheses about the role of salience in shaping perceptions of disorder and crime, scholars should carry out research in neighborhoods or communities where the level and salience of disorder and crime vary. Multi-level and multi-group models of public perceptions across communities (and across nations) would seem especially helpful in light of evidence that these perceptions may be shaped—both in structure and in magnitude—by community characteristics (Gau and Pratt 2010; Sampson and Raudenbush 2004).²²

It is also possible that the differences between our findings and those of previous research are a function of the methodology used—in particular the survey questions and the range of responses. This possibility is an important one, and the extent to which the discriminant validity of perceptual measures of disorder is influenced by methodology should be carefully addressed in future studies. As scholars have argued, the operationalization of disorder will likely vary depending on the research setting and purpose of the study (Johnson et al. in press; Gau and Pratt 2010; Skogan 2015). Moreover, as Wallace et al. (2015: 258) note, survey questions commonly used to measure perceived disorder and crime by asking respondents if certain phenomena are “a problem in one's neighborhood” may conflate the presence or absence of a phenomenon with interpretive assessments about the phenomenon, which may influence tests of discriminant validity. Finally, given that our

²² While we have offered differences in the salience of crime and disorder as a key distinguishing feature, clearly there are other important differences between Trinidad and Tobago and the United States (both at the community level and the national level) that may contribute to the differences between our findings and those of earlier studies.

measures of perceived crime and disorder both come from the same data source, we cannot rule out the possibility of monomethod bias (see Piquero 1999). However, it is worth noting that monomethod bias would have the effect of *increasing* both the correlations between items and the likelihood of finding that perceived crime and disorder are empirically inseparable. Thus, our finding that some aspects of perceived crime and disorder are separable runs counter to the likely effects of monomethod bias.

In addition to raising questions about the external validity of previous research based on data from the US, this study also contributes to ongoing discussions about the meaning of disorder. For example, note that our measure of perceived drugs/social disorder includes behaviors that are illegal but not generally considered to be behaviors that are *mala in se* or necessarily wrong in and of themselves and those that are legal, but disruptive. This is consistent with the specification of broken windows theory offered by Wilson and Kelling (1982) where behaviors such as these are deleterious to the social fabric of a neighborhood when they lead to more serious crime as a result of citizen withdrawal from public spaces. Within this specification, some less serious forms of crime have more in common with indicators of serious social disorder than they have in common with serious criminal activity. This finding is worthy of consideration in future discussions and debates about the conceptual meaning of disorder.

Another of our findings contributes to an emerging theme in studies of disorder. Most discussions of physical disorder tend not to distinguish between whether the source of the disorder is internal or external to the community. Yet, our qualitative data suggests that many residents in the Trinidadian community we studied took into account the *source* of a problem when determining if they perceived it to be disorderly or not (Johnson et al. in press). For example, many residents did not view garbage in the streets in the same light as other indicators of physical disorder which might signal a breakdown of social control. Instead, they viewed garbage in the streets as a sign that the government did not care enough about them or their community to provide regular garbage pick-up or a sufficient number of garbage bins. This may help explain why the trash item had a low factor loading on our physical disorder measure.

This finding is consistent with St. Jean's (2007: 40–41) observations from Chicago: "To many Wentworth residents, the presence of trash on the street often means the absence of sufficient trash cans in an area, or the sporadic services provided by the city's streets and sanitation department." Ross and Mirowsky (1999: 423) noted a similar theme with regard to absentee landlords: "maybe people understand that an abandoned, vacant, or run-down building in their neighborhood... is often a consequence of an absentee landlord who never visits the neighborhood, and the presence of these buildings is not a direct indicator of the breakdown of social control in their neighborhood." Recent research on the impact of foreclosures on perceptions of disorder during the housing crisis further suggests that residents make important distinctions between internal and external causes of disorder and decay in their neighborhood. As Wallace et al. (2012: 643) explain: "It might be the case that individuals do not see foreclosures as a neighborhood problem per se, but instead contextualize it as a problem many neighborhoods and homeowners are experiencing across the nation. In this sense, foreclosures are not perceived as 'deviant' or 'disorderly' but rather as something that can happen to anyone in any neighborhood."

More research on residents' causal explanations for neighborhood problems may help clarify conceptualizations of disorder. St. Jean (2007) argues that the conceptualization of disorder in the broken windows thesis suffers from a middle class bias. He notes that typical indicators of disorder have multiple meanings, particularly for residents of communities characterized by concentrated disadvantage. For such residents, whether physical

conditions of the community were internally or externally generated is an important distinction. That this same distinction was made by residents in Chicago and in a disadvantaged community in Trinidad is instructive, and speaks to the potential universality of the meaning of disorder. As scholars continue to refine the conceptualization and operationalization of disorder, it would be wise to incorporate qualitative approaches that can help contextualize the findings of the quantitative analyses that tend to dominate this line of research. Such approaches can contribute significantly to our understanding of these phenomena.

A Few Cautions

Given the polemics of the disorder/crime debate, we suggest that great care be exercised when considering the implications of our findings for theory and practice. Since this analysis was primarily concerned with the capacity of citizens to distinguish between crime and disorder, we are unable to contribute directly to the theoretical debate about the *causal* relationship between disorder and crime as outlined in broken windows theory. However, as Gau and Pratt (2008: 181) point out: “disorder cannot cause crime if disorder *is* crime.” Thus, demonstrating the empirical separability or discriminant validity of perceptual measures of crime and disorder is a crucial precursor to the investigation of the causal relationships between them. Findings from this study show that in our sample, perceptions of physical disorder are distinct from perceptions of general crime. This criterion alone, however, is insufficient to establish causality. Our findings are limited to subjective or perceptual measures of disorder and crime and should not be confused with actual or objective measures of these phenomena. While our findings may be instructive for theorizing about the relationships between crime and disorder, nothing in our findings speaks directly to the causal relationships between them (see Sampson and Raudenbush 1999; Skogan 1990; Taylor 2001). Disentangling a causal relationship would require a very different research design than the one used here.

Similarly, while some may be tempted to draw inferences about the appropriateness of order maintenance policing from this study; we discourage anyone from making such connections. Broken windows theory has been offered as the justification for order maintenance policing. As a policy innovation, order maintenance policing may have impacts through incapacitation, deterrence, or through the proposed theoretical link between disorder and crime. At the same time, ill-conceived policing practices implemented under the banner of order maintenance policing may also have the effect of alienating citizens from the police and increasing crime through a number of potential causal pathways (Gau and Brunson 2010; Harcourt 2001; Tyler 2006). Therefore, conclusions regarding the efficacy of order maintenance or broken windows policing are best left to actual studies of the effect of policy changes (e.g. Rosenfeld et al. 2007; Weisburd et al. 2015) and should not be drawn based on the results of this research. Knowing the correlation between perceptions of disorder and crime is not sufficient to warrant conclusions about how communities should (or should not) be policed.

Conclusion

Citizen perceptions of crime and disorder in communities play an important role in multiple theoretical traditions within criminology. Thus, research intended to clarify the nature and correlates of these perceptions is a worthwhile investment for criminologists. Most

research from the US on whether citizens distinguish between crime and disorder has found that measures of these perceptions are not empirically separable. Our research from a high crime community in a developing nation in the Caribbean suggests that citizens do distinguish between some aspects of crime and disorder while other aspects tend to blend together perceptually. The challenge for future research is to continue providing greater clarity about the conceptual meaning of disorder for citizens, as well as the role of context in shaping both the magnitudes and structures of these perceptions.

Acknowledgments Funding for this research was provided by the Ministry of National Security of Trinidad and Tobago. The points of view expressed in this paper are those of the authors alone and do not represent the official policies or positions of the Ministry of National Security or the Trinidad and Tobago Police Service. This research was approved by the Human Subjects Review Board at George Mason University (HSRB Protocol #4846).

References

- American Association for Public Opinion Research (2008) Standard definitions: final dispositions of case codes and outcome rates for surveys, 5th edn. AAPOR, Lenexa, Kansas
- Armstrong T, Katz C (2010) Further evidence on the discriminant validity of perceptual incivilities measures. *Justice Q* 27:280–304
- Asparouhov T, Muthén B (2009) Exploratory structural equation modeling. *Struct Equ Model* 16:397–438
- Beauducel A, Herzberg PY (2006) On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct Equ Model* 13(2):186–203
- Boggess LN, Maskaly J (2014) The spatial context of the disorder–crime relationship in a study of Reno neighborhoods. *Soc Sci Res* 43:168–183
- Bowen NK, Guo S (2012) Structural equation modeling. Oxford University Press, New York
- Braga A, Welsh B, Schnell C (2015) Can policing disorder reduce crime? A systematic review and meta-analysis. *J Res Crime Delinq* 52:567–588
- Brown TA (2006) Confirmatory factor analysis for applied research. The Guilford Press, New York
- Brown BB, Perkins DD, Brown G (2004) Incivilities, place attachment and crime: block and individual effects. *J Environ Psychol* 24:359–371
- Browne MW (2001) An overview of analytical rotation in exploratory factor analysis. *Multivar Behav Res* 36:111–150
- Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. In: Kenneth BA, Long JS (eds) Testing structural equation models. Sage, Newbury Park, pp 136–162
- Campbell DT, Fiske DW (1959) Convergent and discriminate validation by the multitrait-multimethod matrix. *Psychol Bull* 56:81–105
- Cattell RB (1966) The scree test for the number of factors. *Multivar Behav Res* 1(2):245–276
- Chen F, West SG, Sousa KH (2006) A comparison of bifactor and second-order models of quality of life. *Multivar Behav Res* 41(2):189–225
- Chen F, Curran P, Bollen K, Kirby J, Paxton P (2008) An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociol Methods Res* 36(4):462–494
- Doran BJ, Lees BG (2005) Investigating the spatiotemporal links between disorder, crime, and the fear of crime. *Prof Geogr* 57(1):1–12
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* 4:272–299
- Flora DB, Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 9:466–491
- Garrido LE, Abad FJ, Ponsoda V (2016) Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychol Methods* 21(1):93–111
- Gau JM, Brunson RK (2010) Procedural justice and order maintenance policing: a study of inner city young men's perceptions of police legitimacy. *Justice Q* 27(2):255–279
- Gau JM, Pratt TC (2008) Broken windows or window dressing? Citizens' (in)ability to tell the difference between disorder and crime. *Criminol Public Policy* 7(2):163–194

- Gau JM, Pratt TC (2010) Revisiting broken windows theory: examining the sources of the discriminant validity of perceived disorder and crime. *J Crim Justice* 38:758–766
- Gault M, Silver E (2008) Spuriousness or mediation? Broken windows according to Sampson and Raudenbush (1999). *J Crim Justice* 36:240–243
- Gerbing DW, Anderson JC (1993) Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In: Bollen KA, Long JS (eds) *Testing structural equation models*. Sage, Newbury Park
- Guttman L (1954) Some necessary conditions for common-factor analysis. *Psychometrika* 19:149–161
- Harcourt BE (2001) *Illusion of order: the false promise of Broken Windows policing*. Harvard University Press, Cambridge
- Hinkle JC, Yang S-M (2014) A new look into broken windows: what shapes individuals' perceptions of social disorder? *J Crim Justice* 42:26–35
- Hu L, Bentler PM (1999) Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6:1–55
- Innes M (2005) Why disorder matters: antisocial behaviour and incivility as signals of risk. Social contexts and responses to risk conference, January 2005. <http://www.kent.ac.uk/scarr/events/finalpapers/Innes.pdf>
- Johnson D, Maguire ER, Kuhns JB (2014) Public perceptions of the legitimacy of the law and legal authorities: evidence from the Caribbean. *Law Soc Rev* 48(4):947–978
- Johnson D, Maguire ER, Maass SA, Hibdon J (in press) Systematic observation of disorder and other neighborhood conditions in a distressed Caribbean community. *J Community Psychol*
- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Kamata A, Bauer DJ (2008) A note on the relation between factor analytic and item response theory models. *Struct Equ Model* 15:136–153
- Kline RB (2013) Exploratory and confirmatory factor analysis. In: Petscher Y, Schatsschneider C (eds) *Applied quantitative analysis in the social sciences*. Routledge, New York, pp 171–207
- Kochel TR, Parks R, Mastrofski SD (2013) Examining police effectiveness as a precursor to legitimacy and cooperation with police. *Justice Q* 30(5):895–925
- Korbin CE (2008) Making order of disorder: a call for conceptual clarity. *Criminol Public Policy* 7(2):203–214
- Kubrin CE, Messner SF, Deane G, McGeever K, Stucky TD (2010) Proactive policing and robbery rates across U.S. cities. *Criminology* 48:57–97
- MacCallum RC, Roznowski M, Necowitz LB (1992) Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol Bull* 11:490–504
- Maguire ER, Johnson D (2010) Measuring public perceptions of the police. *Polic Int J Police Strateg Manag* 33(4):703–730
- Maguire ER, Willis JA, Snipes J, Gantley M (2008) Spatial concentrations of violence in Trinidad and Tobago. *Caribb J Criminol Public Saf* 13(1&2):44–83
- Markowitz FE, Bellair PE, Liska AE, Liu J (2001) Extending social disorganization theory: modeling the relationships between cohesion, disorder and fear. *Criminology* 39(2):293–320
- Marsh HW, Hau K-T, Wen Z (2004) In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler's (1999) findings. *Struct Equ Model* 11:320–341
- Muthén B (1983) Latent variable structural equation modeling with categorical data. *J Econom* 22:48–65
- Muthén B (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49:115–132
- Muthén L (2013) Administrator posting on the Mplus discussion board (January 4). <http://www.statmodel.com/discussion/messages/11/11403.html?1357431506>
- Muthén B, Asparouhov T (2002) Latent variable analysis with categorical outcomes: multiple-group and growth modeling in Mplus. Muthén & Muthén, Los Angeles
- Muthén LK, Muthén BO (1998–2012) *Mplus user's guide*, 7th edn. Muthén & Muthén, Los Angeles
- Muthén L, Muthén B (2000) *Mplus short courses: introductory latent variable modeling using Mplus*. Muthén & Muthén, Los Angeles
- Muthén B, du Toit SHC, Spisic D (1997) Robust inference using weighted-least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished paper
- Perkins DD, Taylor RB (1996) Ecological assessments of community disorder: their relationship to fear of crime and theoretical implications. *Am J Community Psychol* 24(1):63–107
- Perkins DD, Meeks JW, Taylor RB (1992) The physical environment of street crime: defensible space, territoriality and incivilities. *J Environ Psychol* 13:29–49

- Perkins DD, Wandersman AH, Rich RC, Taylor RB (1993) The physical environment of street crime: defensible space, territoriality and incivilities. *J Environ Psychol* 13(1):29–49
- Piquero A (1999) The validity of incivility measures in public housing. *Justice Q* 16(4):793–818
- Pride in Gonzales Committee (2005) Gonzales community profile, East Port of Spain, Trinidad. Unpublished report
- Reisig MD, Lloyd C (2009) Procedural justice, police legitimacy, and helping the police fight crime: results from a survey of Jamaican Adolescents. *Police Q* 12:42–62
- Rhemtulla M, Brosseau-Liard PE, Savalei V (2012) When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods* 17(3):354–373
- Rosenfeld R, Fornango R, Rengifo AF (2007) The impact of order-maintenance policing on New York City homicide and robbery rates: 1988–2001. *Criminology* 45(2):355–384
- Ross CE, Mirowsky J (1999) Disorder and decay: the concept and measurement of perceived neighborhood disorder. *Urban Aff Rev* 34:412–432
- Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *Am J Sociol* 105:603–651
- Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of ‘Broken Windows’. *Soc Psychol Q* 67:319–342
- Schmitt TA (2011) Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeduc Assess* 29(4):304–321
- Schreiber JB, Nora A, Stage FK, Barlow EA, King J (2006) Reporting structural equation modeling and confirmatory factor analysis results: a Review. *J Educ Res* 99(6):323–337
- Skogan W (1990) *Disorder and decline: crime and the spiral of decay in American cities*. Free Press, New York
- Skogan W (2015) Disorder and decline: the state of the research. *J Res Crime Delinq* 52(4):464–485
- St. Jean PKB (2007) *Pockets of crime: broken windows, collective efficacy, and the criminal point of view*. The University of Chicago Press, Chicago
- Takane Y, DeLeeuw J (1987) On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 52:393–408
- Tankebe J (2009) Public cooperation with the police in Ghana: does procedural fairness matter? *Criminology* 47(4):1265–1293
- Taylor R (1999) The incivilities thesis: theory, measurement and policy. In: Langworthy R (ed) *Measuring what matters: proceedings from the Policing Research Institute meetings*. National Institute of Justice, Washington, pp 65–88
- Taylor R (2001) *Breaking away from broken windows: Baltimore neighborhoods and the nationwide fight against crime, grime, fear, and decline*. Westview, New York
- Timmerman ME, Lorenzo-Seva U (2011) Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol Methods* 16(2):209–220
- Tyler TT (2006) *Why people obey the law*. Princeton University Press, Princeton
- Wallace D, Hedberg EC, Katz CM (2012) The impact of foreclosures on neighborhood disorder before and during the housing crisis: testing the spiral of decay. *Soc Sci Res* 93(3):625–647
- Wallace D, Louton B, Fornango R (2015) Do you see what I see? Perceptual variation in reporting the presence of disorder cues. *Soc Sci Res* 51:247–261
- Weisburd D, Hinkle J, Braga A, Wooditch A (2015) Understanding the mechanisms underlying broken windows policing: the need for evaluation research. *J Res Crime Delinq* 52(4):589–608
- Weng LJ, Cheng CP (2005) Parallel analysis with unidimensional binary data. *Educ Psychol Meas* 65(5):697–716
- Wilson JQ, Kelling G (1982) Broken windows. *Atl Mon* 211:29–38
- Worrall JL (2006) The discriminant validity of perceptual incivility measures. *Justice Q* 23(3):360–383
- Xu Y, Fiedler ML, Flaming KH (2005) Discovering the impact of community policing: the broken windows thesis, collective efficacy, and citizens’ judgment. *J Res Crime Delinq* 42:147–186
- Yang SM (2010) Assessing the spatial-temporal relationship between disorder and violence. *J Quant Criminol* 26(1):139–163
- Yu C (2002) *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Doctoral dissertation, UCLA, Los Angeles, CA