❧ 6 ❧

A MULTIVARIATE, MULTILEVEL RASCH MODEL WITH APPLICATION TO SELF-REPORTED CRIMINAL BEHAVIOR

Stephen W. Raudenbush*
Christopher Johnson*
Robert J. Sampson†

In studying correlates of social behavior, attitudes, and beliefs, a measurement model is required to combine information across a large number of item responses. Multiple constructs are often of interest, and covariates are often multilevel (e.g., measured at the person and neighborhood level). Some item-level missing data can be expected. This paper proposes a multivariate, multilevel Rasch model with random effects for these purposes and illustrates its application to self-reports of criminal behavior. Under assumptions of conditional independence and additivity, the approach enables the investigator to calibrate the items and persons on an interval scale, assess reliability at the person and neighborhood levels, study the correlations among crime types at each level, assess the proportion of variation in each crime type that lies at each level, incorporate covariates at each level, and

accommodate data missing at random. Using data on 20 item responses from 2842 adolescents ages 9 to 18 nested within 196 census tracts in Chicago, we illustrate how to test key assumptions, how to adjust the model in light of diagnostic analyses, and how to interpret parameter estimates.

1. INTRODUCTION

The task of combining information from multiple item responses arises widely in studies of behavior, beliefs, attitudes, exposure to risk, and symptoms of disease. In many of these studies, participants are clustered within social settings such as schools or neighborhoods, and item-level missing data are often unavoidable. To meet these challenges, we contend that a multivariate, multilevel Rasch model with random effects will often be useful. The model is based on strong assumptions. Using data from a large-scale survey of crime, we show how to test these assumptions and illustrate advantages of the approach when the assumptions are realistic.

The self-report method is widely used in the social sciences for measuring criminal involvement at the individual level. Studies of self-reported offending inquire about many aspects of delinquent and criminal behavior, including physical assaults, robbery, arson, vandalism, theft, drug sales, and fraud. In a typical survey, interviewers may ask scores of questions to obtain information about the number, severity, and types of crimes respondents have participated in during some interval of time. A common goal is to construct a relatively small number of outcome variables from a much larger number of item responses.

It is impractical and statistically unsound, however, to view *each* item in a self-report survey as an outcome: there would be as many explanatory models as items, and, assuming that small numbers of respondents endorse many of these items, multiple analyses would be statistically imprecise. Interpreting the evidence from multiple imprecise analyses would be difficult at best. Clearly, some reduction in the dimensionality of the data is essential. On the other hand, a summary that is too simple—for example, one that conceives of a single dimension of criminal behavior—may conceal important individual differences in crime.

Consider some strategies for combining information. Perhaps most common, one might count the total number of crimes

committed of a certain type for each respondent. Alternatively, one might create a binary variable indicating whether the respondent had committed one or more crimes of a certain type. Yet such *ad hoc* approaches presuppose a given dimensionality of crime; that is, they assume that the number of relevant types of crime is known *a priori*. Such approaches may also be criticized for treating more and less severe crimes equally.

Osgood, McMorris, and Potenza (2002) make a compelling argument for the application of item response theory (IRT) to combine information yielded by surveys of self-reported crime. They find that IRT scaling creates a meaningful metric that appropriately reflects the varying seriousness of criminal behavior while reducing the skewness that commonly arises in composite measures of crime. IRT provides person-specific standard errors of measurement and creates useful diagnostic information to assess the information provided by the scale as a whole and by each item. In a sequel to their paper, Osgood, Finken, and McMorris (2002) demonstrate the useful properties of IRT-based analyses based on a Tobit regression model.

The current paper builds on the work of these and other authors by extending the IRT approach in four ways that we think will be useful for the study of crime in particular, but also more generally for any phenomena (e.g., health, personality, family relationships) where the self-report method is used to construct multi-item scales.

1. There is a good reason to make the model multivariate. This move helps assess a key model assumption: unidimensionality—that is, that each IRT measure taps a single interval scale. It also enables study of the covariation of types of offending net the effects of measurement error.
2. If respondents are nested within social settings such as neighborhoods or schools, the IRT model should be multilevel. This allows the study of variation and covariation of propensity to offend at each level.
3. Models might accommodate data missing at random (MAR). Data are MAR when the probability of missingness is independent of the missing outcome given the observed data. This is a comparatively weak assumption that will be approximately correct when the observed data contain substantial information about the probability of missingness (Little and Rubin 1987).

4.   We extend the model to incorporate covariates, person and
     neighborhood characteristics that predict criminal behavior.
     Within this framework, it is possible to test whether covariates relate
     differently to different types of crime and even to different items.

Expanding the IRT model in these four ways inevitably makes
it more complex. This is particularly true if the IRT model selected is a
two-parameter model, with a location parameter and a discrimination
parameter for each item. We opt for a one-parameter or Rasch model
(Rasch 1980; Wright and Masters 1982), not only for its simplicity but
because of the interpretability of the information it yields when its
assumptions hold.

The Rasch model makes a strong assumption: that each item is
equally discriminating. When this assumption is true, the resulting
scale has several appealing features. Item location can be interpreted
as "severity," giving the scale a clear interpretation: persons scoring
higher on the scale display more severe levels of criminality than do
persons scoring lower, and the relative severity of the items is identical
for all persons. More importantly, however, we illustrate later how to
test the assumption of equal discrimination using a two-parameter
IRT model and how to use these results to diagnose item performance
and improve the scale.

Our Rasch model departs from the classical fixed-effects
Rasch model (Rasch 1980). The classical model requires complete
data and discards those respondents whose responses are all negative
(or all positive). By conceiving person propensity to be random
rather than fixed, missing data and invariant responses vectors are
accommodated. These specifications lead us to propose a multivari-
ate, multilevel Rasch model with random effects as a useful tool for
studying crime and other self-reported behavior. Our approach is
formally equivalent to a three-level hierarchical logistic regression
model. The first level entails item responses, which depend on item
difficulties and person propensities. The second level describes
variation and covariation between person propensities within
clusters (e.g., neighborhoods). The third level describes variation and
covariation between clusters.

Our approach builds on the strategy for assessing
ecological settings proposed by Raudenbush and Sampson (1999).
The current paper extends this work by showing how to test key

model assumptions, how to revise the scales and the model in light of these tests, and how to incorporate explanatory variables at each level. Moreover, this explanatory model is tailored to fit self-reported crime data, which follows a nonlinear "age-crime curve" (Gottfredson and Hirschi 1990).

The remainder of this paper is organized into five sections. Section 2 conveys our conceptual framework for measuring criminal behavior. Section 3 reviews the basic Rasch model and shows how to extend it to be multilevel and multivariate. Section 4 illustrates application to data from 2842 young people in Chicago, ages 9 to 18. It illustrates how to test key assumptions, how to modify the model and discard poorly performing items, and how to interpret all parameter estimates. Section 5 considers the incorporation of covariates at the person and neighborhood level. A final section draws conclusions and considers implications for future research.

2. CONCEPTUAL FRAME

Criminologists are not united on how to conceptualize individual differences in criminal activity. Many would agree, however, that a predatory crime occurs in time and space with the confluence of three conditions: (1) the presence of a motivated offender, (2) the presence of a viable target, and (3) the absence of capable guardians (Cohen and Felson 1979). It follows that the propensity to commit crimes will depend on personal characteristics that predispose respondents to criminal activity as well as differences in local environments associated with the availability of targets for crime and the prevalence of guardians. The distinction in criminology between propensity approaches and time-varying contextual approaches (see Cohen and Vila 1986) is central to our understanding of the measurement problem.

2.1. Personal Differences in "Propensity"

Many criminologists emphasize stable differences between persons in a global propensity to offend (e.g., see Gottfredson and Hirschi 1990; Moffitt 1993). In this view, such a propensity is a stable trait, and high levels of the trait increase the probability of virtually every type of crime, even though expression of specific types of crime might reflect

local variation in opportunities for successfully committing various crimes. An alternative view might postulate multiple traits, such as a propensity for violent crime versus property crime. In fact, the notion of individual-level specialization has not found much support in the literature (Wolfgang et al. 1972; Blumstein et al. 1986; Piquero 2000).

In opposition to the notion that criminal propensity reflects deep-rooted, stable, traitlike differences is an alternative view that propensity is contextually and temporally situated. Contextual factors might include the presence of antisocial peers (Warr 1993) and opportunities for employment or stable romantic relationships (Sampson and Laub 1993; Horney, Osgood, and Marshall 1995). Persons who are similar in stable predictors of crime such as temperament, family background, and cognitive skill may nonetheless vary substantially in their propensity to offend as a function of these contextually and temporally varying circumstances. It is, of course, difficult to separate social causation from social selection. Long-term differences in propensity to offend and contextual circumstances are likely to be reciprocally related.

On balance, the literature suggests that individual variation in the "propensity" to offend comprises a single underlying factor that is expressed in different types of criminal behavior as a function of contextually and temporally varying circumstances. This view contradicts a notion that there are deep-rooted differences between, say, burglars and robbers, though a given person may be more likely to commit burglary than robbery in a specific situation. The resulting notion of offending propensity as one-dimensional may reflect limitations in available data. Few normative samples of self-reports are large enough to reliably distinguish tendencies to commit offenses that occur rarely, such as sex offenses, from other types of crime. To say that we have a single dimension statistically does not imply that only one dimension of crime exists in reality. The data may be inadequate to reliably reveal the multidimensionality of the phenomenon.

3. THE CONTRIBUTION OF LOCAL VARIATION

Neighborhoods are important distributors of social and cultural capital that may mold the growth and development of individuals living within them (Bourdieu 1977). Neighborhoods may also provide differential role models that expose their inhabitants to various peer

influences that affect the development of delinquency (Bowles 1977). Norms in some neighborhoods may allow more tolerance of criminals and criminal activity than norms in other neighborhoods. Some neighborhoods provide opportunities for individuals to achieve pro-social goals through community programs, organizations and resources, and guardianship, whereas other neighborhoods do not provide these opportunities. Neighborhoods may differ further in the availability of targets of crime and the collective efficacy of neighbors to act as guardians (Sampson, Raudenbush, and Earls 1997). These differences may encourage local differences in the amount and type of criminal activity. For example, police presence or the lack thereof, the presence of subway stations and other forms of mass transportation, and the lack of garages that mandates vehicles being parked on the streets all create opportunities for crime to exist. We suspect that local geographic variation may give rise to a multi-dimensional conception of crime that has not found support in studies confined to the individual level of analysis.

Based on these basic assumptions, we model crime indicators in a way that anticipates both individual differences and contextual variation. We leave it for later research to estimate the causes of this variation. At present, our definition will shape the way we conceptualize the measurement of crime. Our discussion of contextual variation suggests that it would be unreasonable to conceptualize crime as a single dimension that measures every type of crime. Such a one-dimensional characterization of crime would assume that all crimes have a common set of causes. Likewise it would be unreasonable to conceptualize each specific criminal act as having unique causes. In this case, we would be failing to understand that crimes like robbery, theft, and assault may be related to each other and, studied together, give us more precise measurement and a better understanding of each of these crimes. A useful conceptualization of crime must also recognize the fact that crimes vary inherently in seriousness. This implies the need for a continuous metric or scale upon which we may rate crimes.

Past research has illustrated that even before we determine the metric of the scale we must first consider the type of data we are collecting, be it official crime records, victimization reports, or self-reported offending. Although the comparison of the different types of crime reports is beyond the scope of this paper, differences in sources of crime reports have varied uses and sources of error.

Early survey research illustrated that self-reported offending produced different results than did patterns based on official crime records (e.g., Short and Nye 1957; Gould 1969; Elliott and Ageton 1980). In contrast, other research has found these two sources of data tend to be comparable when an adequate measure was conceptualized and operationalized (Hindelang et al. 1979, 1981). Although this literature is far from conclusive, criminologists largely agree that varied sources of data enrich understanding of crime and its relationships to covariates. We focus here on self-reported crime with the understanding that, while self-reports alone do not paint the full picture of crime, they do allow us to analyze crime from the perspective of the individual. For one of the best discussions of the self-report method, including a review of research, see Huizinga and Elliott (1986). Farrington et al. (1996) provide one of the most recent and comprehensive analyses comparing official and self-report data. The methods we propose here can be applied to official as well as to self-reported crime data, and they can be adapted to study other forms of social behavior, attitudes, or mental health outcomes.

4. THE MODEL

We first consider how the Rasch model may be applied to the modeling of self-reported crime, clarifying key model assumptions. Next, we show that a Rasch model with random effects is a special case of a two-level hierarchical logistic regression model. Finally, we extend that model to the case of multivariate, multilevel data.

The simplest and most easily interpretable item response model is the Rasch model. Consider a series of questions about violent crime, each asking whether the respondent has committed a specific act ("yes" or "no") during the past year. According to the Rasch model, the log-odds of a "yes" response depends on the severity of the act and the propensity of the respondent to commit violent crime. Key assumptions are that item severity and person propensity are additive in their effects and that item responses are conditionally independent given severity and propensity. These assumptions imply that the item set measures a unidimensional trait—for example, "propensity to commit violent crime." If these assumptions hold, model estimates yield a readily interpretable ordering of items and

persons on an interval scale (Rasch 1980; Wright and Masters 1982). We shall illustrate how to test the required assumptions and suggest a way to proceed when item response behavior does not fit these assumptions.

When applied to binary items tapping acts of crime, the Rasch model locates item severities, ψ_m, and person propensities to offend, π_j, on a log-odds ("logit") scale. Let $Y_{mj} = 1$ if person j responds affirmatively to item m and $Y_{mj} = 0$ if the response is negative for items $m = 1, \ldots, M$ and persons $j = 1, \ldots, J$. Let $\mu_{mj} = Prob(Y_{mj} = 1 \mid \psi_m, \pi_j)$ denote the conditional probability that person j will affirmatively respond to item m, and let $\eta_{mj} = \log[\mu_{mj}/(1 - \mu_{mj})]$, the natural log-odds of affirmatively responding. Then, under the Rasch model,

$$\eta_{mj} = \pi_j - \psi_m. \tag{1}$$

In words, the log-odds of a "yes" response is the simple difference between person j's propensity to offend, π_j, and item m's severity, ψ_m. Two key assumptions are made:

(i) *Local independence*: Conditional on item severity and person propensity, item responses Y_{mj} are independent Bernoulli random variables and thus have conditional mean μ_{mj} and conditional variance $\mu_{mj}(1 - \mu_{mj})$.

(ii) *Additivity*: Item differences and person differences contribute additively to the log-odds of an affirmative response.

A key condition for local independence to hold is that the M items in a set tap a single underlying dimension of crime ("unidimensionality"). For example, suppose that, unbeknownst to the researcher, a set of items assessing violent crime actually tapped two dimensions, such as violence in service of robbery (armed robbery, purse-snatching) and interpersonal aggression (e.g., hitting a family member in anger, hitting a peer in anger). Local independence would thus fail because covariation would arise among the items of each subtype. Local independence would also fail if the ordering of items created an autocorrelation.

Assumption (ii), if valid, gives credence to the idea that less frequently occurring acts of a given type are more severe. If (ii) falls, a two-parameter model might be formulated:

$$\eta_{mj} = \lambda_m(\pi_j - \psi_m). \tag{2}$$

In (2), each item is characterized not only by a location parameter ψ_m but also by the discrimination parameter, λ_m. Item and person characteristics enter multiplicatively into the model, and the severity of the item depends on the propensity to offend of the person. This idea is depicted in Figure 1(a), which displays the item characteristic curves (ICC) of three items that follow a Rasch model (or one-parameter model) as contrasted to Figure 1(b), which displays three item characteristic curves under the two-parameter model. The ICC expresses the probability of positive endorsement—that is, $Pr(Y_{mj}=1)$, as a function of the underlying latent propensity, to offend, π_j, of person j. The location parameter, ψ_m, is the point on the horizontal scale for which the probability of an affirmative response is 50. The slope, λ_m, is the slope of the ICC at that same point.

In Figure 1(a), item difficulties are, respectively, $\psi_1=-2$, $\psi_2=-1$, and $\psi_3=0$. Note that the Rasch model is a special case of the two-parameter model with $\lambda_1=\lambda_2=\lambda_3=1$. Under the Rasch model, a person with propensity of 0 is quite likely to respond affirmatively to the most frequently endorsed item (that is, item 1) and somewhat unlikely to respond affirmatively to the least frequently endorsed item (item 3). Under the Rasch assumptions, the fact that only the most serious offenders are likely to respond affirmatively to the least frequently endorsed items leads to the interpretation of ψ_m as "item severity."

Under Figure 1(b), the Rasch assumptions fail. Here we have $\psi_1=-2$, $\psi_2=-1$, and $\psi_3=0$ as before, but now the discrimination parameters (or "slopes") are not all equal. Rather $\lambda_1=.3$, $\lambda_2=1$, $\lambda_3=1$. Though the item location parameters ψ_m are the same as in Figure 1(a), they cannot be interpreted unambiguously as item *severities*, because now the relative likelihood of endorsement of the item depends on the criminality of the respondent. Those with very high propensities to offend are more likely to endorse item 2 than item 1 while those with lower propensities are more likely to endorse item 1 than item 2.

The parallelism of the curves in Figure 1(a) reflects the additivity assumption, lending the notion of severity to item location parameters. In contrast, the crossing of the item characteristic curves in Figure 1(b) reflects the multiplicative relationship between items and persons and undermines the notion that the item location parameters reflect severity.

(a) $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\psi_1 = -2$, $\psi_2 = -1$, $\psi_3 = 0$.
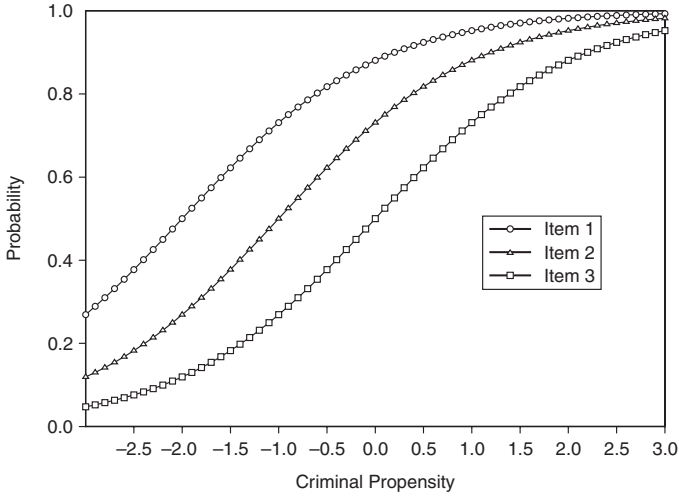
(B) $\lambda_2 = \lambda_3 = 1$, $\lambda_1 = 0.3$, $\psi_1 = -2$, $\psi_2 = -1$, $\psi_3 = 0$.
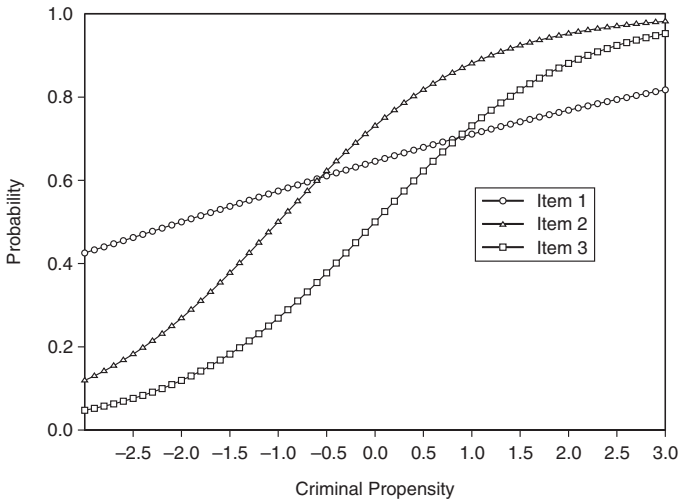
FIGURE 1. Probability of an affirmative response (vertical axis) as a function of propensity to offend when (a) discrimination parameters are equal, and (b) when they are not. The location parameter, ψ_m, is the point on the horizontal scale for which the probability of an affirmative response is .50. The slope, λ_m, is the slope of the curve at that same point.

We therefore test the additivity assumption and the associated clear interpretation of item location parameters by studying a two-parameter model, comparing its fit and its description of item behavior to that afforded by the Rasch model. We also check item fit by examining item-total correlations and standardized residuals. In the illustrative example, we find that items fitting the model poorly also produce departures from the Rasch assumptions. Discarding those few items creates a more coherent scale that also displays approximately Rasch behavior.

To test dimensionality, we extend the model to incorporate a multivariate structure (see Section 4.2). This leads to statistical tests of the dimensionality assumption.

4.1. *The Rasch Model as a Two-Level Logistic Regression Model*

Item response data can be viewed as having a two-level structure with items nested within persons. At the first level, we model the log-odds of an affirmative response, η_{ij}, as a linear function of item indicators. Let the index i denote an arbitrary item response and a_{mij} be an indicator variable taking on a value of 1 if the ith item response is to item m and zero otherwise. We then write

$$\eta_{ij} = \pi_j + \sum_{m=1}^{M-1} \alpha_{mj} \, a_{mij} \, . \tag{3}$$

Note there are $M - 1$ item indicators, and the item having no indicator is defined as the reference item. This rather general model allows the association between each a and η to vary across people. To fit the Rasch assumptions, we impose constraints on the level-2 model—that is, the model that describes variation across people:

$$\begin{aligned} \pi_j &= \gamma_0 + u_{0j} \\ \alpha_{mj} &= \alpha_m, m = 1, \dots, M - 1. \end{aligned} \tag{4}$$

Under (4) the associations between each a and η are invariant over respondents.

This standard two-level logistic regression model is equivalent to the Rasch model (Equation 1) with

$$\text{Person propensity} = \pi_j = \gamma + u_{0j}$$

$$\text{Item severity} = 0 \text{ for the reference item} \qquad (5)$$

$$\text{Item severity} = -\alpha_m \text{ for items } m = 1, \ldots, M - 1.$$

4.2. Extending the Rasch Model

We now extend the Rasch model to include randomly varying personal propensities, to include multiple dimensions of crime, and to incorporate a nested design, with persons nested within clusters (neighborhoods or schools, for example).

4.2.1. Random Effects

In the classical Rasch model, item severities and person propensities are fixed effects. A nice result is that, given complete data, item totals and person totals are sufficient statistics, increasing the ease of estimation (Rasch 1980). However, this approach does not handle missing data well, and respondents saying yes to all questions or no to all questions must be discarded. A random-effects model estimated by maximum likelihood incorporates data MAR and retains all respondents (Mislevy and Bock 1997). Retaining respondents with invariant response patterns is particularly important when covariates are included in the model to account for between-person variation in propensity to offend. The random effects specification does, however, require a parametric assumption about the distribution of the person propensities—i.e., that u_{0j} in equation (5) is distributed independently as $N(0, \tau)$.

4.2.2. Multivariate Structure

We further extend the model to include multiple crime types. In the illustrative example below, we shall consider a model with two crime types: violent crime and property crime. Under this model, each person will have two propensities: π_{VC} and π_{PROP} for violent crime and property crime, respectively. This allows study of the correlation structure and dimensionality of the crime types. Estimated correlations are adjusted for measurement error attributable to item inconsistency.

4.2.3. *Multilevel Structure*

Extending the model to include neighborhood effects (or school effects) allows contextual factors to contribute to individual propensities to offend, in line with our theoretical understanding (see Section 2). The correlation structure and dimensionality of crime types may differ at the person and neighborhood levels, and such differences can be studied using a multilevel approach. Finally, the multilevel approach naturally accommodates covariates measured on neighborhoods as well as persons, yielding standard errors that appropriately reflect the nested structure of the data and increasing the efficiency of estimation.

4.3. *The Full Model*

We add the index $p = 1, \ldots, P$ to allow for P types of crime and the index $k = 1, \ldots, K$ to allow nesting of persons within K clusters. Our level-1 model (equation 3) is thereby extended to

$$\eta_{ijk} = \sum_{p=1}^{P} D_{pijk} \left(\pi_{pjk} + \sum_{m=1}^{M_p - 1} \alpha_{pmjk} a_{pmijk} \right), \qquad (6)$$

where

$\eta_{ijk} = \log [\mu_{ijk}/(1 - \mu_{ijk})]$ is the log-odds that person j in neighborhood k will respond affirmatively to the ith item, where $\mu_{ijk} = Prob(Y_{ijk} = 1 | \pi_{pjk})$, with conditioning on all fixed effects and predictors also implicit;

D_{pijk} is an indicator taking on a value of 1 if the ith item is in the scale that measures crime type p, 0 otherwise;

π_{pjk} is the log-odds of an affirmative response by person j in neighborhood k to the reference item within crime type p;

$a_{pmijk} = 1$ if item i is the mth item within scale p, 0 otherwise; and

α_{pmjk} is the discrepancy between the log-odds of an affirmative response to the mth item in scale p for person jk and the reference item within that scale, holding constant π_{pjk}.

At the second level (between persons), we allow the person propensities to vary randomly within a neighborhood but require the item effects to be invariant across persons:

$$\pi_{pjk} = \beta_{pk} + u_{pjk}$$
$$\alpha_{pmjk} = \alpha_{pmk} \quad p = 1, \ldots, P; \; m = 1, \ldots M - 1, \tag{7}$$

where

β_{pk} is the mean "propensity," that is the mean of π_{pjk} within neighborhood k; and
u_{pjk} is a random person effect.

The random effects u_{pjk} for scales $p = 1, \ldots, P$ form a P-variate normal distribution with zero means, variances $\tau_{\pi pp}$ and covariances $\tau_{\pi pp'}$ between random effects in scales p and p'. These variances are assembled into the P by P matrix T_π.

At the third level (between neighborhoods), we allow the neighborhood mean propensities to vary randomly over neighborhoods but fix the average item effects as

$$\beta_{pk} = \gamma_{p0} + \upsilon_{pk}$$
$$\alpha_{pmk} = \alpha_{pm} \tag{8}$$

where

γ_{p0} is the population average log-odds of an affirmative response to the reference item in scale p; and
υ_{pk}, for scales $p = 1, \ldots, P$ form a P-variate normal distribution with zero means, variances $\tau_{\beta pp}$ and covariances $\tau_{\beta pp'}$ between random effects in scales p and p'. These covariance components are assembled into the P by P matrix τ_β.

The multivariate three-level model accords the following definitions:

Person propensity $= \pi_{pjk} = \gamma_{p0} + \upsilon_{pk} + u_{pjk}$
Neighborhood mean propensity $= \beta_{pk} = \gamma_{p0} + u_{pjk}$
Item severity $= 0$ for the reference item in scale p
Item severity $= -\alpha_{pm}$ for items $m = 1, \ldots M - 1$

4.3.1. Inclusion of Covariates

Person-level predictors of offending propensity may be included in the level-2 model (equation 7) and neighborhood-level predictors may be included in the level-3 model (equation 8). We illustrate this below.

5. ILLUSTRATIVE EXAMPLE

5.1. *Sample and Data*

The sample design involved two stages. At the first stage, Chicago's 343 neighborhood clusters (NCs) were cross-classified by seven levels of ethnic mix and three levels of socioeconomic status.[1] Within the 21 strata so constructed, NCs were sampled with the aim of producing a nearly balanced design. The resulting sample is described in Table 1 with census tracts as units.[2] The number of tracts in each stratum is shown in parentheses. The table shows that the confounding of ethnic mix and neighborhood socioeconomic status (SES) precludes study of certain combinations: There are no predominantly white and poor

TABLE 1

Number of Census Tracts (N = 196) by Socioeconomic Status (SES) and Racial/Ethnic Composition in the PHDCN Design[a]

Racial/Ethnic Strata	SES							
	Low		Medium		High		Total	
75% Black	31	(9)	10	(4)	9	(4)	50	(17)
75% White	0	(0)	7	(4)	18	(8)	25	(12)
75% Latino	12	(4)	12	(4)	0	(0)	24	(8)
20% Latino and 20% White	11	(4)	14	(5)	10	(4)	35	(13)
20% Hispanic and 20% Black	7	(4)	7	(4)	0	(0)	14	(8)
20% Black and 20% White	3	(2)	4	(4)	10	(4)	17	(10)
NCs Not classified above	8	(4)	14	(4)	9	(4)	31	(12)
TOTAL	72	(27)	68	(29)	56	(24)	196	(80)

[a] SES was defined by a six-item scale that summed standardized neighborhood-level measures of median income, percent college educated, percent with household income over $50,000, percent families below the poverty line, percent on public assistance, and percent with household income less than $50,000 based on the 1990 decennial census. In forming the scale, the last three items were reverse coded.

Note: The 80 sampled NCs are shown in parentheses.

[1] See Sampson, Raudenbush, and Earls (1997) for a detailed description of the construction of the 343 NCs.
[2] Our analysis uses the census tract (N = 196) rather than the NC (N = 80) as the analytic unit to increase statistical power at the between-neighborhood level.

tracts, nor are there any predominantly Hispanic and high-SES tracts. Nevertheless, there is substantial variation with ethnic mix by SES. (Note the presence of "Low," "Medium," and "High" SES tracts that are predominantly Black and many ethnically heterogeneous tracts that vary in SES.) Table 1 confirms the racial and ethnic segregation in Chicago while rejecting the common stereotype that minority neighborhoods in the United States are homogeneous.

At the second stage, dwelling units were enumerated ("listed") within each NC. In most instances, all dwelling units were listed, though for particularly large NCs, census blocks were selected for listing with probability proportional to size. Within listed blocks, dwelling units were selected systematically from a random start. Within selected dwelling units, all households were enumerated. Age-eligible participants were selected with certainty. To be age-eligible, a household member must have had an age within six months of one of seven ages: 0, 3, 6, 9, 12, 15, and 18 years of age. The analysis reported here used cohorts 9, 12, 15, and 18. Each child was administered a Self-Report of Offending questionnaire to determine participation in certain delinquent and criminal acts. Questions were of the form: Have you ever hit someone you lived with? (Yes = 1 and No = 0), followed by questions on last year prevalence and incidence. In this paper we focus on whether or not the respondent reported being involved in each item during the past year.

Home-based interviews with parents and children in each cohort were conducted over 30 months from 1994 to 1997. Though each family is being followed over a period of five to seven years for repeated assessments, we continue our attention in this paper to wave-1 data. Sample members were approximately half female; 45 per cent were of Hispanic origin while 36 per cent were Black and 15 per cent White (Table 2). Frequencies for other ethnic groups were small. To make use of all of the data, our analyses classified participants as Hispanic, Black, or other, with the understanding that "others" are overwhelmingly White. At the neighborhood level, we are interested in concentrated disadvantage, which is a weighted factor score constructed from the data of the 1990 decennial census of the population to reflect differences in poverty, race and ethnicity, the labor market, age composition, and family structure. Neighborhood concentrated disadvantage was created from five variables: (1) percentage of population below the poverty line, (2) percentage of population

TABLE 2
Descriptive Statistics

Person-level Data (N = 2977)	
Age	m = 13.28, sd = 3.33
Female	m = .500
Hispanic	m = .452
Black	m = .362
White	m = .146
Asian	m = .013
Pacific Islander	m = .002
American Indian	m = .010
Other	m = .014
Neighborhood-level data (J = 196)	
Concentrated disadvantage[a]	m = 0.04, sd = 0.82

[a] Concentrated disadvantage is a weighted factor score constructed from the data of the 1990 decennial census of the population to reflect differences in poverty, the labor market, age composition, and family structure: (1) percentage of population below the poverty line, (2) percentage of population that is on some form of public assistance, (3) percentage of population that is unemployed, (4) percentage of population that is less than 18 years of age, and (5) percentage of population whose households are female headed. Sampson, Raudenbush, and Earls (1997) provides a detailed description of the construction of the scale.

that is on some form of public assistance, (3) percentage of population that is unemployed, (4) percentage of population that is less than 18 years of age, and (5) percentage of population whose households are female headed. Sampson, Raudenbush, and Earls (1997) provide a detailed description of the construction of the scale.

We begin tentatively with a two-dimensional notion of crime. The resulting scales, displayed in Table 3, include *violent crime* (nine items) and *property crime* (six items) as self-reported during personal interviews. The violent crime items indicate acts of physical aggression (hitting someone you lived with in the past year with the intent of hurting them, similarly hitting someone you did not live with, throwing objects at others, robbery, purse snatching, pick pocketing, setting fires, gang fighting, and carrying a hidden weapon). Property crimes include purposely damaging property, breaking into a building to steal, stealing from a store, stealing from a car, stealing from a household member, and knowingly buying or selling stolen goods.

TABLE 3
Item Responses for Crime

Variable	Category	Frequency
Violent Crime		
Hit someone with whom you lived in the	0 = no	2469
past year with the idea of hurting them	1 = yes	353
Hit someone with whom you did not live in the	0 = no	2129
past year with the idea of hurting them	1 = yes	685
Thrown objects such as bottles or rocks at	0 = no	2495
people in the past year	1 = yes	333
Ever carried a hidden weapon in the last year	0 = no	2565
	1 = yes	258
Ever maliciously set a fire in the last year	0 = no	2809
	1 = yes	28
Ever snatched a purse/picked a pocket in the	0 = no	2821
last year	1 = yes	18
Ever attacked with a weapon in the last year	0 = no	2739
	1 = yes	95
Ever used a weapon to rob someone in the	0 = no	2822
last year	1 = yes	10
Ever been in a gang fight in the last year	0 = no	2667
	1 = yes	167
Property Crime		
Ever purposely damaged property in the	0 = no	2559
last year	1 = yes	269
Ever broke into a building to steal in the	0 = no	2819
last year	1 = yes	13
Ever stolen from a store in the last year	0 = no	2513
	1 = yes	317
Ever stolen from a household member in	0 = no	2440
the last year	1 = yes	381
Ever stolen from a car in the last year	0 = no	2776
	1 = yes	68
Ever knowingly bought/sold stolen goods in	0 = no	2662
the last year	1 = yes	164

5.2. Checking Model Assumptions: The Additivity Assumption

We seek a model with "Rasch properties": with item location parameters interpreted as "item severities" and person propensities lying on the same scale. Under the Rasch assumptions, person propensity and item severity combine additively to determine the log-odds of

item endorsement. We can check these assumptions by comparing results based on one-parameter and two-parameter models. Therefore, before considering our multivariate multilevel model, we estimated a Rasch model (equation 1) using a computer program named BILOG (Mislevy and Bock 1997) and compared the results with those based on a two-parameter model (equation 2), again using BILOG. The results appear in Table 4.

5.2.1. *Violence Scale*

Note that all "slopes" are constant in the one-parameter ("Rasch") results, but they are allowed to vary in the case of the two-parameter model.[3] Looking at the two-parameter results, we see that one item, "hitting someone you live with," has a particularly flat slope estimate (1.234) as compared to the average slope of 2.046. That same item has an unusually high standardized residual (5.075) under the one-parameter model as well as exhibiting the lowest biserial correlation of any item (.506). Using a Bayesian information criterion (BIC) to compare models, the two-parameter model fits the data very slightly better than does the one-parameter model.

We display parameter estimates in Figures 2(a) and 2(b). Note that all the ICCs are forced to be proportional in the case of the one-parameter model (Figure 2a). For the two-parameter model (Figure 2a), the ICCs tend to be nearly proportional except for the case of item 20 (hitting someone you live with). Clearly, this item, with a slope of 1.234, is less discriminating than are the other items.

It may be that violent acts committed at home constitute a different dimension than those committed outside the home. For whatever reason, item 20 behaves in a way that appears different from the other items. We therefore excluded item 20 and re-ran the two analyses (Table 5). With item 20 removed, the one-parameter

---

[3]BILOG sets the slope in the one-parameter model to a constant, not necessarily 1.0, and constrains the person propensities to have a mean of 0 and variance 1.0. An alternative and statistically equivalent parameterization would constrain the slope to 1.0 and allow the propensities to have a constant variance other than 1.0. See Mislevy and Bock (1997) and http://www.ssicentral.com/irt/bilog.htm for a copy of the BILOG program and user's manual. To obtain the BILOG code used in our analysis, contact the second author of this chapter at cjque@umich.edu.

TABLE 4
Results of the Rasch Model for Violence Scale

| Item	Item Biserial Correlation	Violence All Items					
		1 Parameter Model			2 Parameter Model		
		Threshold (s.e.)	Slope (s.e.)	Standardized Residual	Threshold (s.e.)	Slope (s.e.)	Standardized Residual
20 (hit someone you live with)	0.506	3.377 (0.122)	2.046 (0.072)	5.075	1.993 (0.121)	1.234 (0.103)	2.154
21 (hit someone you don't live with)	0.656	0.940 (0.038)	2.046 (0.072)	3.337	0.887 (0.039)	2.417 (0.194)	0.922
24 (throw objects at someone)	0.719	1.565 (0.047)	2.046 (0.072)	0.742	1.621 (0.071)	1.895 (0.146)	1.102
3 (carried hidden weapon)	0.804	1.759 (0.051)	2.046 (0.072)	1.822	1.675 (0.069)	2.345 (0.201)	1.310
6 (set fire)	0.824	3.172 (0.111)	2.046 (0.072)	2.053	3.285 (0.256)	1.924 (0.259)	1.945
11 (snatched purse)	0.694	3.377 (0.122)	2.046 (0.072)	1.798	3.816 (0.483)	1.651 (0.330)	1.439
22 (attacked with weapon)	0.961	2.436 (0.075)	2.046 (0.072)	3.337	2.204 (0.101)	2.693 (0.306)	0.771
23 (used force to rob)	0.870	3.747 (0.174)	2.046 (0.072)	1.417	3.739 (0.434)	2.072 (0.467)	1.649
25 (gang fight)	0.850	2.082 (0.060)	2.046 (0.072)	2.045	1.964 (0.085)	2.399 (0.226)	1.121
BIC (max BIC [in bold] is best)		−5426.39			**−5426.37**		

(a)

(b)
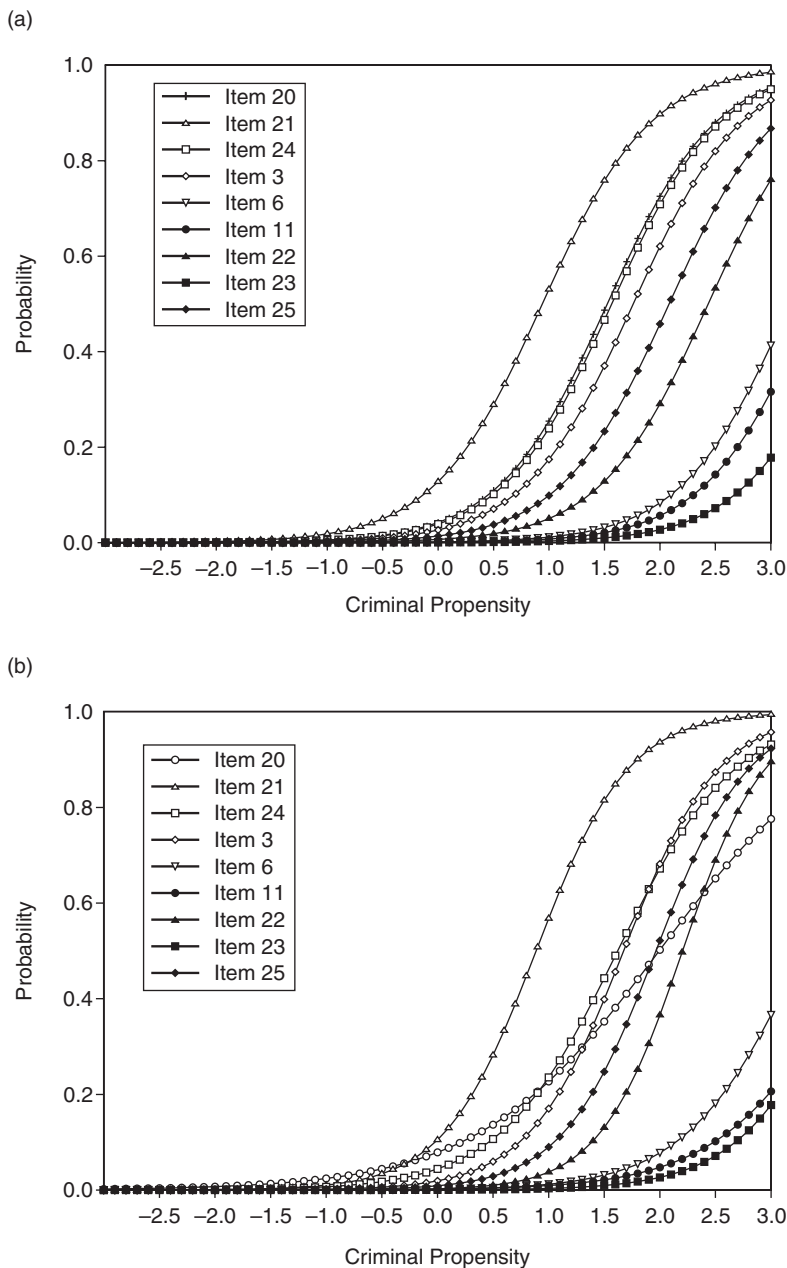
FIGURE 2. Graphical comparison of 1- and 2-parameter models for violent crime.

model now fits better than does the two-parameter model. The graphs of the ICCs look quite similar in both cases, as shown in Figures 3(a) and 3(b). We are inclined to remove item 20 and tentatively plan to consider violence at home to be different from the other aggression items. Using the revised scale, "Rasch" assumptions are reasonable. Item location parameters can reasonably be interpreted as item severities, and items and persons arguably are calibrated on a common scale.

5.2.2. *Property Scale*

We repeated the same procedures for the property crime. Remarkably, the item with the flattest slope, largest standardized residual (under Rasch), and by far the smallest biserial correlation was the single item in the scale that related crime within the household. Specifically, item 9 ("ever stolen from a household member") appears to discriminate less well than do the other items. Once again, the two-parameter model fit better with this item included, but when this "home" item was deleted, the one-parameter model fit better. Once again, under the two-parameter model the ICCs were nearly proportional once this item was removed. With this single item deleted, Rasch assumptions are again sensible.

　　　In sum, looking across the two scales, there is some reason to suspect that items relating to crimes committed within the household of the participants behave differently than do the other items. Specifically, at least for this sample and this wording of items, "domestic" crimes are less discriminating than are other crimes. Given this different behavior, and in light of the potentially meaningful differences between domestic acts and other acts of crime, we removed these items from our scale. A reasonable next step might be to construct and evaluate a "domestic crime" scale, but we do not pursue this in the current paper.

5.3. *Checking Model Assumptions: The Local Independence Assumption*

The Rasch model assumes that, given person propensity and item severity, item responses are independently sampled from a Bernoulli distribution. One way to test for violations of this assumption is to estimate a model with extra-binomial dispersion. If the within-participant variance is more or less than expected under an assumption

TABLE 5
One- and Two-Parameter Models for Violence One Item Removed

|		Violence (W/O Item 20)					
		1 Parameter			2 Parameter		
Item	Item Biserial Correlation	Threshold (s.e.)	Slope (s.e.)	Standardized Posterior Residual	Threshold (s.e.)	Slope (s.e.)	Standardized Posterior Residual
22 (attacked with weapon)	0.984	2.336 (0.074)	2.272 (0.086)	2.154	2.208 (0.099)	2.699 (0.304)	0.597
21 (hit someone you don't live with)	0.607	0.902 (0.038)	2.272 (0.086)	0.038	0.914 (0.042)	2.197 (0.177)	0.782
24 (throw objects at someone)	0.687	1.503 (0.048)	2.272 (0.086)	2.224	1.655 (0.076)	1.810 (0.143)	1.283
3 (carried hidden weapon)	0.844	1.689 (0.052)	2.272 (0.086)	1.488	1.639 (0.065)	2.501 (0.226)	0.949
6 (set fire)	0.840	3.030 (0.105)	2.272 (0.086)	2.223	3.300 (0.254)	1.912 (0.254)	1.887
11 (snatched purse)	0.684	3.220 (0.113)	2.272 (0.086)	2.416	3.877 (0.496)	1.611 (0.320)	1.424
23 (used force to rob)	0.912	3.562 (0.161)	2.272 (0.086)	1.392	3.707 (0.431)	2.113 (0.494)	1.821
25 (gang fight)	0.899	1.999 (0.061)	2.272 (0.086)	1.643	1.937 (0.081)	2.509 (0.242)	1.006
BIC (max BIC [in bold] is best)		**−4405.22**			−4428.16		

(a)

(b)
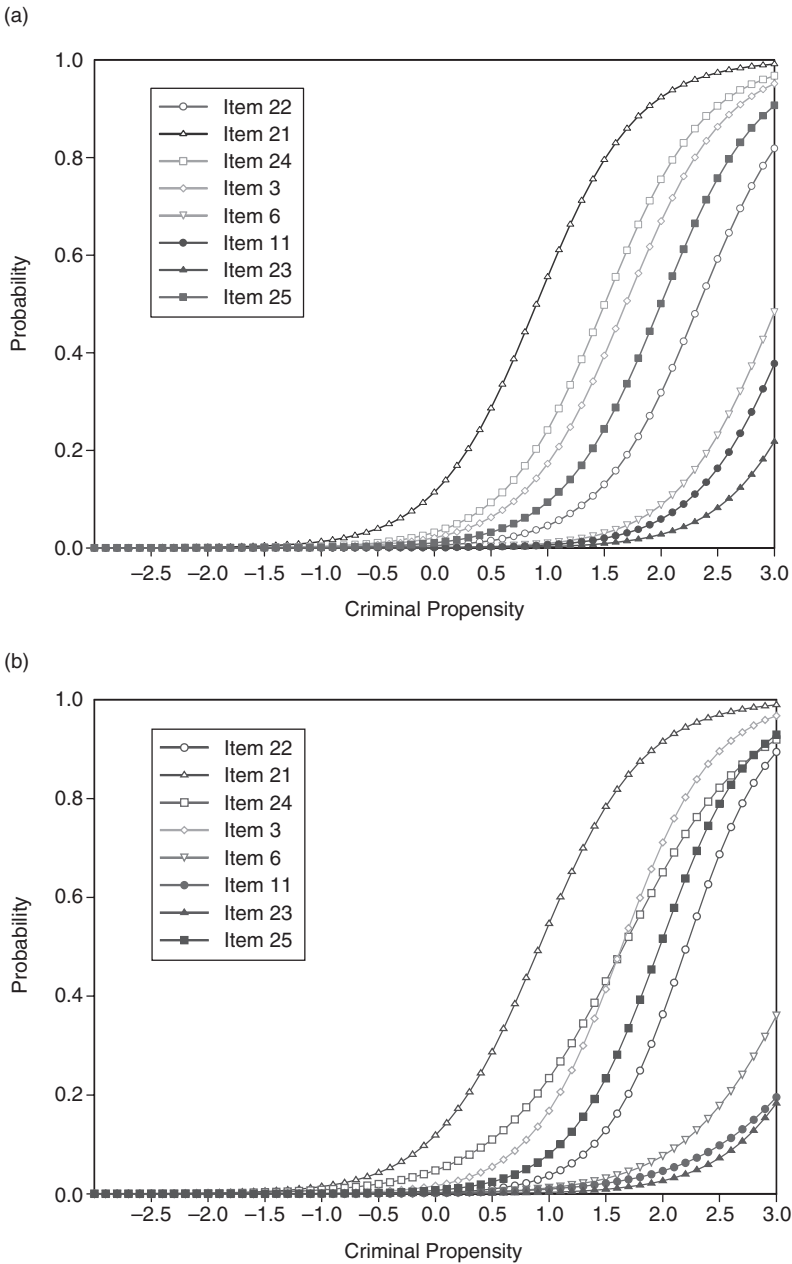
FIGURE 3. Graphical comparison of 1- and 2-parameter models for violent crime without item 20.

of independent Bernoulli trials, we have evidence against the local independence assumption. This can be accomplished using penalized quasi-likelihood (see Raudenbush et al. 2000).[4] We found that the conditional variance of the item responses was significantly lower than expected under a Bernoulli model. This result suggests that the local independence assumption may not hold. We therefore modified the Rasch model to allow for underdispersion in all subsequent analyses.

5.4. *Model*

We now apply the three-level model (equations 6, 7, and 8) to the case in which $P = 2$ types of crime: violent crime and property crime. The level-1 model views the log-odds of endorsement on item i as depending on which type of crime is of interest and which item was endorsed. Let $D_{VCijk} = 1$ if item i is an indicator of violent crime and 0 otherwise. Let $D_{PROPijk} = 1$ if item i is an indicator of property crime and 0 otherwise. We now have

$$
\eta_{ijk} = D_{VCijk} \left(\pi_{VCjk} + \sum_{m=1}^{7} \alpha_{VCmjk} a_{VCmijk} \right)
$$
$$
+ D_{PROPijk} \left(\pi_{PROPjk} + \sum_{m=1}^{4} \alpha_{PROPmjk} a_{PROPmijk} \right),
\tag{9}
$$

where

a_{VCmijk} and $a_{PROPmijk}$ are indicator variables representing the items in violent crime and property crime scales respectively;
π_{VCjk} and π_{PROPjk} are the adjusted log-odds of the endorsement of the crime on an "average item." They are the latent "traits" or "true scores" of a particular respondent on the corresponding scale;

---

[4]We estimated the three-level Rasch model using a very accurate approximation to maximum likelihood (ML) and also using penalized quasi-likelihood (PQL) with and without extra-binomial dispersion. Item severities were nearly perfectly correlated across the two analyses. The PQL results revealed evidence of substantial under-dispersion. PQL with under-dispersion produced slightly larger between-person variances than did the ML approach. Under this model the level-1 variance is $Var(Y_{ijk}|\eta_{ijk}) = \sigma^2 \mu_{ijk}(1 - \mu_{ijk})$, $\sigma^2 > 0$. Under the Bernoulli model, $\sigma^2 = 1$.

α_{VCmkj} and $\alpha_{PROPmjk}$ represent the "difficulty" or "level of severity" of crime item m. Note now that seven item indicators represent the eight violent crimes and four item indicators represent the five property items. For the sake of parsimony, these item "difficulties" will be again fixed across the children and across the tracts; that is $\alpha_{VCmjk} = \alpha_{VCm}$ and $\alpha_{PROPmjk} = \alpha_{PROPm}$.

Level-2 Model. The level-2 model accounts for variation between children within tracts

$$\pi_{VCjk} = \beta_{VCk} + u_{VCjk},$$

$$\pi_{PROPjk} = \beta_{PROPk} + u_{PROPjk} \tag{10}$$

on the latent measures of the problem behaviors, where β_{VCk} and β_{PROPk} are the pooled neighborhood averages for tract k on the two underlying latent traits respectively. The random effects u_{VCjk} and u_{PROPjk} are assumed bivariate normally distributed with zero means and person-level variance covariance matrix

$$\begin{bmatrix} \tau_{\pi VC} & \tau_{\pi VCPROP} \\ \tau_{\pi VCPROP} & \tau_{\pi PROP} \end{bmatrix}. \tag{11}$$

Level-3 Model. The level-3 model accounts for variation between tracts on the underlying latent measures of the crimes:

$$\beta_{VCk} = \gamma_{VC} + \upsilon_{VCk},$$

$$\beta_{PROPk} = \gamma_{PROP} + \upsilon_{PROPk}, \tag{12}$$

where γ_{VC} and γ_{PROP} are the grand mean levels of the latent traits of each type of crime respectively. The random effects ν_{VCK} and υ_{PROPk} are assumed bivariate normally distributed with zero means and a tract-level variance covariance matrix

$$\begin{bmatrix} \tau_{\beta VC} & \tau_{\beta VCPROP} \\ \tau_{\beta VCPROP} & \tau_{\beta PROP} \end{bmatrix}. \tag{13}$$

Combined Model. All of the above models can be combined through a series of substitutions of terms and displayed as follows in equation (14):

$$\eta_{ijk} = D_{VCijk}\left(\gamma_{VC} + \sum_{m=1}^{7}\alpha_{VCm}a_{VCmijk} + u_{VCjk} + \upsilon_{VCk}\right)$$
$$+ D_{PROPijk}\left(\gamma_{PROP} + \sum_{m=1}^{4}\alpha_{PROPm}a_{PROPmijk} + u_{PROPjk} + \upsilon_{PROPk}\right).$$

$$(14)$$

The above equation says that the log-odds of endorsement of an item for a particular scale depends on item severities and the unique effects associated with the individual child and tract.

5.5. Results

5.5.1. Item Severities

Table 6 provides the severities of the items within each scale. For the violence scale, as expected, we see that armed robbery, purse snatching, maliciously setting fire, and attacking someone with a weapon are among the rarest and therefore, under Rasch assumptions, most severe crimes. In contrast, hitting someone you don't live with, throwing an object at someone, and carrying a hidden weapon are less severe. In the property scale, breaking into a building to steal and stealing from a car are the most severe while stealing from a store and damaging property are less severe under our model assumptions.

5.5.2. Covariance Component Estimates

We see from Table 7 that the estimated correlation of the two latent scales within neighborhoods is high, $r = .65$, but not so high as to lead us to conclude that the two scales are indistinguishable. Looking now between neighborhoods, we see quite substantial variation in violent crime between neighborhoods. Indeed, about $\hat{\tau}_{\beta VC}/(\hat{\tau}_{\beta VC} + \hat{\tau}_{\pi VC}) = 0.46/(0.46 + 5.74) = 0.07$ or 7 per cent of the reliable variation in violent crime lies between neighborhoods. In contrast, we see no evidence of variation between neighborhoods in property crime with $\tau_{\beta PROP} = 0$. The fact that violent crime appears to vary significantly between neighborhoods while property crime does not provides

TABLE 6
Model Fitting Results: Item Severity

Item	Coefficient Estimate	Standard Error
Violent Crime		
Intercept,[a] γ_{VC}	−1.85	0.08
Ever used a weapon to rob someone in the last year, α_{VC1}	−6.59	0.20
Ever snatched a purse/picked a pocket in the last year, α_{VC2}	−5.91	0.15
Ever maliciously set a fire in the last year, α_{VC3}	−5.37	0.13
Ever attacked with a weapon in the last year, α_{VC4}	−3.72	0.08
Ever been in a gang fight in the last year, α_{VC5}	−2.80	0.07
Ever carried a hidden weapon in the last year, α_{VC6}	−2.10	0.06
Ever thrown objects at someone in the last year, α_{VC7}	−1.63	0.06
Property Crime		
Intercept,[b] γ_{PROP}	−3.61	0.07
Ever broke into a building to steal in the last year, α_{PROP1}	−4.31	0.19
Ever stolen from a car in the last year, α_{PROP2}	−2.14	0.09
Ever knowingly bought/sold stolen goods in the last year, α_{PROP3}	−0.80	0.07
Ever stolen from a store in the last year, α_{PROP4}	0.31	0.06

[a] The reference item is "ever hit someone you don't live with this year."
[b] The reference item is "ever damaged property in the last year."

additional evidence that the two types of crime are not collapsible into a single dimension.

5.6. *Incorporating Covariates*

We now expand the model to incorporate covariates describing differences between persons (age, gender, ethnicity, socioeconomic status)

TABLE 7
Covariance Component Estimate

Parameter Estimates[a]

$$\hat{\mathbf{T}}_\pi = \begin{bmatrix} \hat{\tau}_{\pi VC} & \hat{\tau}_{\pi VCPROP} \\ \hat{\tau}_{\pi PROPVC} & \hat{\tau}_{\pi PROP} \end{bmatrix} = \begin{bmatrix} 5.74 & 3.69 \\ 3.69 & 5.67 \end{bmatrix}$$

$$\hat{\tau}_\beta = \begin{bmatrix} \hat{\tau}_{\beta VC} & \hat{\tau}_{\beta VCPROP} \\ \hat{\tau}_{\beta VCPROP} & \hat{\tau}_{PROP} \end{bmatrix} = \begin{bmatrix} 0.46 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\hat{\sigma}^2 = .30$$

Derived Statistics

Intra-tract correlations	$\hat{\rho}_{VC} = 0.46/(0.46 + 5.74) = 0.07$
	$\hat{\rho}_{PROP} = 0/(0 + 5.67) = 0.00$
Interscale correlation (person level)	$\hat{\rho}_{VC,PROP} = 3.69/(5.74 \bullet 5.67)^{1/2} = .65$
Reliabilities:[b,c] Violence	
Between person	.65
Between tract	.42
Reliability: Property	
Between person	.50
Between tract	.00

---

[a] We can conceive the level-1 model as $Y_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$, where ε_{ijk} is a Bernoulli random error with mean 0 and variance $\mu_{ijk}(1 - \mu_{ijk})$. We can alternatively conceive of this variance as $\sigma^2 \mu_{ijk}(1 \cdot \mu_{ijk})$. If $\sigma^2 = 1$, the independent Bernoulli model holds. If $\sigma^2 \neq 1$, σ^2 represents extra-Bernoulli dispersion (Raudenbush et al. 2000, chapt. 6).

[b] Person-level reliability is the ratio of the variance of the latent "true" person means to the variance of the estimates, conditioning on neighborhood membership. Thus for scale p we have

$$rel(\hat{\pi}_{pjk}) = \frac{\text{var}(\pi_{pjk}|\beta_{pk})}{\text{var}(\hat{\pi}_{pjk}|\beta_{pk})} \approx \frac{\tau_{\pi pp}}{\hat{\tau}_{\pi pp} + \frac{\sigma^2}{\bar{w}_p n_p}},$$

where σ^2 is the extra-Bernoulli variance at level 1 and n_p is the number of items at level 1, and $\bar{w}_p = \bar{Y}_p(1 - \bar{Y}_p)$ with $\bar{Y}_p$ = proportion of "yes" responses in scale p.

[c] Tract-level reliability is the ratio of the variance of the latent "true" tract means to variance of the estimates:

$$rel(\hat{\beta}_{pk}) = \frac{\text{var}(\beta_{pk})}{\text{var}(\hat{\beta}_{pk})} \approx \frac{\tau_{\beta pp}}{\tau_{\beta pp} + \frac{\tau_{\pi pp}}{J_k} + \frac{\sigma^2}{\bar{w}_p n_p J_k}}$$

where J_k is the number of persons in neighborhood k. The reliability reported in the table is the average of the neighborhood-specific reliabilities.

and differences between neighborhoods (concentrated disadvantage). An advantage of the multivariate model is that it enables us to test whether the associations between covariates and outcomes differ across crime types.

More specifically, we expand the level-2 model (equation 10) to include the person-level covariates

$$\pi_{VCjk} = \beta_{0VCk} + \sum_{s=1}^{5} \beta_{sVCk} X_{sk} + u_{VCjk}$$

$$\pi_{PROPjk} = \beta_{0PROPk} + \sum_{s=1}^{5} \beta_{sPROPk} X_{sk} + u_{PROPjk}$$

(15)

where $X_1, \ldots, X_5$ are indicators for African-American and White (or other) ethnicity, female gender, and age-16, and (age-16)^2, respectively. Here Hispanics are the reference group. By modeling the log-odds of each crime as a quadratic function of age, we find that the predicted values on a probability scale produce a nice fit to the age-crime curve (Figure 4). We expand the level-3 model (equation 12) to include neighborhood concentrated disadvantage as a predictor:

$$\beta_{0VCk} = \gamma_{00VC} + \gamma_{01VC} W_k + \upsilon_{VCk}$$
$$\beta_{0PROPk} = \gamma_{00PROP} + \gamma_{01PROP} W_k + \upsilon_{PROPk},$$

(16)

where W_k = the degree of neighborhood concentrated disadvantage in neighborhood k. For simplicity, we fix other level-2 coefficients

$$\beta_{sVCK} = \gamma_{s0VC}$$
$$\beta_{sPROPk} = \gamma_{s0PROP}.$$

(17)

The results are displayed in Table 8 with fitted age-crime curves for males in Figures 4 and 5. We see similarly large gender differences on both violent crimes (Figure 4) and property crimes (Figure 5). We also see similarly shaped age-crime curves. However, ethnicity is associated quite differently from the two outcomes. African-Americans display significantly higher levels of violent crime than do Hispanics, but no such ethnicity effect is apparent with respect to property crime. There is a substantial and significant positive association between neighborhood concentrated disadvantage and violent crime, with no hint of such an effect on property crime.
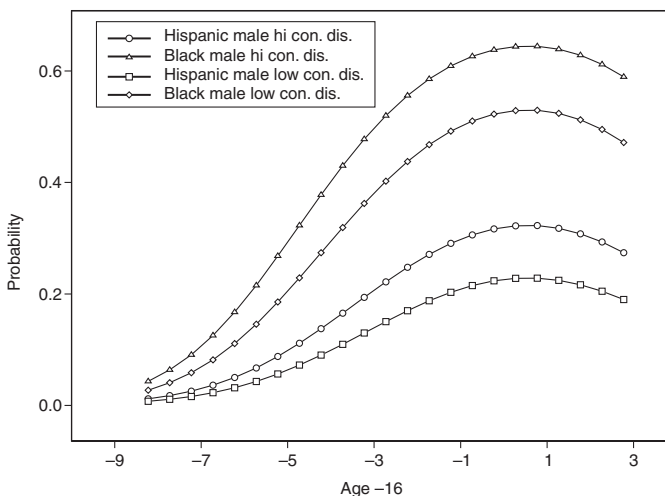
FIGURE 4. Probability of committing violent crime as a function of age, gender, and ethnicity.

Note: The Probability for the reference item "ever hit someone you don't live with" is graphed.

Exploiting the multivariate nature of the model, we can test whether the covariates have the same associations with violent crime as with property crime. We first tested the null hypothesis H_0: $\gamma_{01VC} = \gamma_{01PROP}$ and $\gamma_{s0VC} = \gamma_{s0PROP}$, $s = 1, \ldots, 6$. This yielded a chi-squared statistic of 98.14, d.f. $= 7$, $p = p < .001$, strongly suggesting rejection of the null hypotheses. Clearly the association with covariates differs for violent and property crime. We also tested two specific hypotheses. First, we tested whether the Black-Hispanic difference was the same for violent and property crime—that is, H_0: $\gamma_{20VC} = \gamma_{20PROP}$, yielding a chi-squared statistic of 55.01, d.f. $= 1$, $p < .001$. Clearly, ethnic gaps differ across the two scales. We also tested whether neighborhood concentrated disadvantage related differently to two scales—that is, H_0: $\gamma_{01VC} = \gamma_{01PROP}$, yielding a chi-squared statistic of 6.79, $p < .001$, again a significant result.[5]

---

[5]Note that power for tests of neighborhood effects is substantially less than for tests of person-based effects in these data.

TABLE 8
Association Between Covariates and Crime

(a) Fixed Effects

Predictor	Outcome: Violent Crime			Outcome: Property Crime		
	Coefficient	SE	t-ratio	Coefficient	SE	t-ratio
Intercept, γ_{00}	−0.992	0.123	−8.063	−2.318	0.131	−17.685
Concentrated disadvantage, γ_{01}	0.272	0.098	2.760	−0.025	0.106	−0.233
SES, γ_{10}	0.035	0.048	0.740	0.065	0.053	1.230
African-American, γ_{20}	1.334	0.144	9.320	0.108	0.154	0.070
White and other, γ_{30}	0.288	0.173	1.670	−0.106	0.185	−0.570
Female, γ_{40}	−1.004	0.112	−8.930	−0.940	0.125	−7.498
Age −16, γ_{50}	0.053	0.032	1.650	0.116	0.035	3.320
$(\text{Age} -16)^2$, γ_{60}	−0.048	0.006	−7.710	−0.038	0.007	−5.440

(b) Covariance Components (Controlling Covariates)

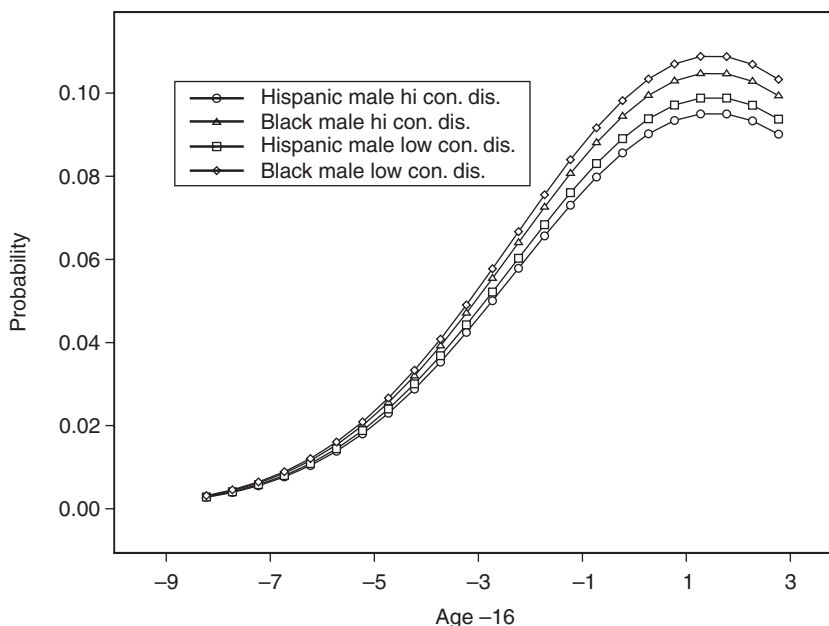|	Violent Crime	Property Crime
Intra-tract variance, $\tau_{\pi p}$	5.24	5.10
Inter-tract variance, $\tau_{\beta p}$	0.07	0

FIGURE 5. Probability of committing property crime as a function of age, gender, and ethnicity.
Note: The probability for the reference item "ever damaged property in the last year" is graphed.

6. DISCUSSION

Measures of criminal behavior yield a list of item responses, each indicating whether a participant has engaged in a specific criminal act. An essential question facing criminologists is how to combine the information from multiple item responses in a coherent way in order to study correlates of criminal behavior. Similar questions face researchers studying other forms of social behavior, attitudes, and mental health conditions.

A procedure for combining multiple responses reflects an implicit or explicit statistical model for the process that generates the responses. Models, of course, require assumptions. We generally prefer explicit models based on clearly stated assumptions that can be evaluated. Our aim in this inquiry has been to propose and illustrate such a model in the case of binary item responses for interviews

regarding self-reported crime, though the general approach is applicable to a broader array of outcomes.

A model, of course, is based on a theory, which also ought to be made explicit. We articulated a rudimentary theory that expects more or less stable differences between persons in propensity to offend but also allows for those propensities to vary over space and time. This leads to a statistical model that is multilevel, with variation in offending depending on neighborhood differences as well as individual differences. The approach readily extends to incorporate temporal change, as we shall discuss below.

The theory holds open the possibility that offenders may prefer to engage in one type of crime rather than another, perhaps as a result of varied opportunities for crime that arise in varied local settings. This aspect of the theory suggests that the model should be multivariate, allowing for the possibility of multiple, correlated dimensions or types of crime.

The conceptual frame also suggests that criminal acts vary in severity and that only those persons with the highest levels of criminal propensity at a given place and time will commit the most severe crimes of a given type. This notion led us to seek a calibration for the severity of criminal acts and the level of criminal propensity of persons during the year before the interview.

6.1. Checking Assumptions

When we combine these requirements, we arrive at a multilevel, multivariate Rasch model. Each aspect of the model is founded on assumptions that can and should be checked. Checking the assumptions not only provides a safeguard against unwarranted conclusions, it can also yield extremely helpful insights about the data and the theory. We therefore checked our Rasch assumptions as well as our multilevel and multivariate assumptions.

6.1.1. Checking the Rasch Assumptions

Under the Rasch model, the probability that a person will commit a given act depends additively on the criminal propensity of the person and the severity of the act. Persons with high propensities will commit a criminal act with higher probability than will persons with low propensity (holding constant the severity of the act). And more severe

acts of a given type will occur with lower probability than will less severe acts (holding constant the offender's propensity). These are strong assumptions that hold if and only if all items are equally discriminating. We therefore checked these assumptions by estimating a "two-parameter model"—that is, a model that allows the item discriminations to vary. The results were quite remarkable. Each of our two scales ("violent crime" and "property crime") contained one crime act that occurs domestically—inside the respondent's home. In both cases, the "domestic" crime item was less discriminating than were the other items. And in both cases, once the domestic item was removed, the simpler Rasch model fit the data better (according to a Bayesian information criterion) than did the more complex two-parameter model. These results suggested that domestic crimes may be of a different type than nondomestic types. The reasoning works as follows. Persons of comparatively low criminal propensity in the realm of nondomestic crimes appeared to engage in domestic crimes almost as frequently as did persons of high propensity in the non-domestic realm. This suggests that the "realms" are different or at least that the specific items labeled "domestic" are not tapping the same latent trait as are the other items. The results also suggest that, once these domestic crimes are removed from the two scales, the strong "Rasch" assumptions are reasonable for those scales, meaning that personal propensity and item severity can be calibrated on a common scale. Whether the "domestic" versus "nondomestic" distinction holds up over other items or across other samples is a topic for further research.

6.1.2. Checking the Multilevel Assumptions

We found statistically significant variation between neighborhoods in violent crime but not property crime. Moreover, after fitting our model that incorporated covariates at each level, we found a significant positive association between neighborhood concentrated disadvantage and violent crime but no association between such disadvantage and property crime. One might conclude that our multi-level model was essential for capturing contextual variation in violent crime but provided no advantage if one were interested only in property crime. Of course, one cannot assess the need for the multi-level aspect unless one has access to a multilevel modeling strategy.

6.1.3. *Checking the Multivariate Assumptions*

The multivariate nature of the model enabled us to assess dimensionality. We posed and tested a hypothesis of unidimensionality—the hypothesis that the items nominally tapping violent and property crime in reality reflect a single underlying criminal propensity. We reject that hypothesis for three reasons: (1) the correlation between the latent crime dimensions between persons within neighborhoods was large ($r = .65$) but not large enough to convince us that the two were indistinguishable; (2) as mentioned above, violent crime but not property crime displayed significant variation between neighborhoods; and (3) the two dimensions related differently to covariates. The multivariate nature of the model enabled us to test these differences. We found that the difference between Blacks and Hispanics was significantly larger for violent than for property crime and that the association between concentrated disadvantage and violent crime was significantly larger than the association between concentrated disadvantage and property crime.[6] Indeed, the point estimates of the ethnic gap and the concentrated disadvantage effect were large enough to be practically significant in the case of violent crime (Figure 4) and nearly null in the case of property crime (Figure 5). That the difference between coefficients defined on the two scales was highly significant statistically undermines a speculation that measurement unreliability might account for these differences.

6.1.4. *Checking the Multivariate Normality Assumption*

We assumed the random effects at the person level and neighborhood level to be multivariate normal in distribution. This assumption can be checked graphically (see Raudenbush and Bryk 2002, ch. 9). In general, estimates of fixed effects (that is, item difficulties and effects

---

[6]This result provides stronger evidence of bidimensionality than would be apparent from separate univariate analyses. For example, separate univariate analyses might show a significant effect of concentrated disadvantage for violent crime and no corresponding significant effect for property crime. But this pair of findings does not imply that the two effects are significantly different from each other. We also note that the common scaling of the outcomes in the logit metric facilitates the multivariate tests: The coefficients in the model for violent crime are on the same scale as the coefficients in the model for property crime, which provides a sound basis for interpreting their difference as evidence of bidimensionality.

of covariates) are robust to failure of this assumption as are point estimates of variance parameters.

6.2. *Interpreting Correlates of Crime*

How might we interpret regression coefficients when the outcome variable has been defined by our model? We offer two alternative interpretations.

First, we can view π_{pjk} as a latent propensity of person j in neighborhood k to engage in crime type p. This is a continuous variable scaled on a logit metric, so interpretations of regression coefficients follow conventional practice. In this view, the first level of our model is a *measurement model* describing how the observed binary outcome Y_{ijk} varies given an underlying probability jointly determined by the latent π_{pjk} and the item severities. The "top two-levels" of the model constitute a *structural model* describing how person and neighborhood characteristics combine to explain variation in the latent variables π_{pjk}.

Second, we might interpret regression coefficients by avoiding the latent variable terminology. Instead, we say that graphs like Figure 4 portray the association between covariates and the probability of committing a specific type of crime. In Figure 4, the y-axis is the probability of "hitting someone you do not live with," the reference item in the violent crime scale. This modeling of associations between covariates and item-specific probabilities, is, however, subject to strong constraints. In particular, we impose the constraint that the covariates have the same association with each odds of an affirmative response within scale p. However, the scale-specific intercepts do vary as a function of the item. This means that, had we changed the reference item to, say, "attacked someone with a weapon," Figure 4 would have exactly the same shape it currently has; only the metric of the y-axis would change. We might refer to this as a "proportional odds" assumption: the odds of committing each crime within a common scale p change in proportion as we change the value of a covariate, holding constant the remaining covariates.

The two ways of interpreting the regression coefficients are equivalent mathematically. However, the first emphasizes an under-lying continuity in criminal propensity while the second emphasizes

assumptions made about constraints imposed on the association between covariates and specific acts of crime. We tend to prefer the second for these data because many persons in our sample report committing no crimes. To assume that a continuous latent variable underlies their observable behavior makes sense in a theoretical sense, but the data provide no information to check such an assumption. In contrast, to emphasize the modeling of item-specific probabilities (or odds) while stipulating the proportional odds assumption makes sense to us. It emphasizes the importance of an assumption that can be checked.

6.3. *Checking the Proportional Odds Assumption*

In our level-2 and level-3 models—that is, the "person-level" and "neighborhood-level" models of equations (15) and (16)—we incorporated covariates as predictors of the intercept for each scale. However, we did not incorporate covariates as predictors of the item location parameters (the item "severities"). Nor did we allow these item effects to vary randomly over persons or neighborhoods. It was this decision that enforced a proportional odds structure on our results.

Had we allowed covariates to predict item location parameters we would implicitly have departed from the assumption of unidimensionality. For example, if we allow neighborhood disadvantage to relate differently to "hit someone you do not live with" and "attacked with a weapon," we would be saying that these items measure somewhat different things (because they relate differently to a covariate). Such a violation of unidimensionality is known in the educational testing literature as "differential item functioning" (Holland and Wainer 1993). It is often used to detect "item bias," for example to detect items that relate differently to gender or ethnicity, holding constant the "trait" of the person assessed. In our context, modeling item parameters as a function of covariates offers a way to check for differential item functioning, thus checking the proportional odds assumptions.

6.4. *Incorporating Time*

As mentioned, our conception views criminal propensity as varying not only among people but also over time and space. We have

incorporated spatial heterogeneity via a multilevel model with random effects of neighborhoods. This approach might be augmented by allowing spatial dependence in these random effects (c.f., Langford 1999). Incorporating time is straightforward. After obtaining repeated measures data, we define Y_{itjk} as a binary variable that takes on a value of 1 if person j in neighborhood k responds affirmatively to item i at time t. Then μ_{itjk} is the changing probability of committing a crime of the type indicated by item i, and η_{tijk} is the log-odds of an affirmative response. The level-1 model represents variation in item responses within an occasion for a given person, and defines the latent variable π_{tjk} as the criminal propensity of person j in neighborhood k at time t. The level-2 model is then a model for within-person change in this latent propensity, possibly as a polynomial function of age or time. This model defines person-specific "change parameters" (e.g., an intercept, a linear change rate, and a quadratic change rate) that vary over persons in the level-3 model and over neighborhoods in the level-4 model. In sum, the first level of the model is a measurement model and the "top" three levels of the model constitute a model for individual change among persons nested with social settings (c.f., Raudenbush and Bryk 2002, chap. 8). We shall illustrate application of this approach in future work as we follow the Chicago sample across subsequent waves of data collection.

6.5. Comparison to a More Conventional Approach

One might ask whether the extra effort needed to implement the approach we propose here yields much of a difference in bottom-line results. Perhaps a less-rigorous approach would produce similar results. To check out this possibility, we created a simple count of the number of violent crimes committed from our original list of nine (Table 3) and used this count as an outcome in a two-level linear model with the same covariates as specified in Table 8. The sign of every coefficient in the "simple" analysis was the same as in the "complex" analysis. However, in every case, the absolute value of the t-ratio was smaller in the simple analysis. For example, the t-ratio associated with neighborhood concentrated disadvantage was 2.76 in the complex analysis but only 2.02 in the simple analysis, so that the simple analysis suggested only a marginally significant association

with violent crime. Part of this discrepancy is explainable by the fact that 66 cases with item-missing data were dropped from the simple analysis. Such cases are easily incorporated into the more complex analysis and could also be incorporated into the simpler analysis using imputation of missing values. But resorting to imputation makes the simple analysis less simple and therefore less appealing. The discrepancy in t-ratio is not likely due entirely to this small reduction in sample size, however, and the assumptions underlying the t-ratio are more realistic in the complex analysis, using a logit link function, than in the simple analysis, which uses a linear model for a highly skewed outcome. The more realistic assumptions of the complex analysis lend greater credence to its inferential results. In addition, the fitted values based on the simple analysis were negative for 166 (6 per cent) of the cases, casting doubt on the metric of the estimated effect sizes. Negative predictions of the probability of crime are not possible using the more sensible logit link function.[7]

Finally, and perhaps most importantly, the simple approach omits a step in the analysis that we have come to view as key: the careful assessment of item functioning that can lead to the modification of scale construction. For example, in the present case, our item analysis led to the discovery of a clear ill-fitting pattern for in-home self-reported criminal behavior. The more complex analysis makes explicit assumptions about the measurement model, provides ways to check those assumptions, and thereby encourages precision in specification of the outcome variable and the interpretation of results that is absent when using more *ad hoc* procedures. Indeed, the incapacity of the simpler analysis to discern ill-fitting items may in part account for its noisier results. A careful simulation study would help clarify the conditions under which the two analyses will differ, but our sense is that the greater effort to implement the more complex analysis is well rewarded.

[7]The problem of negative fitted values based on the simple model could be addressed by using a logit or log link for the counted data. But such strategies rob the simple approach of some of its simplicity.

REFERENCES

Blumstein, Alfred, Jacqueline Cohen, Jeffrey Roth, and Christy Visher, eds. 1986. *Criminal Careers and "Career Criminals."* Washington, DC: National Academy Press.

Bourdieu, P. 1977. "Cultural Reproduction and Social Reproduction." Pp. 487–511 in *Power and Ideology in Education*, edited by J. Karabel and A. H. Halsey. New York: Oxford University Press.

Bowles, S. 1977. "Unequal Education and the Reproduction of the Social Division of Labor." Pp. 137–53 in *Power and Ideology in Education*, edited by J. Karabel and A. H. Halsey. New York: Oxford University Press.

Cohen, L. E., and M. Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44:588–608.

Cohen, L. E., and B. J. Vila. 1996. "Self-Control and Social Control: An Exposition of the Gottfredson-Hirschi/Sampson-Laub Debate." *Studies on Crime and Crime Prevention* 5:12550.

Elliott, D., and S. Ageton. 1980. "Reconciling Race and Class Differences in Self-Reported and Official Estimates of Delinquency." *American Sociological Review* 45:95–110.

Farrington, D., R. Loeber, M. Stouthamer-Loeber, W. Van Kammen, and L. Schmidt. 1996. "Self-Reported Delinquency and a Combined Delinquency Seriousness Scale Based on Boys, Mothers, and Teachers: Concurrent and Predictive Validity for African-Americans and Caucasians." *Criminology* 34:493–518.

Gottfredson, M. R., and T. Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Gould, L. 1969. "Who Defines Delinquency? A Comparison of Self-reported and Officially Reported Indices of Delinquency for Three Racial Groups." *Social Problems* 16:325–36.

Hindelang, M., T. Hirschi, and J. Weis. 1979. "Correlates of Delinquency: The Illusion of Discrepancy Between Self-Report and Official Measures." *American Sociological Review* 44:995–1014.

———. 1981. *Measuring Delinquency*. Beverly Hills, CA: Sage.

Holland, P., and H. Wainer, eds. 1993. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Horney, J., D. W. Osgood, and I. H. Marshall. 1995. "Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances." *American Sociological Review* 60:655–73.

Huizinga, D., and D. Elliott. 1986. "Reassessing the Reliability and Validity of Self-Report Measures." *Journal of Quantitative Criminology* 2:293–327.

Langford, I. H. 1999. "Multilevel Modeling of the Geographical Distributions of Diseases." *Applied Statistics* 48(2):253–268.

Little, R., and D. Rubin, 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

Mislevy, R. J., and R. D. Bock. 1997. *BILOG Version 3.11*. Chicago, IL: Scientific Software International.

Moffitt, T. 1993. "Adolescence-Limited and Life-Course Persistent Antisocial Behavior: A Developmental Taxonomy." *Psychological Review* 100:674–701.

Osgood, D. W., B. J. McMorris, and M. T. Potenza. 2002. "Analyzing Multiple-item Measures of Crime and Deviance 1: Item Response Theory Scaling." *Journal of Quantitative Criminology* 18(3):267–296.

Osgood, D. W., L. L. Finken and B. J. McMorris. 2002. "Analyzing Multi-Item Measures of Crime and Deviance II: Tobit Regression Analysis of Transformed Scores." *Journal of Quantitative Criminology* 18(4):319–347.

Piquero, A. R. 2000. "Frequency, Specialization, and Violence in Offending Careers." *Journal of Research in Crime and Delinquency* 37:392–418.

Rasch, G. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests.* Chicago: University of Chicago.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Newbury Park, CA: Sage.

Raudenbush, S., M. L. Yang, and M. Yosef. 2000. "Maximum Likelihood for Hierarchical Models via High-order, Multivariate LaPlace Approximation." *Journal of Computational and Graphical Statistics* 9(1):141–57.

Raudenbush, S. W., and R. Sampson. 1999. "Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." Pp. 1–41 in *Sociological Methodology 1999*, edited by Michael E. Sobel. Cambridge, MA: Blackwell Publishers.

Sampson, R. J., and J. H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

Sampson, R. J., S. Raudenbush, and F. Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918–24.

Short, J. F. Jr., and L. Nye. 1957. "Reported Behavior as a Criterion of Deviant Behavior." *Social Problems* 5:207–13.

Warr, M. 1993. "Age, Peers, and Delinquency." *Criminology* 31:17–40.

Wolfgang, M., R. Figlio, and T. Selfin. 1972. *Delinquency in a Birth Cohort.* Chicago, IL: University of Chicago Press.

Wright, B., and G. Masters. 1982. *Rating Scale Analysis: Rasch Measurement.* Chicago, IL: MESA Press.