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

Christopher Johnson and Stephen W. Raudenbush

University of Michigan

September 30, 2002

The research reported here was funded by grants from the John D. and Catherine T. MacArthur Foundation and the National Institute of Justice to the Project on Human Development in Chicago Neighborhoods (Felton Earls, Principal Investigator). Previous versions were presented at the Johns Hopkins University and the American Society of Criminology annual meeting.

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

Abstract

Repeated measures data generated from longitudinal designs are often used when studying correlates of individual change. Such studies pose several challenges: (1) The measurement scale must be invariant over time; (2) covariates of interest are often multilevel (e.g., measured at the person and neighborhood level); (3) some item-level missing data can be expected. To cope with these challenges, we propose a repeated measures, multilevel Rasch model with random effects. Under assumptions of conditional independence, additivity, and measurement invariance over time, the approach enables the investigator to calibrate the items and persons on an interval scale, incorporate covariates at each level, and accommodate data missing at random. Using data on 8 items tapping violent crime from 2,842 adolescents ages 9-18 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.

Longitudinal data enable us to repeatedly measure the status of individuals over a specified time period. We can then fit models to estimate the individual change taking place over that time. In fitting these models we must assume that measurement invariance is present across our time points; that is, our items tap the same underlying latent trait at each time of measurement. We can test this assumption by assesing item behavior over time. This suggests application of a longitudinal item response theory (IRT) model, the focus of this chapter.

A second rationale for a longitudinal IRT model arises when the number of items is small and /or the item responses are highly skewed. Available methods for modeling growth typically assume a continuous outcome variable. Often this assumption will fail. For example, a scale score consisting of crime items with a low probability of endorsement would not approximate continuity even if it included a very large number of items. This is also the case for many other types of data that arises widely in studies of behavior, beliefs, attitudes, exposure to risk, and symptoms of disease. A longitudinal IRT model does not require that the outcome in growth studies be continuous or normally distributed.

Raudenbush, Johnson, and Sampson (in press) showed how to embed an IRT model in a hierarchical structural model using cross sectional data. They also illustrated how to make this IRT model multivariate and multilevel. They found that this new methodology avoids problems such as negative predicted probabilities of the outcome variable and incorrect t-statistics while providing for a careful assement of item functioning. In this chapter, we intend to extend their methodology to incorporate repeated measures data.

Raudenbush et al.'s first extension was to make the IRT model multivariate. This allows for the simultaneous study of different types of crime, allowing assessments of the key model assumption of unidimensionality which requires that each IRT measure taps a single interval scale. The multivariate approach enabled them to study whether covariates relate differently to different types of offending net effects of measurement error.

The second extension of the IRT model was made to make the model multilevel. This extension reflected the fact that much of the data collected within social settings is in a natural nested structure. For example, in the study of crime we would like to take into account that the respondent is nested within a neigborhood, which is in turn nested within a city. Without taking this nested structure into account, the study of the variation and covariation of the propensity to offend might be biased due to the incorrect specification of the variance components. The multilevel approach is particularly important when explanatory variables are measured at higher levels, e.g., the neigborhood or city level.

The third extension of the IRT model had to do with accommodating the data missing at random (MAR). Data are missing at random when the probability of missing this is independent of the missing data, 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). To cope with this problem, the IRT model specifies person effect as random rather than fixed. The fourth extension of the IRT model was to add covariates at different levels of our model in order to predict criminal behavior. Having set up the model in this way, Raudenbush, et al., (2002) were able to determine whether covariates related differently to different scales of crime and were are also able to determine whether covariates related differently to different items.

Raudenbush et al., opted for a one-parameter IRT or Rasch model. The Rasch model, which specifies a location parameter for each item, is simpler than a two-parameter which specifies a location parameter and a discrimination parameter for each item. This brings to light one of the strong assumptions of the Rasch model: each item is equally discriminating. This assumtion implies that the relative severity of the items is identical for all persons. This is a very appealing feature, allowing us to view items as more or less severe, anchoring the scale in a conceptually meaningful way.

In this chapter, we extend the Rasch model in a different way. While allowing the model to be a multilevel random effects model, we allow the model to encompass repeated measures on a set of items. This paper is organized as follows. We begin with a brief review of the Rasch model, describing its application to self-reported criminal behavior. Second, we show how the Rasch model with random effects can be formulated as a special case of a two-level hierarchical logistic regression model. Third, we show how the two-level model formulation of the Rasch model readily extends to three levels incorporating multilevel data. Fourth, we extend the model to include repeated measures. Fifth, we illustrate an application of this model by analyzing data on 2977 adolescents nested within 196 and cents census tracks and interviewed at two-time points. We show how to specify mean propensity and growth in propensity to commit violent crime and how to test hypotheses concerning person level and neighborhood level predictors of latent status and change in criminal propensity. We close with suggestions for future research.

The Rasch Model Applied to Self-Reported Crime

As mentioned above the easiest IRT model to interpret is the Rasch model. Consider a cross-sectional survey asking a series of questions about violent crime. For each question the respondent indicates whether he or she 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, e.g., "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)

The additivity assumption of the Rasch model implies 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 formally, 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 $m = 1,...,M$ and persons $j = 1,\ldots,J$. Let $\mu_{mj} = Prob(Y_{mj} = 1|\psi_m,\pi_j)$ 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,

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

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. Key assumptions are:

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"). Suppose, for example that, unbeknownst to the researcher, a set of items assessing violent crime actually tapped two dimensions, e.g., 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). Then local independence would fail because covariation would arise among the items of each sub-type. Local independence would also fail if the ordering of items created an auto-correlation in the responses.

A Two-Parameter Model

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:

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

In (2), each item is characterized not only by a location parameter ψ_m but also by the "discrimination parameter," λ_m. Under (2) 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 la, which displays the item characteristic curves (ICC) of three items that follow a Rasch model (or "one-parameter model") as contrasted to Figure 1b, which displays three item characteristic curves under the two-parameter model. The item characteristic curve 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.

Insert Figures la, 1b About Here

In Figure 1a, item severities 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 lb, 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 1a, 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 1a reflects the additivity assumption. In contrast, the crossing of the item characteristic curves in Figure lb reflects the multiplicative relationship between items and persons and also undermines the notion that the item location parameters reflect severity.

Raudenbush et al. (in press) test the additivity assumption by studying a two-parameter model. They compared the fit of the two-parameter model to that of the Rasch model. They also checked item fit by examining item-total correlations and standardized residuals. Using the first wave of data we analyze in this chapter, they found that one item fit the scale poorly and also produced departures from the Rasch assumptions. Discarding that item created a more coherent scale that also displays approximately Rasch behavior.

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. Viewed this way, the Rasch model is a special case of a two-level logistic regression model. 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 *i*th item response comes from item *m* and zero otherwise. Then we write

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

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:

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

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

$$person\ propensity = \pi_j = \gamma_0 + u_{0j}$$

$$item\ severity = 0\ \ for\ the\ reference\ item$$

$$item\ severity = -\ \alpha_m\ for\ items\ m = 1,...,M-1.$$

The Rasch Model as a Multilevel Model

Our Item response data can also be viewed as having a three-level structure with items nested within persons, and persons nested within neighborhoods (or schools). At the first level, we model the log-odds of an affirmative response, η_{ijk}, as a linear function of item indicators. As before, let the index *i* denote an arbitrary item response and a_{mij} be an indicator variable taking on a value of 1 if the *i*th item response comes from item *m* and zero otherwise. Let *j* denote the person and *k* denote the neighborhood. Then we write

$$\eta_{ijk} = \pi_{jk} + \sum_{m=1}^{M-1} \alpha_{mjk}\, a_{mijk}. \qquad \mathbf{5}$$

The level-2 model describes variation across people within neighborhoods:

$$\begin{aligned} \pi_{jk} &= \beta_{0k} + u_{0jk} \\ \alpha_{mjk} &= \alpha_{mk},\ m = 1,...,M-1. \end{aligned} \qquad \mathbf{6}$$

where

π_{jk} is the log-odds of an affirmative response by person *j* in neighborhood *k;* and

u_{0jk} is a random person effect.

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

$$\begin{aligned} \beta_{0k} &= \gamma_0 + \upsilon_{0k} \\ \alpha_{mk} &= \alpha_m \end{aligned} \qquad \mathbf{7}$$

where

γ_0 is the population average log-odds of an affirmative response; and

υ_{0k} is a random person effect.

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

$$person\ propensity = \pi_{jk} = \gamma_0 + \upsilon_{0k} + u_{0jk}$$

$$item\ severity = 0\ \ for\ the\ reference\ item \qquad \mathbf{8}$$

$$item\ severity = -\ \alpha_m\ for\ items\ m = 1,...,M-1.$$

In addition this three-level model, which includes person and neighborhood effects (or school effects), allows contextual factors to contribute to individual propensities to offend. The correlation structure may differ at the person and neighborhood levels, and such differences can be studied using a multilevel approach. This 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.

The Rasch Model as a Repeated Measures Multilevel Model

We now extend the three-level model of the previous section to incorporate longitudinal data[1]. Here we add the index $t = 1,...,T$ to allow for repeated measures. Our model for growth is a polynomial of degree P:

$$\eta_{tijk} = \pi_{0jk} + \pi_{1jk} d_{tijk} + \pi_{2jk} d_{tijk}^2 + ... + \pi_{pjk} d_{tijk}^P + \sum_{m=1}^{M-1} \alpha_{mjk} a_{mtijk} \qquad \mathbf{9}$$

where

[1] Note that we could also extend our model to incorporate multiple types of crime as shown in Raudenbush, Johnson, and Samson (2002) but for simplicity we chose to examine a single type of crime.

$\eta_{tijk} = \log [\mu_{tijk} / (1 - \mu_{tijk})]$ is the log-odds of the conditional probability μ_{tijk} that person *j* in neighborhood *k* will respond affirmatively to the *i*th item at time *t*;

$\mu_{tijk} = Prob(Y_{tijk} = 1 | \pi_{jk})$, with conditioning on all fixed effects and predictors also implicit;

d_{tijk} is a measure of time or age;

π_{0jk} is a person-specific intercept, the log-odds of a "yes", at time $d_{tijk} = 0$, by person *j* in neighborhood *k* to the reference item;

π_{pjk} , p>0 is the p^{th} order polynomial effect (e.g. linear, quadratic, cubic, etc.) on the reference item;

$a_{mtijk} = 1$ if item *i* is the *m*th item at time *t*, 0 otherwise; and

α_{mjk} is the discrepancy between the log-odds of an affirmative response to the *m*th item for person *j* in neighborhood *k* at time *t* and the reference item, holding constant the changing propensity

$$\pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} d_{tijk}^{p} .$$

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:

$$\begin{aligned} \pi_{pjk} &= \beta_{pk} + u_{pjk} \\ \alpha_{mjk} &= \alpha_{mk} \; m = 1, \ldots M - 1. \end{aligned} \qquad \textbf{10}$$

where

β_{pk} is the mean of the change parameter π_{pjk} within neighborhood *k*; and

u_{pjk} is a random person effect.

The person-specific random effects are assumed independent across people but correlated within people, having a multivariate normal distribution with means of zero and variance-covariance matrix T_π .

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

$$\begin{aligned} \beta_{pk} &= \gamma_p + \upsilon_{pk} \\ \alpha_{mk} &= \alpha_m \end{aligned} \qquad \textbf{11}$$

where

γ_p is the population average change parameter of degree p. The neighborhood specific random effects υ_{pk} are independent across neighborhoods but correlated within neighborhoods with a *P-1*-variate normal distribution with means of zero and variance-covariance matrix T_β . Also the three-level model accords the following definitions:

item severity = 0 *for the reference item*

item severity = $-\alpha_m$ *for items* $m = 1,\ldots M-1$.

Person-level predictors of offending propensity may be included in the level-2 model (Equation 10) and neighborhood-level predictors may be included in the level-3 model (Equation 11). In the illustrative example below, we shall consider a model with one type of crime: violent crime consisting of items measures at two different time points. With two different time points the polynomial must be linear, that is of the form $\pi_{0jk} + \pi_{1jk}\, d_{tijk}$, which is equation 9 with p=1.

Illustrative Example

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.[2] Within the 21 strata so constructed, NC's were sampled with the aim of producing a nearly balanced design. The resulting sample is described in Table 1 with census tracts as units.[3] The number of tracts in each stratum is shown in parentheses. The table shows that the confounding of ethnic mix and neighborhood SES precludes study of certain combinations: there are no predominantly white and poor 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.

Insert Table 1 About Here

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

[2] See Sampson, Raudenbush, and Earls (1997) for a detailed description of the construction of the 343 NCs.
[3] 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.

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 in two waves over 54 months from 1994-99. On average, approximately two years elapsed between waves. Though each family was followed until 2002, we confine our attention in this paper to wave-1 and wave-2 data. Sample members were approximately half female; 45% were of Hispanic origin while 36% were Black and 15% 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 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.

Insert Table 2 About Here

For simplicity we will consider only the violent crime items presented in Raudenbush, Johnson, and Sampson (2002). The violent crime scale, displayed in Table 3, includes items self-reported during personal interviews at wave 1 and wave 2. This scale indicates acts of physical aggression (hitting someone you did not live with in the past year with the intent of hurting them, throwing objects at others, robbery, purse snatching, pick pocketing, setting fires, gang fighting, and carrying a hidden weapon).

Insert Table 3 About Here

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. Raudenbush et al. (in press) performed such a comparison using the wave-1 data, leading them to drop one ill-fitting item from the scale. We now examine item functioning of wave-2. Specifically, we estimated a Rasch model (Equation 1) using the program BILOG (Mislevy and Bock, 1997) and compared the results to those based on a two-parameter model (Equation 2), again using BILOG. The results appear in Table 4.

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 .[4] Using a Bayesian Information Criterion (BIC) to compare models, the one-parameter model fits the data very slightly better than does the two-parameter model. That is to say, the BIC of one-parameter model is higher than that of the two-parameter model. The results were similar for wave 1 in that the one-parameter model had a better BIC.

Insert Table 4 and Figure 2 About Here

We graphically display parameter estimates in Figure 2. Note that all the ICCs are forced to be proportional in the case of the one-parameter model (Figure 2a). Even in the case of the two-parameter model, however, (Figure 2b), the ICC's tend to be nearly proportional. Within this scale, "Rasch" assumptions are reasonable thus we opt for the Rasch model. Item location parameters can reasonably be interpreted as item severities, and items and persons arguably are calibrated on a common scale.

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

[4] "BILOG" sets the slope in the one-parameter 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

within-participant variance is more or less than expected under an assumption of independent Bernoulli trials, we have evidence against the local independence assumption. This can be accomplished using the "HLM5" software (see Raudenbush et al., 2000).[5] 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 under-dispersion in all subsequent analyses.

Assessing Item Invariance

A meaningful study of growth requires a constant metric over time. In the Rasch modeling framework, the item severities must be invariant across waves. To assess this assumption, we compare item severities estimated separately for the wave-1 and wave-2 data using Bilog. The results (Table 5) show quite remarkable agreement. Discrepancies between item severities are uniformly small relative to the estimated standard errors.

Insert Table 5 about here

& Bock (1997) and http://www.ssicentral.com/irt/bilog.htm for a copy of the Bilog program and users manual. To obtain the Bilog code used in our analysis please contact the second author at cjque@umich.edu.

[5] 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, $\sigma^2 = 1$.

Fitting The Repeated Measures IRT Model

Setting Up The Data

In order to help us understand the multilevel and repeated measures model, let us first look at the structure of the level-1data, that is, the repeated measures data for each participant. Table 6 contains these repeated measures data for three persons residing in census tract 6109. The first column, labeled "Tract ID" gives the identification number of the census tract in which the participant resides. The second column, labeled "Person ID" is the personal identifier of the participant. The third column labeled "Dwave2" is a dummy variable taking on a value of 0 at wave 1 and 1 at wave 2. The fourth ("Age") column gives the age of the participant at each wave. "ΔAge," the fifth column, is the deviation of the participants age from the mean of the two ages of that participant across the two waves (See note to table 4 for an example on how to compute this). The sixth column contains the values of the outcome variable, Y, for each item at each wave. Column seven through 15 are dummy variables indicating the item with which each outcome Y is associated. Thus, for example, in the first row for participant 5081, we see that Y = 1 and d1 = 1, revealing that person 5081 said "yes" when asked if she had committed the crime of ever caring a hidden weapon in the last year labeled "item 1" at wave 1. Similarly we see that for the second row in the record of person 5081 that Y = 0 and d2 = 1, showing that, at wave 1, participant 5081 said "no" when asked if she had "ever hit someone you don't live with in the last year" (item 2).

Insert Table 6 about here

Table 7a displays the level-2 data, that is, the person level data, for the same 3 participants. Just as in the level-1 data the first two columns are the "Tract ID" and the "Person ID." Thus, for example, we see that person 5081 in tract 6109 had a mean age of 16.29 (we can see from table 6 that person 5081 was 15.19 at wave 1 and 17.39 at wave 2 so that mean age = [15.19 +17.39]/2=16.29). In the analytic model, however, we shall work with mean age-16 and its square (columns 4 and 5) rather than working with the uncentered mean age. Centering age in this way sharply reduces collinearity between age and age-squared. It also defines the level-1 intercept, π_{ojk}, to be the propensity for violent crime at age 16, approximately the peak age of offending. Column six is an indicator for female gender while columns seven and eight indicate ethnicity (Column seven indicates African-American ethnicity while column eight indicated White or other ethnicity). Column nine contains the value for SES which is centered with a mean of zero. Thus, we see for example, that person 5081 is a black female having SES = – 0.074.

Insert Table 7 about here

Table 7b contains the level-3 data, that is the data that varies at the level of the census tract. In this case there is just one variable of interest, concentrated disadvantage that has a mean of zero. Thus we see that tract 6109 has a value of 0.10 on concentrated disadvantage, very slightly above the overall mean.

Applying The Three-Level Model To The Data

We now estimate the three-level model (Equations 9, 10 and 11), expanded to include covariates. The level-1 model views the log-odds of endorsement on item *i* as depending which item was endorsed. We now have

$$\eta_{tijk} = \pi_{0jk} + \pi_{1jk} d_{tijk} + \sum_{m=1}^{7} \alpha_{mjk} a_{mtijk} \qquad \textbf{12}$$

where

a_{mtijk} is an indicator variable representing item;

π_{0jk} is the average log-odds of average criminal propensity of person *j* (averaged over wave 1 and wave2). We will refer to this quantity as person mean propensity;

π_{1jk} is the rate of change per year in the log-odds of endorsing a crime for person *j*;

d_{tijk} is the change in age from wave 1 to wave 2 measured in years;

α_{mkj} represents the average item "severity" of crime item *m* over wave 1 and wave 2 .

Note now that 7 item indicators represent the 8 violent crimes items. For the sake of parsimony these item severities will be again fixed across the children and across the tracts; that is $\alpha_{mjk} = \alpha_m$ for all *j*, *k*.

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

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + \sum_{s=1}^{5} \beta_{0sk} X_{sk} + u_{0jk} \\ \pi_{1jk} &= \beta_{10k} + \sum_{s=1}^{5} \beta_{1sk} X_{sk} + u_{1jk} \\ \alpha_{mjk} &= \alpha_m \; for\ all\ j,k \end{aligned} \qquad \textbf{13}$$

on the measures of violent behaviors, where β_{00k} and β_{10k} are the neighborhood intercepts for tract *k* of person status and change in person mean propensity respectively. β_{0sk} and β_{1sk} are the effects of the indicator variables X_{sk}, $s = 1,\ldots,5$ which are indicators for ethnicity, gender, the mean age of person j (mean-age), and mean person age centered squared $(\text{mean age-16})^2$. The random effects u_{0jk} and u_{1jk} are assumed bivariate normally distributed with zero means and person-level variance covariance matrix

$$\begin{bmatrix} \tau_{\pi 00} & \tau_{\pi 01} \\ \tau_{\pi 10} & \tau_{\pi 1} \end{bmatrix} . \qquad \textbf{14}$$

Level-3 Model. The level-3 model accounts for variation between tracts on person mean propensity and the change in person mean propensity respectively:

$$\begin{aligned} \beta_{00k} &= \gamma_{00} + \gamma_{01} W_k + \upsilon_{0k}, \\ \beta_{10k} &= \gamma_{10} + \gamma_{11} W_k + \upsilon_{1k}, \\ \beta_{0sk} &= \gamma_{0s} \ , \ s > 1 \\ \beta_{1sk} &= \gamma_{1s} \ , \ s > 1 \end{aligned} \qquad \textbf{15}$$

where

γ_{00} and γ_{10} are the overall intercepts for status in propensity and the rate of change per year in propensity respectively;

W_k is the degree of neighborhood concentrated disadvantage in neighborhood k;

and the random effects ν_{0K} and υ_{1k} are assumed bivariate normally distributed with zero means and tract-level variance covariance matrix

$$\begin{bmatrix} \tau_{\beta 00} & \tau_{\beta 01} \\ \tau_{\beta 10} & \tau_{\beta 11} \end{bmatrix}. \qquad \textbf{16}$$

For simplicity we fix β_{0sk} and β_{1sk}.

Results

Item severities. Table 8 provides the severities of the items within each scale (note that these item severities are averaged over wave 1 and wave 2). 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.

Insert Table 8 About Here

In contrast, hitting someone you don't live with, throwing objects at someone and carrying a hidden weapon are less severe.

Association between covariates and mean propensity. Our three-level model specifies mean propensity and change in propensity as depending on covariates measured at the level of the person and the neighborhood. Correlates of mean propensity appear in Table 9 (under "mean propensity"). We see a strong quadratic effect of age, to be discussed in more detail below (see "describing the age-crime curve"). Controlling for age and the other covariates, we find significant effects of ethnicity: African Americans display higher propensity than do Hispanics (coeff. = 1.099, odds ratio = exp{1.099} = 3.00, t = 10.54). Whites also display higher propensity than do Hispanics (coeff = 0.253, odds ratio = 1.29, t = 2.02). We can deduce that the odds for African Americans, relative to Whites is 3.00/1.29 = 2.33. Similarly, females display lower propensity than do males (coeff. = -.767, odds ratio = .46, t = -9.27). Neighborhood concentrated disadvantage increases the risk of crime (coeff = 0.214, t = 2.98). Given the standard deviation of 0.82 sd, a two sd increase in concentrated disadvantage is associated with an increase of 2*(0.82)*(0.21) = .351, yielding an odds ratio of 1.42. Interestingly, personal SES is unassociated with violent crime after controlling concentrated disadvantage.

Insert Table 9

Association between covariates and change in crime (see Table 9, under "change in propensity"). We note first an overall reduction in propensity per year of age, evaluated at the age of 16, coeff = -0.133, t = -2.06. However, the expected change depends on age. Controlling for gender, ethnicity, and concentrated disadvantage, the expected change in propensity = -0.133 - 0.041*(age-16) + (.002)*(age - 16)^2. We can easily tabulate the expected change in propensity, therefore, for a typical participant in each of our four cohorts. Specifically, a typical person in the first cohort would have been 9 at wave 1 and 11 at wave 2. For the second, third, and fourth cohorts, the corresponding ages would be: 12 and 14 for cohort 2; 15 and 17 for cohort 3; and 18 and 20 for cohort 4. Using the mid point of each interval (e.g., using 10 for cohort 1, 13 for cohort 2, etc.), we calculate the expected change for each "typical" cohort member:

Cohort	Age at wave 1	Age at wave 2	Expected change/year
1	9	11	1.85
2	12	14	0.01
3	15	17	-0.13
4	18	20	-0.24

The expected trajectories are graphed for four Hispanic males who are of typical age for their cohort (Figure 3a). The Y axis is the probability of committing violent crime, defined here in terms of the reference item ("ever hit someone you do not live with during the past year"). We see a rather sharp increase in the probability of committing crime for younger participants (those less than 13). However, the rate of increase becomes negative thereafter and falls sharply for the oldest cohort.

Insert Figure 3 About Here

Describing the age-crime curve. The foregoing discussion suggests that our participants' rates of increase in committing violent crime tended to be large in early adolescence. These rates of increase eventually diminish as a function of age, reaching zero at about age 13, suggesting that the probability of violent crime is at its maximum at around 13. This seems to contradict conventional wisdom about the age-crime curve, suggesting that violent crime tends to peak later. However, we know that Chicago and many other large cities experienced a sharp drop in the rate of violent crime during the years of our data collection, 1995-1999. In a longitudinal study, age and history are confounded. Presumably, the rates of change by age reflect combined effects of age and history. To probe these puzzles further, Figure 3b graphs the expected age-crime curve for Hispanic males (of average SES and average neighborhood concentrated disadvantage) as a function of age in 1996 and 1998 (roughly the midpoints of waves 1 and 2, respectively).[6] Note that the age-crime curves have the "classic" forms at each wave. The wave 1 curve peaks at about age 17 while the wave 2 curve peaks at about age 16. Vertical distances between the curves at any age can be interpreted as combined history and cohort effects. These effects are most pronounced at the later ages. Within each wave, the age-crime curves reflect cohort and age effects. To reduce cohort effects, we have

[6]The wave 1 curve is computed by substituting -1.0 for age-change in our model, allowing age to vary from 9 to 18 and setting other covariates to zero. The wave 2 curve is computed by setting age-change to 1.0, allowing age to vary from 11 to 20, again holding other covariates at zero.

controlled for ethnicity, gender, concentrated disadvantage, and SES. Increasingly precise control for demographic covariates would presumably eliminate most of the cohort effects, yielding graphical interpretations of history effects (vertical distances between the graphs at each age) and age effects (associations between cross-sectional age and crime at each wave. The collection of data on multiple cohorts at multiple ages plus the availability of rich demographic data lay the basis for separating age, history, and cohort effects.

Conclusions

A meaningful metric is essential in studying individual change and correlates of change. The construction of such a metric requires that the variable that is repeatedly measured have the same meaning at each age. Item response theory (IRT) provides a sensible way to create and assess metrics used to assess change.

Our approach to constructing such a metric had three distinct stages. First, we used available theory to tentatively assign items to common scales. In Raudenbush et al. (in press), we assigned nine items to the violent crime scale. Second, we assessed the fit of the Rasch model within each wave. In Raudenbush et al. (in press), we discarded an item that did not fit a Rasch scale, leaving eight items that did fit the scale. A scale that fits the strong Rasch assumption of additivity has an extremely nice property: crime items more rarely endorsed can be interpreted as more "severe" than crime items more readily endorsed in that only the most serious criminals endorse those rare items. The severity of the item does not depend on the criminal propensity of the person, an extremely helpful property when trying to measure changes in propensity.

Having scales within each wave that meet Rasch assumptions does not insure that the metric is invariant as a function of age. To assess this requirement, we scaled the items separately within waves. The results were encouraging (Table 5); the item severities showed little evidence of depending on age.

In taking the next step, it was tempting to create a single scale score for each participant at each wave, to view these scale scores as continuous, and to model correlates of change in a hierarchical model. We resisted this temptation because of the small number and skewed nature of our item responses. Many of our participants never endorsed a single item! To assign these participants a scale score would involve an extrapolation with large error. Moreover, the data would never approximate continuity. To model correlates of change, we therefore used a three-level logistic regression model, with item responses at level 1, persons at level 2 and neighborhoods at level 3. In essence, we modeled the probability of committing each of eight crimes as a function of covariates, but imposing Rasch assumptions: that the log-odds of each crime was equal apart from a constant. This can be viewed as a proportional odds assumption. Our checks on Rasch assumptions within each wave and across the two waves gives support to this constraint. Alsothis the constraint adds statistical power and ease of interpretation to the results.

Although longitudinal data afford many advantages, interpretation of results can nonetheless be quite ambiguous. In a longitudinal study of a single cohort, age and history are perfectly confounded. Cross-sectional studies, of course, confound age and cohort. Collecting data on multiple cohorts and multiple aspects of demography creates the possibility of separating age, period, and cohort effects. In our study, we know that young people were maturing at a time when the rate of violent crime was dropping in Chicago and many other cities. Thus, within-participant change confounds maturation and history effects. The age-crime curves at each wave give some indication of the magnitude of these history effects (see Figure 3), though cohort effects are implicated as well. Precise control for participant demography in conjunction with data on multiple cohorts repeatedly observed over time creates potential, in principle, of separating age, history, and cohort effects in longitudinal studies.

REFERENCES

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

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. Scientific Software International.

Osgood, D. W., B. J. McMorris, and M. T. Potenza. In press(a). "Analyzing Multiple-item Measures of Crime and Deviance 1: Item Response Theory Scaling." To appear in Journal of Quantitative Criminology.

Osgood, D. W., L. L. Finkin and B. J. McMorris. (in press b). Analyzing Multi-Item Measures of Crime and Deviance II: Tobit Regression Analysis of Transformed Scores. To appear in Journal of Quantitative Criminology.

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, Second Edition. Newbury Park, CA: Sage.

Raudenbush, S., M. L. Yang, & M. Yosef. 2000. "Maximum Likelihood for Hierarchical Models via High-order, Multivariate LaPlace Approximation." Journal of Computational and Graphical Statistics 9(1): 141-157.

Raudenbush, S. W. and R. Sampson, . 1999. "Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods." Sociological Methodology 29: 1-41.

Sampson, R. J. S. Raudenbush, and F. Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." Science 277: 918-924.

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

Table 1. Number of census tracts (N = 196) by Socio-Economic Status (SES)[a] and Racial/Ethnic Composition in the PHDCN Design. The 80 sampled NC=s are shown in parentheses.

	SES							
Racial/Ethnic Strata	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 & 20% White	11	(4)	14	(5)	10	(4)	35	(13)
20% Hispanic & 20% Black	7	(4)	7	(4)	0	(0)	14	(8)
20% Black & 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.

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) percent of population below the poverty line, 2) percent of population that is on some form of public assistance, 3) percent of population that is unemployed, 4) percent of population that is less than 18 years of age, and 5) percent of population whose households are female headed. Sampson, Raudenbush, and Earls (1997) provides a detailed description of the construction of the scale.

Table 3. Item responses for Violent crime by wave

VARIABLE	CATEGORY	FREQUENCY	
		Wave 1	Wave 2
Hit someone with whom you did not live with in the past year with the idea of hurting them	0 = no	2129	1960
	1 = yes	685	451
Thrown objects such as bottles or rocks at people in the past year	0 = no	2495	2201
	1 = yes	333	210
Ever carried a hidden weapon in the last year	0 = no	2565	2221
	1 = yes	258	191
Ever maliciously set a fire in the last year	0 = no	2809	2402
	1 = yes	28	9
Ever snatched a purse/picked a pocket in the last year	0 = no	2821	2401
	1 = yes	18	10
Ever attacked with a weapon in the last year	0 = no	2739	2342
	1 = yes	95	69
Ever used a weapon to rob someone in the last year	0 = no	2822	2403
	1 = yes	10	8
Ever been in a gang fight in the last year	0 = no	2667	2307
	1 = yes	167	97

Table 4. Results for one-parameter and two-parameter models for violence at wave 2

Violence							
		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
1 (attacked with weapon)	1.075	2.524 (0.089)	2.355 (0.104)	2.118	2.429 (0.124)	2.674 (0.348)	0.888
2 (hit some-one you don't live with)	0.668	1.286 (0.049)	2.355 (0.104)	0.685	1.280 (0.059)	2.397 (0.258)	0.685
3 (throw objects at someone)	0.756	1.846 (0.061)	2.355 (0.104)	1.893	1.995 (0.097)	1.958 (0.183)	1.143
4 (carried hidden weapon)	0.854	1.909 (0.064)	2.355 (0.104)	0.191	1.931 (0.087)	2.308 (0.229)	0.300
5 (set fire)	0.904	3.561 (0.161)	2.355 (0.104)	3.350	4.004 (0.591)	1.887 (0.551)	3.665
6 (snatched purse)	0.967	3.511 (0.162)	2.355 (0.104)	1.379	3.790 (0.457)	2.035 (0.471)	1.284
7 (used force to rob)	1.016	3.616 (0.177)	2.355 (0.104)	0.814	3.839 (0.509)	2.108 (0.566)	1.273
8 (be in a gang fight)	0.941	2.329 (0.078)	2.355 (0.104)	0.617	2.381 (0.121)	2.254 (0.242)	0.757
BIC (max BIC [in bold] is best)			**-3214.85**			-3244.87	

Table 5. Results of the one parameter model for violence at wave1 and the one parameter model for violene at wave 2.[a]

Item	Wave 1 Threshold (s.e.)	Wave 2 Threshold (s.e.)
1 (attacked with weapon)	2.627 (0.072)	2.524 (0.089)
2 (hit some-one you don't live with)	1.240 (0.037)	1.286 (0.049)
3 (throw objects at someone)	1.821 (0.046)	1.846 (0.061)
4 (carried hidden weapon)	2.001 (0.050)	1.909 (0.064)
5 (set fire)	3.298 (0.102)	3.561 (0.161)
6 (snatched purse)	3.482 (0.109)	3.511 (0.162)
7 (used force to rob)	3.812 (0.156)	3.616 (0.177)
8 (be in a gang fight)	2.301 (0.059)	2.329 (0.078)

[a]Wave-1 item severities were re-scaled to have the same mean and standard deviation as the wave-1 severities.

Table 6: Level 1 Record for Subjects 4061, 5081, 5121.

Tract	Person ID	Dwave2	Age	ΔAge[a]	Y	d1	d2	d3	d4	d5	d6	d7	d8
6109	4601	0	12.45	-1.09	0	1	0	0	0	0	0	0	0
6109	4601	0	12.45	-1.09	0	0	1	0	0	0	0	0	0
6109	4601	0	12.45	-1.09	0	0	0	1	0	0	0	0	0
6109	4601	0	12.45	-1.09	0	0	0	0	1	0	0	0	0
6109	4601	0	12.45	-1.09	0	0	0	0	0	1	0	0	0
6109	4601	0	12.45	-1.09	0	0	0	0	0	0	1	0	0
6109	4601	0	12.45	-1.09	0	0	0	0	0	0	0	1	0
6109	4601	0	12.45	-1.09	0	0	0	0	0	0	0	0	1
6109	4601	1	14.63	-1.09	0	1	0	0	0	0	0	0	0
6109	4601	1	14.63	-1.09	0	0	1	0	0	0	0	0	0
6109	4601	1	14.63	-1.09	0	0	0	1	0	0	0	0	0
6109	4601	1	14.63	-1.09	0	0	0	0	1	0	0	0	0
6109	4601	1	14.63	-1.09	0	0	0	0	0	1	0	0	0
6109	4601	1	14.63	-1.09	0	0	0	0	0	0	1	0	0
6109	4601	1	14.63	-1.09	0	0	0	0	0	0	0	1	0
6109	4601	1	14.63	-1.09	0	0	0	0	0	0	0	0	1
6109	**5081**	**0**	**15.19**	**-1.1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**
6109	**5081**	**0**	**15.19**	**-1.1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
6109	**5081**	**1**	**17.39**	**1.1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**
6109	**5081**	**1**	**17.39**	**1.1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
6109	5121	0	18.32	-1.24	0	1	0	0	0	0	0	0	0
6109	5121	0	18.32	-1.24	0	0	1	0	0	0	0	0	0
6109	5121	0	18.32	-1.24	0	0	0	1	0	0	0	0	0
6109	5121	0	18.32	-1.24	0	0	0	0	1	0	0	0	0
6109	5121	0	18.32	-1.24	0	0	0	0	0	1	0	0	0
6109	5121	0	18.32	-1.24	0	0	0	0	0	0	1	0	0
6109	5121	0	18.32	-1.24	0	0	0	0	0	0	0	1	0
6109	5121	0	18.32	-1.24	0	0	0	0	0	0	0	0	1
6109	5121	1	20.80	1.24	0	1	0	0	0	0	0	0	0
6109	5121	1	20.80	1.24	0	0	1	0	0	0	0	0	0
6109	5121	1	20.80	1.24	0	0	0	1	0	0	0	0	0
6109	5121	1	20.80	1.24	0	0	0	0	1	0	0	0	0
6109	5121	1	20.80	1.24	0	0	0	0	0	1	0	0	0
6109	5121	1	20.80	1.24	0	0	0	0	0	0	1	0	0
6109	5121	1	20.80	1.24	0	0	0	0	0	0	0	1	0
6109	5121	1	20.80	1.24	0	0	0	0	0	0	0	0	1

[a]ΔAge is defined as the difference between age and mean age where mean age = (age at time 1 + age at time 2) divided by 2. For example, in the case of person 5081 within tract 6109 (see above table) mean age = (15.19 + 17.39) / 2 = 16.29, and ΔAge = 15.19 – 16.29 = -1.10 at wave 1 and 17.39 – 16.29 = 1.10 at wave 2.

Table 7. Corresponding Level 2 and Level 3 Records for Subjects 4601, 5081, 5121.

a. Level 2 Record for Subjects 4601, 5081, 5121.

Tract ID	Person ID	Mean Age	Mean Age-16	$(\text{Mean Age-16})^2$	Female	Black	White & Other	SES
6109	4601	13.54	-2.46	6.04	1	0	1	1.00
6109	**5081**	**16.29**	**.29**	**0.08**	**1**	**1**	**0**	**-0.74**
6109	5121	19.56	3.56	12.68	1	0	1	2.11

b. Level 3 Records for Subjects 4601, 5081, 5121.

Tract	Concentrated Disadvantage
6109	**0.10**

Table 8. Model fitting results: Item Severity

ITEM	COEFFICIENT ESTIMATE
Intercept[a], γ_0	-1.92
Ever used a weapon to rob someone in the last year, α_1	-7.03
Ever snatched a purse/picked a pocket in the last year, α_2	-6.54
Ever maliciously set a fire in the last year, α_3	-6.23
Ever attack with a weapon in the last year, α_4	-4.43
Ever been in a gang fight in the last year, α_5	-3.79
Ever carried a hidden weapon in the last year, α_6	-3.01
Ever thrown objects at someone in the last year, α_7	-2.71

[a]The reference item is "ever hit someone you don't live with this year."

Table 9. Association between covariates and crime

a) Fixed Effects

Predictor	Mean Propensity			Change in Propensity		
	Coeff.	SE	t-ratio	Coeff.	SE	t-ratio
Intercept	-1.919	0.090	-21.331	-0.133	0.065	-2.061
Concentrated Disadvantage	0.214	0.071	2.983	-0.003	0.055	-0.070
SES	0.031	0.035	0.896	-0.040	0.024	-1.660
African American	1.099	0.104	10.539	-0.039	0.077	-0.508
White & Other_0	0.253	0.125	2.020	-0.160	0.090	-1.775
Female	-0.767	0.083	-9.265	-0.067	0.058	-1.170
Mean Age-16	0.059	0.018	3.357	-0.041	0.011	-3.635
$(\text{Mean Age-16})^2$	-0.042	0.004	-9.688	0.002	0.003	0.743

b) Covariance Component

Variance component

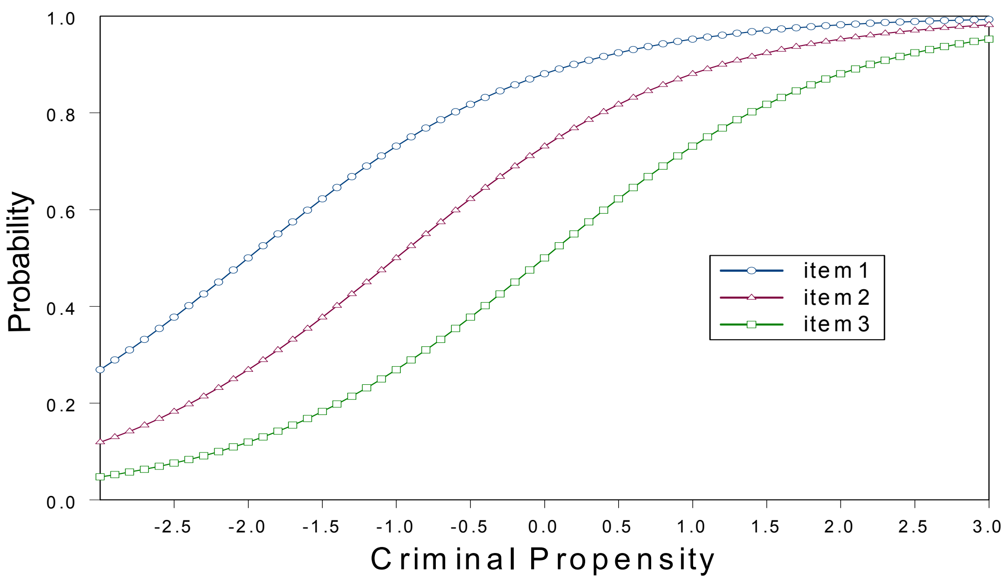
Level-1, σ^2: 0.336

Level-2, τ_π: $\begin{bmatrix} 3.12 & 0.03 \\ 0.03 & 0.81 \end{bmatrix}$

Level-3, τ_β: $\begin{bmatrix} 0.009 & -0.013 \\ -0.013 & 0.019 \end{bmatrix}$

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.

(a) $\lambda_1=\lambda_2=\lambda_3=1$, $\psi_1= -2$, $\psi_2 = -1$ $\psi_3 = 0$.



(b) $\lambda_2=\lambda_3=1$, $\lambda_1=0.3$, $\psi_1= -2$, $\psi_2 = -1$ $\psi_3 = 0$.

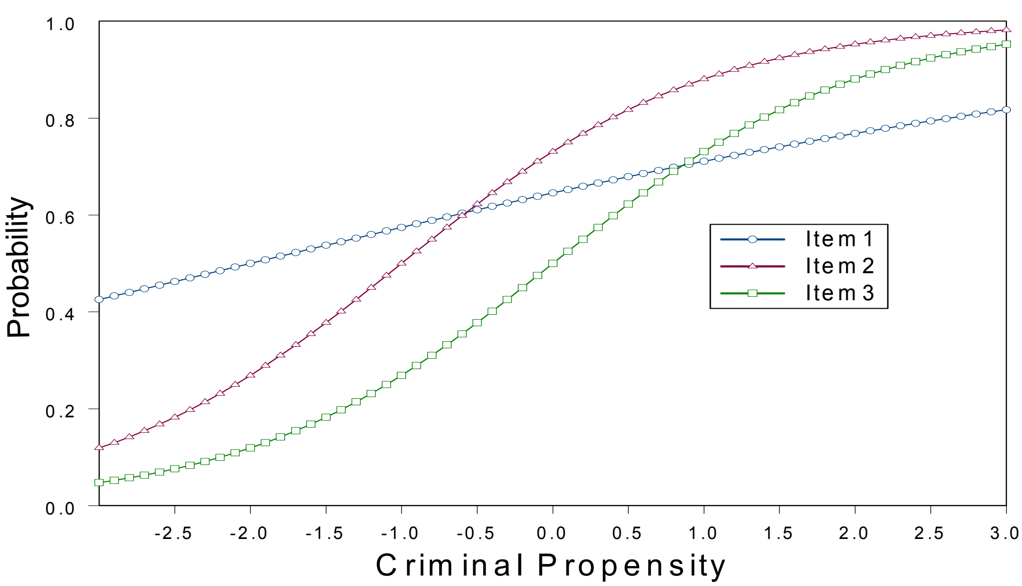
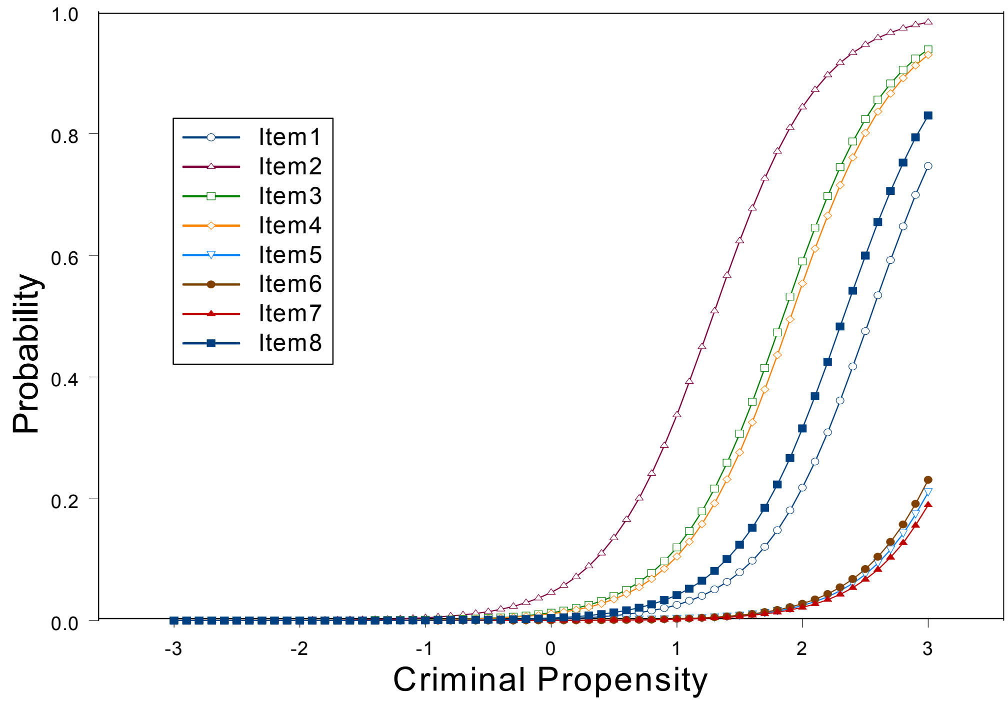


Figure 2. Graphical comparison of 1- and 2-parameter models for violent crime at wave 2.

(a)



(b)

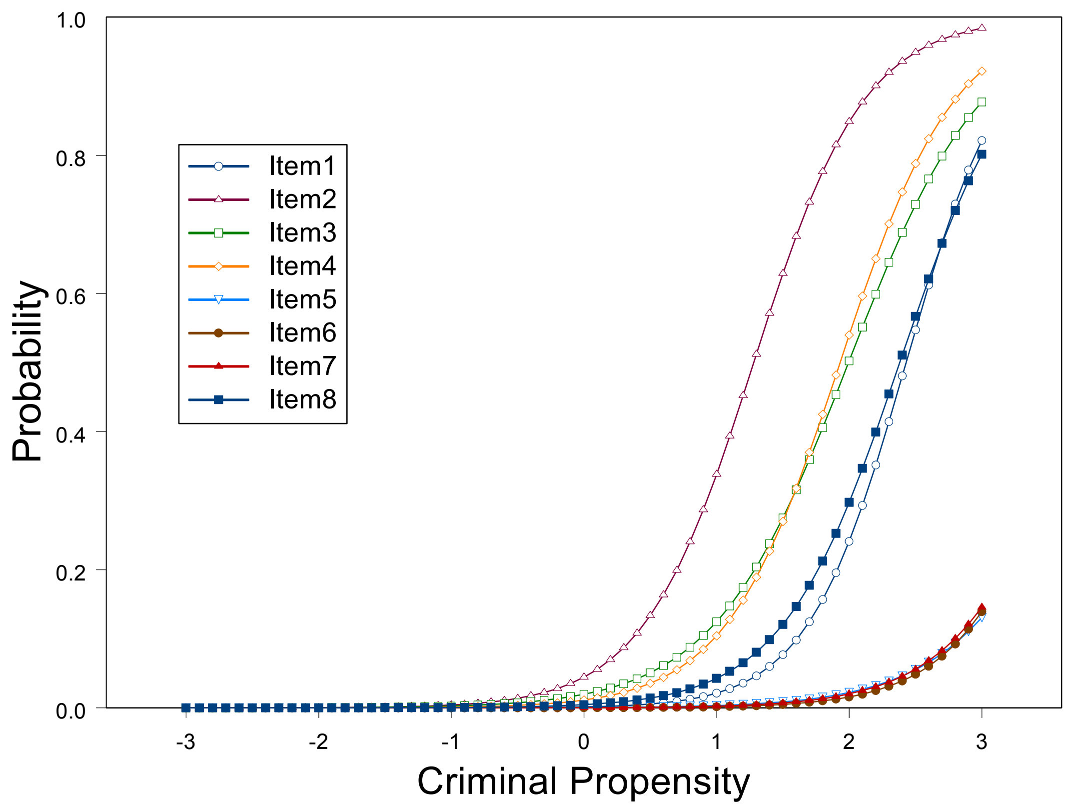


Figure 3a. Change in the probability of offending of respondents between wave 1 and wave 2

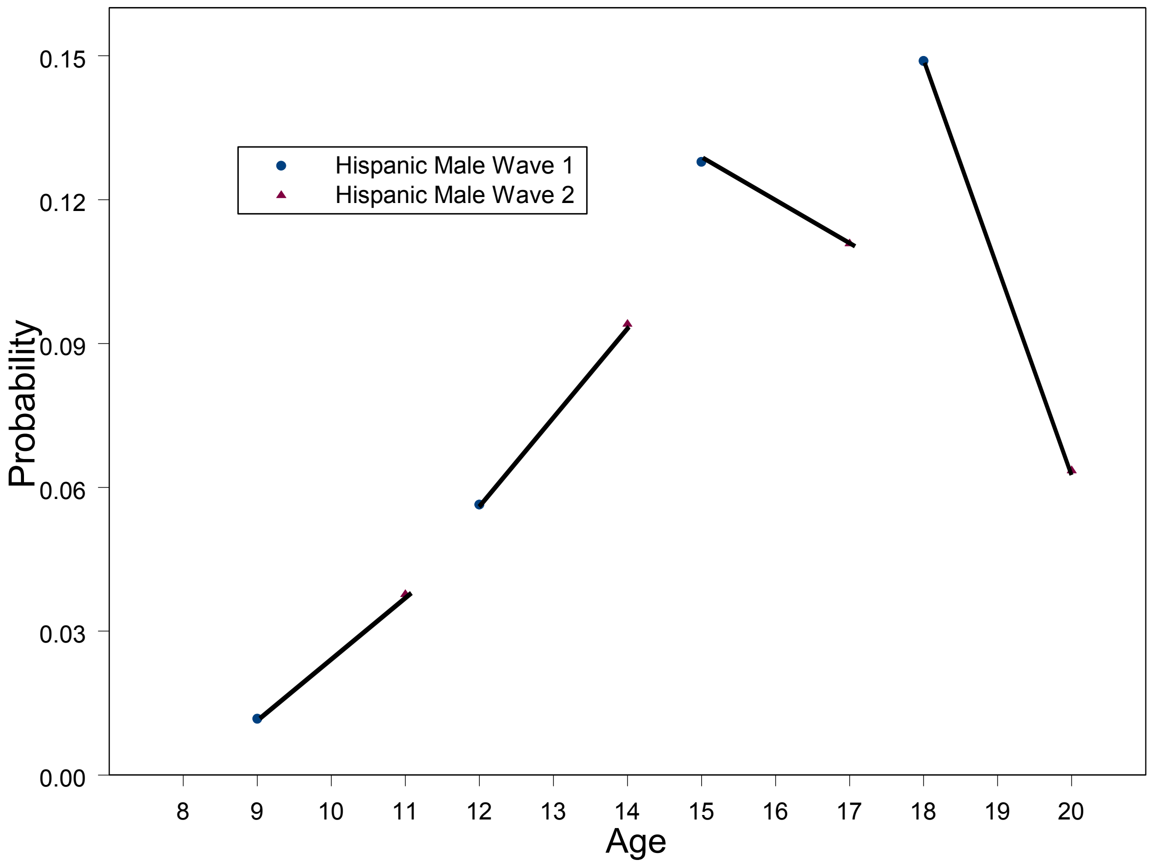


Figure 3b. Change in the probability of offending of respondents between wave 1 and wave 2 with cross sectional age crime curves superimposed for each wave.