Published in final edited form as:
Youth Violence Juv Justice. 2004 January 1; 2(3): 273. doi:10.1177/1541204004265878.

THE SELF-REPORT OF OFFENDING AMONG SERIOUS JUVENILE OFFENDERS:

Cross-Gender, Cross-Ethnic/Race Measurement Equivalence

George P. Knight,
Arizona State University

Michelle Little,
Temple University

Sandra H. Losoya, and
Arizona State University

Edward P. Mulvey
University of Pittsburgh

Abstract

This article evaluates the measurement equivalence of a self-report of offending measure among female and male juveniles, as well as Hispanic, African American, and White male juveniles. The findings indicate (a) considerable functional equivalence across gender and ethnically/racially diverse groups of juvenile offenders, and (b) scalar equivalence across Hispanic and White male juvenile offenders, but (c) that researchers should be careful making either mean difference or association comparisons across genders or African American/White boys.

Keywords

self-report of offending; measurement equivalence; juvenile offending

The empirical search for factors that are associated with, and perhaps lead to, desistance among serious juvenile offenders is a topic of great scientific and practical importance, and the primary focus of the articles presented in this volume. The empirical search for these factors requires that we have some way of operationally defining (i.e., measuring) the end state and the process of desistance. As Mulvey et al. (2004 [this issue]) point out, this is not a simple task with one clearly accepted approach. Although one could use the absence of subsequent arrests, or adjudications/convictions, as an indicator of desistance, the problem here is that the probability of any single juvenile offender being arrested, let alone adjudicated or convicted, for any single illegal offense is very low (Brame, Fagan, Piquero, & Schubert, 2004 [this issue]; Cohen, 1986). Hence, in any sample of serious juvenile offenders, many may be subsequently, and erroneously, identified as desisting simply because they had not been caught or adjudicated/convicted of engaging in illegal activities.

One typical solution to this problem is to use self-reports of offending (SRO) as an indicator of involvement in antisocial or criminal behavior; that is, researchers (e.g., Elliott & Ageton, 1980; Hindelang, Hirschi, & Weis, 1981; Huizinga & Elliott, 1986; Thornberry & Krohn, 2000) have asked juveniles to report whether they have committed specific acts (that happen

to be illegal) within a past specified time period. Then a substantial decline in the SRO over time associated with some absolute low threshold of offending can be used to identify juveniles who are desisting relative to those who are not desisting (Bushway, Piquero, Broidy, Cauffman, & Mazerolle, 2001).

For the study of desistance, then, much rests on the accuracy of these self-reports. Fortunately, the available empirical evidence suggests that these types of self-reports of delinquent activities have considerable reliability and validity (for a review of this literature see Thornberry & Krohn, 2000). However, as Thornberry and Krohn (2000) noted, "There is an unresolved issue of differential validity" across diverse subgroups of juvenile offenders, and that this "is perhaps the most important methodological issue concerning the self-report method" (p. 58) of assessing delinquent behaviors. The current article addresses this issue. Specifically, we consider the cross-group measurement equivalence of a variety score derived from a SRO measure for female and male juveniles, and for Hispanic, African American, and White juveniles.

Although currently there is no clearly agreed-on single approach for definitively assessing measurement equivalence, a number of authors (e.g., Hines, 1993; Hughes, Seidman, & Williams, 1993; Hui & Triandis, 1985; Knight & Hill, 1998; Malpass & Poortinga, 1986) have described different forms of measurement equivalence that require a range of analytical approaches. Hui and Triandis (1985) categorized these forms of equivalence into item equivalence, functional equivalence, and scalar equivalence. Item equivalence exists when the items on a measure have the same meaning across the groups of interest. For example, if respondents are asked if they have "taken something from another person by force," it is important that (a) individuals from diverse groups each mean the same thing when thinking about the use of force and (b) similar relationships between the response to this item and responses to other items measuring SRO exist across groups. Functional equivalence exists when the scores generated by a measure have similar precursors, consequents, and correlates across the groups of interest; that is, if the scores derived from a SRO measure correlate with scores derived on measures of other constructs (such as arrest history, exposure to delinquent peers, parental monitoring, etc.) in a manner consistent with the theory regarding the nature of the SRO construct (i.e., construct validity coefficients), and if these correlations are consistent with the available theory and empirical knowledge regarding the nature of the SRO construct within the specific groups of interest, then there is some evidence of functional equivalence of the SRO measure. Scalar equivalence exists when a given score on a measure refers to the same degree, intensity, or magnitude of the target construct across the groups of interest; that is, if two individuals from different groups show the same decline in the variety of offenses they report over a defined period of time, then these two individuals should be desisting to the same degree. If, however, there is a differential bias in reporting associated with group membership, such as if the members of one gender or one race/ethnic group are more likely to exaggerate or minimize their illegal activities, then any specific SRO score is not indicative of the same degree of desistance across these groups. Scalar equivalence is the most important, and most difficult, level of equivalence to demonstrate empirically.

Although there are many useful analytical procedures for examining the cross-group equivalence of measures (see Knight & Hill, 1998), much of the writing and empirical investigations of measurement equivalence have suffered from three very important limitations. First, researchers have often assumed that cross-group measurement equivalence (scalar and all lower levels of equivalence) is only necessary when one is interested in the examination of mean differences across groups (e.g., Steenkamp & Baumgartner, 1998). Essentially, if the comparison of scores from different groups is to reflect any real difference between groups, it is necessary that any given score on the measure on which the groups are being compared reflect the same magnitude of the target construct in each group. However,

measurement equivalence is an important consideration even if one does not wish to make cross-group mean comparisons (see Knight, Tein, Prost, & Gonzales, 2002, for a more thorough discussion of this point).

Lack of measurement equivalence can also bias the estimate of the strength of a relationship between constructs in different groups. Consider the case of researchers examining the correlation between measures of two constructs in a sample that includes Latino and non-Latino White participants. There is evidence that Latinos, compared to non-Latino Whites, more frequently endorse extreme responses (Hui & Triandis, 1985; Marin, Gamba, & Marin, 1992). Assuming that the measures of the two target constructs have similar response formats that include a wide array of responses, this ethnically linked response bias could result in an inflated estimate of the correlation between these constructs because the two measures share a common measurement artifact. Alternatively, if the measures of the two constructs have quite different response formats, then the ethnically linked response bias may simply increase the unexplainable variance in one or both of the measures. This additional measurement error will attenuate, or underestimate, the correlation between the two measures. Thus, even when no group comparisons are made, the absence of measurement equivalence for a central construct, such as offending measured using a SRO scale, can result in differential errors in measurement across groups, and even inaccurate conclusions about the relations between a SRO measure and other constructs.

Second, researchers have often failed to distinguish between measurement invariance and measurement equivalence (e.g., Steenkamp & Baumgartner, 1998). The examination of measurement invariance involves an assessment of the cross-group similarity of the interitem or item-total relationships, usually through confirmatory factor analysis or item response theory (differential item functioning) analyses. However, Hui and Triandis (1985) pointed out that the evaluation of measurement equivalence across groups involves multiple forms of analytical evaluations beyond just item-level tests of measurement invariance. Examination of the equivalence of measurement must also be done at the various conceptual levels that might be considered in analyses; for example, if general summary measures of a construct are going to be used with different groups, then this variable must be tested for its equivalence across these groups as well. Hence, although item-level analyses may be quite useful for evaluating the item equivalence of a SRO measure, they do not tell the whole story. These types of analyses provide only limited information regarding functional and scalar equivalence. The empirical examination of the construct validity coefficients associated with a SRO measure within the specific groups in which the measure will be used (see Knight & Hill, 1998; Knight et al., 2002, for more detail) is needed to get a full picture of functional and scalar equivalence.

Third, researchers have often failed to acknowledge and understand the critical role of theory and previous knowledge regarding the nature of the construct being measured and the groups to which the measure will be administered (see Knight & Hill, 1998; Steenkamp & Baumgartner, 1998, for examples of such failures); that is, theory and previous knowledge regarding the nature of the construct being measured should specify the indicators of this construct as well as the nomological net of the constructs that should, and should not, be related to that construct. Moreover, this theory/knowledge should be informed by an understanding of the populations in which the measure is administered (see Knight et al., 2002 for a more thorough discuss of this point). This means that, in this case, theory should create an expectation regarding exactly how the individual SRO item scores should be interrelated or related to the SRO total score in the population(s) being studied. Similarly, theory should create an expectation regarding what constructs should, and should not, be related to SRO in the population(s) being studied.

This failure to differentiate between measurement invariance and measurement equivalence, associated with the absence of consideration of the role of theory/knowledge about the construct in the specific groups being assessed, is particularly important because there may well be cases in which partial measurement invariance (i.e., a failure of full measurement invariance) is necessary to ascertain cross-group equivalence. For example, if there are gender-specific forms of offending, it may well be the case that SRO items that correspond to those specific offenses will relate to the total SRO score or the bulk of the SRO item scores differently across genders. For example, in this particular SRO measure, an item such as "Have you ever forced someone to have sex with you?" is likely to be endorsed much less frequently by women. If certain items are endorsed rarely by one gender, then there may well be so little variability on these items for one gender that the correlations of the scores on these items with the remaining items may be low enough because of attenuation to effect the factor loadings (or item characteristic curves in item response theory [IRT] analyses) on these items for one gender but not the other. Nevertheless, it may be necessary to include these specific items in the SRO measure to generate a functional or scalar equivalent measure and to get a more accurate gender comparison of offending. Ultimately a small number of gender-specific illegal activities may well lead to failure of tests of measurement invariance; however, this specific noninvariance may be quite appropriate when trying to accurately measure offending among male and female juveniles.

The current study is designed to evaluate the cross-gender and cross-ethnic/race measurement equivalence of a SRO measure by examining the measurement invariance and construct validity coefficients following the procedures described by Knight and colleagues (Knight & Hill 1998; Knight et al., 2002).[1] Knight and Hill (1998) contended that the evaluation of the equivalence of a measure of any construct, such as SRO, requires item-level analyses as well as scale-level analyses. The item-level analyses may be based on either the interrelations among the SRO items such as in confirmatory factor analyses, or the relations of the SRO item scores to the total SRO score such as in item response theory analyses. The scale-level analyses must be based on the relations among the SRO total scores and scores on other measures of theoretically related constructs (i.e., other indicators of criminal activities and associated psychological constructs) using multiple regression or structural equation modeling types of analyses.

Method

Participants

The full sample for the current study consisted of 1,354 juvenile offenders enrolled in the Research on Pathways to Desistance Project (see Schubert, Mulvey, Cauffman, Steinberg, Losoya, Hecker, et al., 2004 [this issue], for a more detailed description of the enrollment methods and sample). Initial inspection of data showed that data for all 24 items of the SRO were missing at the time these analyses were conducted for 16 participants. Thus, the sample for the current study's analyses was reduced to 1,338 juveniles (181 girls and 1,157 boys). Among the boys, 493 were African American, 393 were Hispanic, 218 were White, 23 were Native American, 2 were Asian American, and 28 were of unknown ethnic/racial origin. Most of the sample (86.5%) were boys, with an age range of 14 to 17 years at the time of baseline interview. One fifth of the sample (19.9%, $n = 266$) were White, two fifths (41.9%, $n = 561$) were African American, one third were Hispanic (33.4%, $n = 447$), 2% ($n = 27$) were Native American, less than 1% ($n = 2$) were of Asian origin, and the remaining 2.6% ($n = 35$) considered themselves to be of another ethnic origin.

[1]Given the difficulty in obtaining female juveniles who are serious offenders, the current article examines measurement equivalence of SRO for ethnically and racially diverse samples of male and female offenders. In addition, because of the size of our sample of female juvenile offenders, the examination cross-ethnic/race equivalence of the SRO is based on only a sample of male offenders.

Measures

SRO—The Self-Report of Offending (Huizinga, Esbensen, & Weiher, 1991) was adapted for the current study to measure the juveniles' account of involvement in antisocial and illegal activities. To assess lifetime offending, 24 items are used to assess whether the juvenile had engaged in criminal activities, including aggressive crimes (e.g., "Taken something from another person by force, using a weapon?"), income-generating crimes (e.g., "Used checks or credit cards illegally?"), and public order offenses (e.g. "Ever driven while intoxicated?"). An overall offending variety score was created and is the number of specific offenses, of the 24 offense items, the person has ever committed (from baseline interview responses). These types of variety scores have often been used to index criminal activity (Hindelang et al., 1981), and there have been some empirical investigations that have examined the validity of such scores (Huizinga & Elliott, 1986; Piquero, MacIntosh, & Hickman, 2002). Given that the current study reports on the cross-ethnic/race psychometric properties of the SRO, as well as the measures of related constructs, reliability and validity coefficients are not reported here for any of the measures.

Offense history—Fourteen items were administered regarding juveniles' offense histories. For the purposes of validating the SRO measure, three items from our Offense History Battery will be included in the current analyses. These items are "How many times have you been arrested in your lifetime?" "How old were you the first time you were arrested?" and "Before the arrest or summons that led to participation in this study, how many times were you locked up in a detention center or jail?" In addition, the number of prior petitions juveniles had received in the juvenile/adult system was obtained from official court records. Juveniles' history of judicial petitions provides an estimate of the number of substantiated charges received by juveniles before their baseline interview for study participation.

Gang membership—Lifetime gang membership was coded from juveniles' responses to two questions. Each juvenile was asked whether he or she was a member of a gang at the time of the arrest that made him or her eligible for the study. If they responded negatively, they were asked whether they had ever been a member of a gang in the past. Positive responses to either question were coded as 1 for lifetime gang membership. This measure was used to examine construct validity of the SRO.

Impulsivity—Juveniles' impulsivity and self-control of aggressive tendencies were assessed using the 8-item Impulse Control scale (e.g., "I say the first thing that comes into my mind without thinking enough about it.") and the 7-item Suppression of Aggression scale (e.g., "People who get me angry better watch out.") from the Weinberger Adjustment Inventory (Weinberger & Schwartz, 1990). Juveniles responded to a 5-point Likert-type scale anchored by *false*, *somewhat false*, *not sure*, *somewhat true*, and *true*.

Moral disengagement—Juveniles' attitudes toward the treatment of others were assessed using a modified version of the Mechanisms of Moral Disengagement measure (Bandura, Barbarnelli, Caprara, & Pastorelli, 1996). This 32-item measure has items bearing on moral justification, euphemistic language, advantageous comparison, displacement of responsibility, diffusion of responsibility, distorting consequences, attribution of blame, and dehumanization (e.g., "If kids fight and misbehave in school it is the teacher's fault," "Some people deserve to be treated like animals."). Juveniles responded to a 3-point Likert-type scale anchored by *disagree*, *neither agree nor disagree*, and *agree*. Mean item total scores are used as an overall scale measure of moral disengagement.

Certainty of punishment and personal rewards of crime—Juveniles' perceptions of their likelihood of being punished and their personal rewards of crime were assessed using a

7-item Certainty of Punishment Scale (adapted from Nagin & Paternoster, 1994) (e.g., "How likely is it that you would be caught and arrested for the following crimes?" "Vandalism," "Robbery with gun," etc.) and a 7-item Personal Rewards of Crime Scale (adapted from Nagin & Paternoster, 1994) (e.g., "How much 'thrill' or 'rush' is it to do any of the following things?" "Vandalism," "Robbery with gun," etc.). Juveniles responded to an 11-point Likert-type scale ranging from 0 (*no chance of being caught/no fun or kick at all*) to 10 (*absolutely certain to be caught/a great deal of fun or kick*). Mean total scores are derived for each scale measure.

Parental monitoring—A 10-item self-report scale was used to measure parental monitoring. Sample items include "How much does your mother know about where you go after school or work is over for the day?" Items are scored on a 4-point Likert-type scale anchored by *doesn't know anything/never*, *knows a little/sometimes*, *knows a lot/usually*, and *knows everything/always*. Mean total scores are used as a measure of parental monitoring. This measure was adapted from a measure used in a large-scale study of authoritative parenting (Steinberg, Lamborn, Dornbusch, & Darling, 1992).

Peer delinquency and antisocial influence—The antisocial behavior of peers and the attempted influence of these peers were assessed using a 10-item Peer Delinquent Behavior Scale (adapted from Menard & Elliott, 1996) (e.g., "How many of your friends have sold drugs?") and a 7-item Antisocial Influence Scale (adapted from Menard & Elliott, 1996) (e.g., "How many of your friends have suggested that you should sell drugs?"). Juveniles responded to a 5-point Likert-type scale anchored by *none of them*, *very few of them*, *some of them*, *most of them*, and *all of them*. Mean total scale scores were used as a measure of peer delinquent behavior and antisocial influence.

Social support—Juveniles' access to a caring adult was assessed using the 8-item Domains of Social Support scale (e.g., "If you have trouble at home, is there someone you could talk to?") from the Contact with Caring Adults measure (CCAD; Institute of Behavioral Science, 1990). The CCAD is a juvenile-report measure consisting of a series of eight sets of six questions concerning the availability of supportive adults in juveniles' lives. The items were originally drawn from the Rutgers Teenage Risk and Prevention Questionnaire (Institute of Behavioral Science, 1990). A Social Support scale score was computed by summing the number of positive responses to the eight Domains of Social Support scale questions.

Data Analytic Procedures

Item equivalence was examined using confirmatory factor analysis to test the hypothesized relations of the observed measure items to their theoretically proposed latent constructs (Kline, 1998). Five fit indices were used to examine the fit of the proposed measurement models: the chi-square (χ^2), Goodness-of-Fit Index (GFI), (Bentler's) Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Because chi-square tests are greatly influenced by sample size, the CFI and GFI indices, which are not biased by sample size, were chosen as model fit criteria, along with the RMSEA and the SRMR. CFI and GFI values that approach 1.0 (i.e., .90 to 1.0) indicate that a measurement model is a good fit to the observed data, and values between .80 and .90 indicate that a model shows an adequate fit to the observed data. The RMSEA and SRMR provide an unbiased estimate of residual covariance between the estimated population covariance and the sample covariance matrices (Browne & Cudeck, 1993). Browne and Cudeck (1993) recommended that RMSEA values of .05 or less indicate that a model shows a close fit to the observed data in relation to the model's degrees of freedom, an RMSEA value in the range of .06 to .08 would indicate a reasonable fit, and values above .10 indicate an inadequate fit.

To compare item equivalence across groups, the factor loadings of items were constrained to be equal across groups, and the relative difference between a constrained and unconstrained model was assessed using chi-square difference criteria. A significant chi-square difference indicates that the fit of the constrained model is significantly different than the fit of an unconstrained model and that item invariance does not exist.

Functional and scalar equivalence was examined using several structural equation models (SEM) testing the moderating effect of gender and ethnicity/race on the slopes and intercepts of the construct validity relationships between groups. Four scale indices of offending history, eight theoretically relevant psychosocial predictors, and juveniles' history of gang membership were used as predictors of juveniles' SRO scores to assess construct validity of the SRO. For these analyses, the homogeneity of slopes and intercepts (again using chi-square difference criteria) across groups was examined by comparing unconstrained models with constrained slope models, and constrained slope models with models that had slopes and intercepts constrained. Comparable slope relationships between theoretically relevant predictors and the criterion variable across groups provided evidence of the measures' functional equivalence across groups. Comparable intercepts and slopes across gender and ethnic/racial groups support scalar equivalence across groups.

Management of missing SRO data—Given that each item of an observed measure is used as a dependent variable within a confirmatory factor analysis, the management of missing item-level data can affect the accuracy of statistical conclusions drawn from the analysis. Kim and Curry (1977) suggested that sample size, proportion of missing data, and the reasons data are missing contribute to the degree to which missing data impedes statistical accuracy. Although the available sample size for the current study was large (N = 1,338), specified multiple group analyses by ethnicity substantially reduced sample sizes for these analyses. The proportion of missing data across SRO items also posed a potential threat to the validity of results. Although the rate of missing responses for most SRO items was extremely low (between 0.1% and 1%), two SRO items ("ever went joyriding?" and "ever broke into a car to steal something from it?") were missing for more than one third of the sample at baseline (38% and 39%, respectively) because these items were added to the baseline interview several months into data collection. The pattern of missing responses on these items was independent of salient predictors used in the item analyses (i.e., race and gender), and thus could be described as "observed at random" (Rubin, 1976). However, inclusion of these items in a confirmatory factor analysis using common methods of managing missing data (i.e., pairwise or listwise deletion) would have substantially reduced the sample sizes for analyses by groups.

To assess the measurement equivalence of the 24-item SRO and preserve power and statistical conclusion validity, maximum likelihood estimation procedures in AMOS were used to manage missing data for all SRO item-level analyses (Arbuckle, 1999). Arbuckle (1996) demonstrated the efficiency and accuracy of using maximum likelihood estimation procedures to estimate models with missing data and found that this procedure, relative to pairwise deletion, provided an improvement in accuracy equivalent to increasing sample size by up to a factor of 2.75.

Results

Item Equivalence

Cross-gender comparisons—Table 1 presents a summary of the test of latent structures across genders for all measures that included more than one item. Because the SRO comprises dichotomous items, all confirmatory factor analyses of dichotomous SRO items were first analyzed using asymptotic distribution free (ADF) method and weighted least squares estimation in LISREL. However, because the ADF procedures require very large sample sizes,

these analyses had very low power for detecting between-group differences. Thus, SRO measurement equivalence was assessed using maximum likelihood estimation procedures.

Initial cross-gender confirmatory factor analysis using maximum likelihood estimation and the summed variety score including the 24 items of the SRO would not converge. Inspection of item base rates across gender groups showed that the prevalence of self-reported sexual assault ("Have you ever forced someone to have sex with you?") by girls was null, so this item was removed from the item analysis, and the resulting 23 item factor analysis did successfully converge.

As indicated by available fit indices of the unconstrained model (CFI = .88; RMSEA = .04), the fit of the full SRO was adequate across gender groups. Chi-square difference between an unconstrained model and a model with factor loadings constrained showed that there was a significant chi-square difference between boys and girls for the 21-item measure, $\chi^2(df = 22) = 55.72$, $p < .001$. Given that this difference is rather small, it is likely that this statistical significance is primarily the result of the relatively large sample size ($N = 1,338$).

Inspection of cross-gender factor-loading differences showed that two items, "ever was paid to have sex with someone" and "ever was in a fight," did not significantly relate to the SRO construct for girls. The largest factor-loading estimate difference across genders was for a particular aggressive crime—fire setting (girls $b = 0.26$, *ns*; boys $b = 0.74$, $p < .001$). The SRO fire-setting slope relationship may have been attenuated across gender groups by disparity in the prevalence of self-reported fire setting between boys and girls. The prevalence of self-reported fire setting among boys (18%) was double the rate reported by girls (9%). The item "ever took something by force without a weapon" also showed a high factor-loading difference across genders. Contrary to the pattern of boy/girl self-reports, however, robbery without a weapon showed a higher factor loading for girls ($b = 1.61$, $p < .001$), than for boys ($b = 1.24$, $p < .001$) in the SRO model. This difference suggested that girls who robbed others without weapons were more likely to commit other illegal acts. Despite these item differences and the significant difference chi-square, the practical fit indices across the constrained and unconstrained factor-loadings models were nearly identical. Thus, the general pattern of fit of the SRO across gender groups supported considerable item equivalence.

Most of the measures of theoretically related constructs also fit well across gender groups (see Table 1). One half of the constructs (impulsivity, peer delinquent behavior, social support, parental monitoring) showed a good fit to the observed data according to all fit indices, and the other four showed a good fit by at least three fit criteria (certainty of punishment, antisocial influence, personal rewards of crime, and moral disengagement). Only two of eight latent constructs (antisocial influence and certainty of punishment) showed significant chi-square fit differences between constrained and unconstrained factor models. However, these chi-square differences were small in relation to the sample size, and visual comparison of the constrained and unconstrained model fit indices suggested that there were negligible differences between the factor loadings across gender groups.

Hispanic vs. White comparisons—Results of confirmatory factor analysis showed that the SRO was a good fit to the observed data across White and Hispanic male groups (see Table 2). However, analysis of chi-square differences between unconstrained and constrained factor models indicated that the fit of the SRO differed across ethnic groups, $\chi^2(df = 23) = 61.59$, $p < .001$.

Neither of the sex-related crimes, "ever forced someone to have sex with you?" nor "ever been paid to have sex with someone?" contributed significantly to the SRO construct for either White boys or Hispanic boys. Inspection of factor-loading differences across Hispanic and European

White groups suggested that factor loadings on the SRO differed across groups for several items. Most notably, the factor loading of receiving/selling stolen property was higher for White boys ($b = 1.69$, $p < .001$) than for Hispanic boys ($b = 1.01$, $p < .001$). "Ever sell marijuana" also showed a much higher factor loading for White boys than Hispanic boys ($b = 1.88$, $p < .001$, vs. $b = 1.12$, $p < .001$) as did "ever drive while intoxicated?" (White boys $b = 1.97$, $p < .001$, vs. Hispanic boys $b = 1.30$, $p < .001$). These factor-loading differences did not reflect substantial differences in the cross-ethnic prevalence of crimes. Results suggest that marijuana dealing, buying/receiving stolen property, and driving while intoxicated are more common crimes for White boys who commit a large variety of crimes.

Initial confirmatory factor analysis of the constructs other than SRO across White and Hispanic male groups showed good fit to the corresponding observed data. With the exception of personal rewards of crime (which demonstrated an adequate fit), all the models for these theoretically related constructs showed a good fit to the data according to at least two fit indices. Comparison of unconstrained and constrained factor models supported measurement invariance across Hispanic and White male groups. Only two of eight constructs (peer delinquent behavior and antisocial influence) showed significant, but small, item-loading differences across Hispanic and White male groups as indexed by the difference chi-square criteria. Despite these significant chi-square differences, the peer delinquent behavior and antisocial influence measures demonstrated a similar fit to the observed data in the constrained and unconstrained models.

African American vs. White comparisons—Initial confirmatory factor analysis of the SRO across African American and White male groups showed a good fit to the observed data by practical fit indices (see Table 3). The comparison of constrained and unconstrained models for the SRO demonstrated a small, but significant, chi-square difference in the factor loadings across African American and White male groups, $\chi^2(df = 23) = 95.96$, $p < .001$, suggesting some item-level differences between groups. Inspection of cross-ethnic factor-loading estimates showed that, again, sex-related crimes did not significantly contribute to the SRO construct for either ethnic group. The pattern of item-level factor loadings showed 10 crimes with marked between-group differences. Most notably, "ever broke into a building to steal something from it?" showed a much higher factor loading for the White male group ($b = 1.48$, $p < .001$) versus African American ($b = 0.91$, $p < .001$). Fire setting also showed a higher factor loading to the SRO for White boys than for African American boys ($b = 1.104$, $p < .001$, vs. $b = .54$, $p < .001$, respectively). This pattern of factor-loading differences was more than likely related to corresponding between-group differences in the frequency of self-reports of fire setting and weapon-related crimes. The rate of self-reported burglary by White boys (54%) was triple the rate reported by African American boys (18%). Self-reported fire setting was also higher among White boys (22%) than African American boys (14%), while African American boys were more likely to report shooting someone (15%), or shooting at someone (27%), than White boys (10% and 18%, respectively).

Results of confirmatory factor analysis showed that the measurement models of the constructs that were theoretically linked to criminal activities fit reasonably well across White and African American male groups. The Impulsivity, Certainty of Punishment, and Personal Rewards of Crime scales showed small, significant chi-square differences in factor loadings across African American and White male groups (see Table 3), however the pattern-of-fit indices supported comparable fit across ethnic groups. The Peer Delinquent Behavior scale showed a moderate chi-square difference, $\chi^2(df = 6) = 65.73$, $p < .05$, suggesting item factor differences across African American and White male groups. Overall, however, the pattern of unconstrained model fit across all of these constructs supported item equivalence for most of these measures across African American and White male groups.

Functional Equivalence

For the next set of analyses, a series of construct validity relations between theoretically relevant predictors and the SRO criterion were tested using structural equation models. All the hypothesized construct validity relations, with the exception of the model predicting SRO scores from social support, proved to be significant and in the anticipated direction for all gender and ethnic/racial groups (see Tables 4 and 5). To examine functional equivalence of selected measures across gender and ethnic groups, the moderating effects of gender and ethnicity on the slopes of the construct validity relations were assessed.

Cross-gender comparisons—Chi-square differences between unconstrained and constrained slope models demonstrated that most of the hypothesized construct validity relations were homogeneous across gender groups (see Table 4). Only 4 of 13 predictive models (certainty of punishment, impulsivity, personal rewards of crime, and peer delinquent behavior) indicated significantly different slopes between gender groups, and the corresponding chi-square differences were very small. Despite these significant chi-square differences in slopes across gender groups, the predictive relations between these psychosocial variables and the SRO were significant and in the appropriate direction, which supported the functional equivalence of the models across gender groups.

Hispanic vs. White comparisons—Results of the tests of slope differences across Hispanic and White groups revealed only one significant slope difference between these two groups (see Table 5). Although significant, this chi-square difference was very small and probably statistically significant primarily because of the large sample size. Alternatively, this difference may be related to differences in item equivalence across ethnic groups (see Table 2). In any event, this between-group difference did not have substantive implications for the cross-ethnic comparability of the predictive relationship between peer delinquent behavior and SRO. Although the slope of the predictive relationship between peer delinquent behavior and SRO was higher in magnitude for White boys ($b = 4.15$) than for Hispanic boys ($b = 2.99$), the slope was positive and significant for Hispanic boys and White boys. Overall, the predominant homogeneity of slope relations across White and Hispanic male groups supported the functional equivalence of the SRO across Hispanic and White male groups.

African American vs. White comparisons—Four of 13 predictive models revealed significant chi-square slope differences across African American and White male groups (see Table 5). The slope of the prediction of SRO from peer delinquent behavior showed the largest between-group chi-square difference, $\chi^2(df = 1) = 11.72$, $p < .001$, between African American and White boys. This chi-square difference was reflected in a larger slope for White boys than African American boys ($b = 4.15$ vs. $b = 2.91$); however, again, the slope relations for both groups were significant and in the expected direction. And again, this slope difference may have been related to between-group item differences (see Table 3). Similarly, chi-square difference criteria showed that the slopes of the prediction models for number of previous secure placements, impulsivity, and antisocial influence were significantly different across African American and White male groups; however, the corresponding relationships were significant and in the expected direction across these groups.

Scalar Equivalence

To examine scalar equivalence, construct validity relations in unconstrained models were compared to corresponding models with intercepts constrained to be equal across gender and ethnic groups. In essence, if the slope and intercept of the relation between each theoretically linked predictor and the SRO scores are the same across groups, then any given score on any predictor is associated with the same SRO score in each group, and the predictor and the SRO measures must be scalar equivalent.

Cross-gender comparisons—In addition to the chi-square difference tests for the slope of the relationships, Table 4 also presents the difference chi-square comparing the intercepts of the construct validity relations across gender groups. These tests indicate that the intercepts of most of the construct validity measurement models (12 of 13) were significantly different across gender groups. Most chi-square differences in intercepts across gender groups were moderate and reflected substantive differences in intercept estimates. Generally, results showed a pattern of lower intercepts for girls than for boys.

Hispanic vs. White comparisons—Three of the 13 construct validity models revealed significant intercept differences across Hispanic and White boys (see Table 5). However, all of the intercept differences across the Hispanic and White groups were negligible in magnitude, as reflected by the relatively small (even though significant) difference chi-squares and intercept differences.

African American vs. White comparisons—Seven of the 13 construct validity models revealed small but significant intercept differences across African American and White male groups (see Table 5). Furthermore, several of these intercept differences were quite substantial.

Discussion

The overall pattern of these results indicates that the SRO measure produces a reasonably good indicator of illegal activities. The confirmatory factor analyses indicate that the items on the SRO appear to be assessing the same underlying construct. The construct validity coefficients indicate that the scores on the SRO are related significantly to all of the theoretically linked constructs examined in this set of analyses. The findings are consistent with the nomological net of expected relationships for a measure of illegal activities.

Cross-Gender Equivalence

The findings generally suggest that the SRO measure has considerable measurement equivalence across genders. Although the confirmatory factor analyses indicate some small violations of measurement invariance (consistent with the differential item functioning observed by Piquero et al., 2002), these noncomparable factor loadings are on items that could reasonably be expected to be gender specific. Fire setting represents a specific type of illegal activity that male and female juveniles in this age group are not equally likely to exhibit. It is not surprising that the differential rate of endorsement of this gender-specific item would lead to quite different factor loadings for boys and girls and that these small differences would then lead to a statistical rejection of invariance in the test of measurement invariance. Hence, our conclusion is that these small differences are reasonably expectable and not a serious indication of nonequivalence of the SRO measure across genders. Furthermore, this SRO measure appears to have considerable item-level equivalence for those items that are not gender specific.

The construct validity analyses also provide considerable evidence of the functional equivalence of this SRO measure. Only 4 of the 13 analyses relating the SRO to the theoretically linked constructs indicated that the slope of the relationship between that other construct (i.e., certainty of punishment, personal rewards of crime, and peer delinquent behavior) and SRO was significantly different for girls and boys. Given the relatively substantial sample sizes used in these analyses, associated statistical power, as well as the relatively small magnitude of most of these differences, it is noteworthy that so few significant differences emerged. For example, although the largest difference chi-square in the slope comparisons was for the relation between SRO and impulsivity, the difference between the correlation between these two variables for boys and girls was very small and probably not very meaningful (for boys the $r = .45$, for girls the $r = .41$). We conclude that these few differences are more likely a function of either the

degree of equivalence of the theoretically related measures (impulsivity, certainty of punishment, personal rewards of crime, and peer delinquent behavior) and/or chance (i.e., Type I errors). In short, these patterns indicate to us that this SRO measure appears to be assessing the same underlying construct among boys and girls.

It is important to note that the small failures of measurement invariance do not appear to substantially affect the slopes in most of our construct validity analyses. Unfortunately, the state of the art for measurement invariance analyses is such that we do not know what constitutes a large enough failure of measurement invariance to undermine observed construct validity relations. As a result, measurement equivalence must be interpreted in light of the construct being measured in the specific groups being studied. In the current case, it was reasonable to expect that a small portion of the SRO item set may represent gender-specific activities and that the factor structure might be slightly different for boys and girls as a result. A finding of partial measurement invariance (i.e., similar factor loading on all items except for small differences on gender-specific illegal activities) is exactly what one would expect in an effort to develop a measure that is equivalent across genders. Documentation of this array of analytical findings certainly provides a richer picture of measurement equivalence than settling simply for analyses of measurement invariance.

One reaction to our measurement invariance findings might be to drop those items that appear to assess gender-specific illegal activities from the SRO measure to produce a state of measurement invariance. However, this can have the undesirable effect of "engineering" any reported gender difference; that is, if the gender-specific items tend to include activities that are more common for one gender than the other, dropping those items could produce a gender difference in SRO scores that is not truly representative of any real differences that may exist. In addition, dropping these items is selectively influencing the variance of the SRO scores for one gender in a way that may well bias the relations of the SRO scores to other constructs differentially for girls and boys

Unfortunately, the significant gender difference for the intercepts describing the relationship of the theoretically linked constructs to the SRO scores makes it impossible for us to conclude that the SRO measure has scalar equivalence across genders; that is, these findings suggest that the exact form of the relationship between the theoretically linked constructs and SRO are not identical for boys and girls. For example, the significant gender difference in the intercept for the relation between prior petitions and SRO indicates that although the increase in the number of offenses reported in the SRO is associated with increases in the number of petitions filed in the same way for boys and girls, girls with no petitions filed would self-report fewer offenses than would boys with no petitions filed.[2]

There are a number of ways in which this specific intercept difference could occur even if the SRO variety score measure were equivalent across genders. For example, if the SRO variety score was equivalent across genders but there were a judicial system difference in the likelihood of a petition being filed for a boy and girl who commit the same illegal act, the slope of the relation between the SRO variety score and petitions would be the same across genders but the intercept for this relation would be lower for the gender for whom there is a lower likelihood of a petition being filed. Similarly, the significant gender difference in the intercept for the relation between parental monitoring and SRO indicates that although a comparable decrease in self-reported offenses is associated with a given increase in parental monitoring for boys and girls, girls with no monitoring would have fewer self-reported offenses than would boys

[2]Although participants in the current study could not have had zero petitions filed against them because participation required being adjudicated/convicted of a serious felony, this intercept difference associated with a common slope indicates a consistent mean difference between boys and girls in the number of self-reported offenses across the range of petitions filed.

with no monitoring. One possibility is that this intercept difference is a function of parents differing in the degree of monitoring they engage in for their female and male juveniles rather than a failure of measurement equivalence of the SRO variety score.

The bottom line is that this SRO measure appears to be assessing the criminal activities of the male and female study participants relatively consistently. However, we cannot definitively conclude that a given SRO score indicates the same degree of criminal activities among these girls and boys. The SRO measure may be scalar equivalent for boys and girls; however, we just cannot conclude this from the current findings.

The implications of these conclusions are pretty straightforward. First, SRO may be a useful tool for indexing criminal activity over time for girls and boys. Second, researchers should be extremely careful in making inferences about gender differences in criminal activities based on this SRO measure, and it might be most appropriate to analyze these groups separately and compare the results at a more global level. Any direct cross-gender comparisons of SRO scores should only be done with the full awareness that any observed difference or similarity may be the result of measurement artifact. Third, researchers should be very careful when making inferences about the relations between SRO scores and scores on other theoretically linked constructs when they have a sample that includes male and female participants. In this latter type of research, it is essential that researchers confirm that any relationships to SRO identified in a gender diverse sample are similar in each gender subsample as well.

Cross-Ethnic/Race Equivalence Among Boys

The findings clearly suggest that the SRO measure has considerable cross-ethnic/race measurement equivalence. Although chi-square tests suggested group differences in the factor loadings on a few items for Hispanics and Whites, these differences were very small, and our results generally support a conclusion of considerable measurement invariance across these groups.

Cross-ethnic measurement invariance was somewhat less affirmed among African American and White male groups. Although a substantial number of small item-level differences in factor loadings occurred across these ethnic groups, the pattern of item variance was understandable in light of social contexts. The item loadings for weapon-related crimes across groups were consistent with different prevalence rates in our data, with African American boys reporting a higher level of weapon-related crimes than White boys. Furthermore, this prevalence disparity was likely related to differences in the relative availability and use of weapons in Philadelphia (where most African American study participants reside), versus Phoenix, Arizona (where most of the White study participants reside). Similarly, item-loading differences reflected prevalence differences in self-reported fire setting, with White boys reporting fire setting more frequently than African American boys. This discrepancy may be related to the use of fire setting in the service of burglary, which is a more common crime (as observed in our data) in Phoenix than in Philadelphia. Given understandable influences of context on item variance, we read our results as indicating that the SRO factor structure is reasonably invariant for African American and White boys.

The construct validity analyses also provide some evidence of the equivalence of the SRO measure in these three ethnic/racial groups. Only 1 (i.e., peer delinquent behavior) of the 13 tests of equality of slopes for the relations of the SRO to the theoretically linked constructs indicated that the slope of the relationship was significantly different for Hispanics and Whites. Given that these analyses have a sample size that produces adequate statistical power for such tests and that this relative frequency of significant effects is only slightly above what one would expect by chance alone (i.e., the Type I error rate), it is appropriate to conclude that this SRO measure is assessing the same underlying construct among Hispanics and Whites. Furthermore,

that only 3 of the 13 intercept differences between the Hispanics and Whites is statistically significant (i.e., for age at first arrest, impulsivity, and social support) suggests that the SRO measure displays considerable scalar equivalence. In short, any given score on the SRO measure indicates the same, or a similar, degree of illegal activity regardless of whether the respondent is Hispanic or White.

Although the findings are not as compelling, the relatively few significant slope differences (4 of 13: number of secure placements, impulsivity, peer delinquent behavior, and antisocial influence) among African Americans and Whites makes a beginning case for equivalence of the SRO measure with these groups. At the same time, this conclusion has to be dampened somewhat by the considerable number of significant intercept differences (7 of 13: age at first arrest, prior petitions, impulsivity, certainty of punishment, peer delinquent behavior, antisocial influence, and parental monitoring) between the African Americans and Whites. Overall, then, these data are only somewhat suggestive of measurement equivalence of the SRO measure among African Americans and Whites.

It is difficult to determine exactly what these differences might mean. It is quite possible that some of these slope and intercept differences are a function of real-world processes that differentially affect African American and White juvenile offenders. For example, the significant slope difference for the number of secure placements related to the number of illegal activities endorsed could be the result of a bias in the likelihood of an African American juvenile being adjudicated and placed in a secure facility, despite having the same level of illegal activities as a White offender (see the analyses of differential court processing in these locales presented in Mulvey et al., 2004 [in this issue]). Similarly, the differences in the intercepts for parental monitoring may indicate neighborhood parenting practices tied to ethnicity. Given the observed significant relationships among constructs and African American/White differences in the relationships among constructs, it is probably reasonable to tentatively conclude that the SRO measure is assessing illegal activities among the African American and White male participants. It also seems reasonable, however, to reserve judgment about whether any given SRO score indicates the same degree of illegal activity among the African American and White participants. These results do not say definitively that it does or does not.

The bottom line is that the SRO measure appears to be assessing criminal activities of the Hispanic and White male research participants, and this measure appears to have some scalar equivalence across these two groups. The implications of this conclusion are that (a) comparisons of these two groups on the SRO measure are likely meaningful, (b) comparisons across these two groups of the relation between the SRO scores and scores on other theoretically linked constructs that are also scalar equivalent are likely meaningful, and (c) the SRO measure can probably be used appropriately in diverse samples that include Hispanic and White boys. With regard to the African American–White male comparisons, the bottom line is probably more similar to that for the cross-gender conclusions; that is, the SRO measure appears to be useful in indexing criminal activities among African American and White boys; however, researchers should be careful when making inferences regarding differences between these two groups in SRO scores or relations of SRO scores to other constructs.

As mentioned at the outset, SRO scales are used widely for characterizing involvement in antisocial and criminal activities, even though there is disturbingly little information on the equivalence of this type of measure in different groups of juvenile offenders. Fortunately, the current study provides some reassurance for those who have nagging fears regarding the meaningfulness of the patterns of offending across diverse samples of offenders observed in the available literature. In general, the SRO measure examined here performed reasonably well in terms of measurement equivalence across gender and race/ethnicity. Furthermore, this evidence supports continued use of the SRO measure in heterogeneous groups of offenders,

provided adequate caution is taken in interpreting particular observed group differences, and effects observed in heterogeneous groups are confirmed within specific gender/ethnic/racial groups.

Acknowledgments

This project was funded jointly by the Office of Juvenile Justice and Delinquency Prevention (grant #2000-MU-MU-0007), the National Institute of Justice (grant #1999-IJ-CX-0053), the John D. and Catherine T. MacArthur Foundation, the William T. Grant Foundation, the Robert Wood Johnson Foundation, the Centers for Disease Control, the William Penn Foundation, the Arizona Governor's Justice Commission, and the Pennsylvania Commission on Crime and Delinquency.

References

Arbuckle, JL. Full information estimation in the presence of incomplete data. In: Marcoulides, GA.; Schumacker, RE., editors. Advanced structural equation modeling techniques. Hillsdale, NJ: Lawrence Erlbaum; 1996. p. 243-277.

Arbuckle, JL. AMOS 4.0 user's guide. Chicago: SPSS; 1999.

Bandura A, Barbarnelli C, Caprara GV, Pastorelli C. Mechanisms of moral disengagement in the exercise of moral agency. Journal of Personality and Social Psychology 1996;71:364–374.

Brame R, Fagan J, Piquero AR, Schubert CA, Steinberg L. Criminal careers of serious delinquents in two cities. Youth Violence and Juvenile Justice 2004;2(3):256–272. [PubMed: 20119514]

Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-162.

Bushway SD, Piquero AR, Broidy LM, Cauffman E, Mazerolle P. An empirical framework for studying desistance as a process. Criminology 2001;39:491–515.

Cohen, J. Research on criminal careers: Individual frequency rates and offense seriousness. In: Blumstein, A.; Cohen, J.; Roth, J.; Visher, CA., editors. Criminal careers and "career criminals". Vol. 1. Washington, DC: National Academy Press; 1986. p. 292-418.

Elliott DS, Ageton SS. Reconciling race and class differences in self-reported and official estimates of delinquency. American Sociological Review 1980;45:95–110.

Hindelang, MJ.; Hirschi, T.; Weis, JG. Measuring delinquency. Beverly Hills, CA: Sage; 1981.

Hines AM. Linking qualitative and quantitative methods in cross-cultural survey research: Techniques from cognitive science. American Journal of Community Psychology 1993;21:729–746.

Hughes D, Seidman E, Williams N. Cultural phenomena and the research enterprise: Toward a culturally anchored methodology. American Journal of Community Psychology 1993;21:687–704. [PubMed: 8085565]

Hui CH, Triandis HC. Measurement in cross-cultural psychology: A review and comparison of strategies. Journal of Cross-Cultural Psychology 1985;16:131–152.

Huizinga D, Elliott DS. Reassessing the reliability and validity of a delinquency scale. Journal of Quantitative Criminology 1986;2:293–327.

Huizinga D, Esbensen F, Weiher A. Are there multiple pathways to delinquency? Denver youth survey. Journal of Criminal Law and Criminology 1991;82:83–118.

Institute of Behavioral Science. Youth Interview Schedule: Denver Youth Survey. Boulder: University of Colorado; 1990.

Kim J, Curry J. The treatment of missing data in multivariate analysis. Sociological Methods and Research 1977;77:215–241.

Kline, RB. Principles and practice of structural equation modeling. New York: Guilford; 1998.

Knight, GP.; Hill, NE. Measurement equivalence in research involving minority adolescents. In: McLoyd, VC.; Steinberg, L., editors. Studying minority adolescents: Conceptual, methodological, and theoretical issues. Mahwah, NJ: Lawrence Erlbaum; 1998. p. 183-210.

Knight, GP.; Tein, J-Y.; Prost, J.; Gonzales, NA. Measurement equivalence and research on Latino children and families: The importance of culturally informed theory. In: Contreras, JM.; Kerns, KA.;

Neal-Barnett, AM., editors. Latino children and families in the United States: Current research and future directions. Westport, CT: Praeger; 2002. p. 181-201.

Malpass, RS.; Poortinga, YH. Strategies for design and analysis. In: Lonner, WJ.; Berry, JW., editors. Field methods in cross-cultural research. Newbury Park, CA: Sage; 1986. p. 47-84.

Marin G, Gamba RJ, Marin BV. Extreme response style and acquiescence among Hispanics: The role of acculturation and education. Journal of Cross-Cultural Psychology 1992;23:498–509.

Menard, S.; Elliott, DS. A report prepared for the MacArthur Foundation by the MacArthur Chicago-Denver Neighborhood Project. 1996. Prediction of adult success using stepwise logistic regression analysis.

Mulvey EP, Steinberg L, Fagan J, Cauffman E, Piquero AR, Chassin L, et al. Theory and research on desistance from antisocial activity among serious adolescent offenders. Youth Violence and Juvenile Justice 2004;2(3):213–236. [PubMed: 20119505]

Nagin DS, Paternoster R. Personal capital and social control: The deterrence implications of a theory of individual differences in criminal offending. Criminology 1994;32:581–606.

Rubin DB. Inference and missing data. Biometrika 1976;63:467–474.

Piquero AR, MacIntosh R, Hickman M. The validity of a delinquency scale. Sociological Methods and Research 2002;30:492–529.

Schubert CA, Mulvey EP, Steinberg L, Cauffman E, Losoya SH, Hecker T, et al. Operational lessons from the Pathways to Desistance Project. Youth Violence and Juvenile Justice 2004;2(3):237–255. [PubMed: 20119515]

Steenkamp JB, Baumgartner H. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research 1998;25:78–90.

Steinberg L, Lamborn SD, Dornbusch SM, Darling N. Impact of parenting practices on adolescent achievement: Authoritative parenting, school involvement, and encouragement to succeed. Child Development 1992;63:1266–1281. [PubMed: 1446552]

Thornberry, TP.; Krohn, MD. U.S. National Institute of Justice. Measurement and analysis of crime and justice: Criminal justice series. Vol. 4. Washington, DC: National Institute of Justice; 2000. The self-report method of measuring delinquency and crime; p. 33-83.

Weinberger DA, Schwartz GE. Distress and restraint as superordinate dimensions of self-reported adjustment: A typological perspective. Journal of Personality 1990;58(2):381–417. [PubMed: 2213473]

Biographies

George P. Knight is a professor in the Department of Psychology at Arizona State University. He received his Ph.D. in social psychology, specializing in social development, from the University of California at Riverside. His research interests include the acculturation/enculturation of Mexican American children and families, cross-ethnic/race measurement equivalence, and prosocial development.

Michelle Little, M.A., is a 3rd-year graduate student in the Developmental Psychology Program at Temple University. She received her undergraduate degree from Yale University. She is currently working as a research interviewer for the Research on Pathways to Desistance Project. Her doctoral research focuses on the psychosocial adaptation of serious juvenile offenders within secure facilities.

Sandra H. Losoya is a research assistant professor at Arizona State University. She received her Ph.D. in developmental psychology at the University of Oregon. Her interests include emotional development and individual differences in emotional responding, emotional coping, and other sources of resilience in high-risk children.

Edward P. Mulvey is a professor of psychiatry and director of the Law and Psychiatry Program at Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine. His

research has centered on issues related to the use of social service treatment as a method of social control, with an emphasis on the prediction of violence in individuals with mental illness and on treatment systems for juvenile serious offenders.

TABLE 1

Tests of Measurement Invariance Across Genders

	Unconstrained Model			Constrained Factor Loadings			
Measure	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*) Difference**
SRO	1,387.81 (452)	NA/.88	NA/.04	1,443.53 (474)	NA/.87	NA/.04	55.72 (22)***
Moral disengagement	2,292.13 (928)	.90/.83	.04/.03	2,319.33 (959)	.90/.83	.04/.03	27.20 (31)
Impulsivity	599.48 (178)	.94/.91	.04/.04	607.54 (191)	.94/.91	.04/.04	8.07 (13)
Certainty of punishment	356.04 (28)	.93/.94	.04/.09	369.76 (34)	.92/.93	.04/.09	13.72 (6)*
Personal rewards of crime	373.51 (28)	.92/.93	.05/.10	377.49 (34)	.92/.93	.05/.09	3.98 (6)
Peer delinquent behavior	582.14 (90)	.93/.95	.04/.07	594.72 (101)	.92/.95	.04/.06	12.57 (11)
Antisocial influence	474.77 (28)	.90/.91	.05/.11	489.69 (34)	.90/.91	.05/.10	14.92 (6)*
Social support	173.33 (40)	.97/.95	.03/.05	179.169 (47)	.97/.95	.03/.05	5.84 (7)
Parental monitoring	165.94 (40)	.97/.97	.05/.05	172.59 (48)	.97/.97	.05/.05	6.64 (8)

NOTE: GFI/CFI = Goodness-of-Fit Index/(Bentler's) Comparative Fit Index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; SRO = self-reports of offending; NA = not applicable.

* $p < .05$.

*** $p < .001$.

TABLE 2

Tests of Measurement Invariance Across Hispanic and White Adolescent Males

	Unconstrained Model			Constrained Factor Loadings			
Measure	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*) Difference**
SRO	887.65 (496)	NA/.90	NA/.04	949.23 (519)	NA/.89	NA/.04	61.59 (23)***
Moral disengagement	1,713.78 (928)	.85/.80	.07/.04	1,756.75 (959)	.84/.80	.07/.04	42.97 (31)
Impulsivity	403.98 (178)	.92/.90	.06/.05	418.94 (178)	.92/.90	.06/.04	14.96 (13)
Certainty of punishment	219.06 (28)	.91/.92	.06/.11	224.03 (34)	.90/.92	.06/.09	4.973 (6)
Personal rewards of crime	346.20 (28)	.86/.86	.07/.14	354.52 (34)	.86/.86	.07/.12	8.32 (6)
Peer delinquent behavior	457.19 (102)	.88/.93	.05/.08	481.05 (113)	.87/.93	.06/.07	23.86 (11)**
Antisocial influence	258.03 (28)	.90/.91	.06/.12	271.93 (34)	.89/.91	.06/.11	13.90 (6)*
Social support	121.72 (40)	.96/.95	.05/.05	132.41 (47)	.96/.96	.06/.05	6.96 (7)
Parental monitoring	118.20 (40)	.95/.96	.08/.06	125.83 (48)	.95/.96	.08/.06	7.63 (8)

NOTE: GFI/CFI = Goodness-of-Fit Index/(Bentler's) Comparative Fit Index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; SRO = self-reports of offending; NA = not applicable.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 3

Tests of Measurement Invariance Across African American and White Adolescent Males

	Unconstrained Model			Constrained Factor Loadings			
Measure	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*)**	**GFI/CFI**	**SRMR/RMSEA**	**χ^2 (*df*) Difference**
SRO	975.54 (496)	NA/.87	NA/.04	1,071.50 (519)***	NA/.85	NA/.04	95.96 (23)***
Moral disengagement	1,683.00 (928)	.87/.81	.07/.03	1,720.70 (959)	.87/.81	.07/.03	37.71 (31)
Impulsivity	372.26 (178)	.93/.91	.06/.04	395.12 (191)	.93/.91	.07/.04	22.89 (13)*
Certainty of punishment	179.63 (28)	.93/.94	.06/.09	198.71 (34)	.92/.94	.08/.08	19.08 (6)**
Personal rewards of crime	220.73 (28)	.92/.92	.07/.10	234.43 (34)	.92/.91	.07/.09	13.70 (6)*
Peer delinquent behavior	521.73 (100)	.87/.91	.05/.08	587.46 (111)	.86/.90	.08/.08	65.73 (11)***
Antisocial influence	242.63 (28)	.91/.90	.06/.10	280.95 (34)	.90/.89	.07/.10	38.32 (6)
Social support	121.72 (40)	.96/.96	.05/.05	130.82 (47)	.96/.95	.06/.05	6.96 (7)
Parental monitoring	109.86 (40)	.96/.96	.08/.05	117.83 (48)	.96/.96	.09/.05	7.97 (8)

NOTE: GFI/CFI = Goodness-of-Fit Index/(Bentler's) Comparative Fit Index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; SRO = self-reports of offending; NA = not applicable.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 4

Cross-Gender Comparisons of Slopes and Intercepts for SRO Construct Validity Relations

Predicted Measure	Girls Slope/Intercept	Boys Slope/Intercept	χ^2 Slope Difference	χ^2 Intercept Difference
Arrests	$0.50^{***}/3.84^{***}$	$0.45^{***}/5.97^{***}$	0.58	38.33^{***}
Age first arrest	$-0.77^{***}/16.70^{***}$	$-0.75^{***}/18.40^{***}$	0.02	34.83^{***}
Secure placements	$0.50^{***}/4.91^{***}$	$0.52^{***}/7.14^{***}$	0.04	46.83^{***}
Prior petitions	$0.53^{**}/4.39^{***}$	$0.71^{***}/5.76^{***}$	1.13	27.80^{***}
Lifetime gang membership	$4.93^{***}/6.44^{***}$	$5.68^{***}/6.63^{***}$	0.68	0.85
Moral disengagement	$4.24^{***}/-0.87$	$5.36^{***}/-0.61$	1.44	34.41^{***}
Impulsivity	$2.85^{***}/-5.59^{***}$	$3.70^{***}/-8.07^{***}$	15.35^{***}	57.24^{***}
Certainty of punishment	$-0.38^{***}/7.96^{***}$	$-0.61^{***}/11.21^{***}$	4.09^{*}	28.08^{***}
Personal rewards of crime	$0.58^{***}/4.25^{***}$	$0.86^{***}/5.90^{***}$	4.32^{*}	45.21^{***}
Peer delinquent behavior	$2.31^{***}/0.69$	$3.33^{***}/0.17$	9.27^{**}	29.53^{***}
Antisocial influence	$2.65^{***}/1.17$	$2.91^{***}/2.84^{***}$	0.43	45.03^{***}
Social support	$-0.19/6.81^{***}$	$-0.17/9.06^{***}$	0.02	47.24^{***}
Parental monitoring	$-2.16^{***}/11.66^{***}$	$-2.49^{***}/14.68^{***}$	0.58	40.18^{***}

NOTE: SRO = self-reports of offending.

$^{*}p < .05$.

$^{**}p < .01$.

$^{***}p < .001$.

TABLE 5

Cross-Ethnic/Race Comparisons of Slopes and Intercepts for SRO Construct Validity Relations Among Adolescent Males

				Hispanic vs. White		African American vs. White	
Predicted Measure	**Hispanic Slope/Intercept**	**African American Slope/Intercept**	**White Slope/Intercept**	**χ^2 Slope Difference**	**χ^2 Intercept Difference**	**χ^2 Slope Difference**	**χ^2 Intercept Difference**
Arrests	$0.41^{***}/6.88^{***}$	$0.43^{***}/5.52^{***}$	$0.48^{***}/6.00^{***}$	0.87	2.04	0.39	3.66
Age first arrest	$-.85^{***}/20.65^{***}$	$-.61^{***}/15.66^{***}$	$-.92^{***}/21.00^{***}$	0.08	4.28^{*}	2.71	7.78^{**}
Secure placements	$0.41^{***}/7.96^{***}$	$0.93^{***}/6.08^{***}$	$0.53^{***}/7.18^{***}$	1.34	1.68	9.30^{***}	1.82
Prior petitions	$0.68^{***}/6.53^{***}$	$0.71^{***}/5.08^{***}$	$0.79^{***}/5.72^{***}$	0.47	1.28	0.00	8.30^{**}
Gang membership	$4.80^{***}/6.87^{***}$	$4.41^{***}/6.72^{***}$	$5.12^{***}/7.19^{***}$	0.11	0.87	0.50	2.44
Moral disengagement	$5.57^{***}/-.48$	$4.72^{***}/-.18$	$5.28^{***}/-.37$	0.06	0.83	0.26	3.42
Impulsivity	$3.89^{***}/-8.88^{***}$	$2.54^{***}/-7.34^{***}$	$4.06^{***}/-8.12^{***}$	1.15	10.21^{*}	10.35^{**}	76.78^{***}
Certainty of punishment	$-.57^{***}/11.86^{***}$	$-.61^{***}/10.18^{***}$	$-.83^{***}/12.90^{***}$	0.90	2.64	2.73	15.18^{**}
Personal rewards of crime	$0.97^{***}/6.08^{***}$	$0.68^{***}/6.13^{***}$	$0.84^{***}/5.70^{***}$	0.63	3.43	1.14	0.00
Peer delinquent behavior	$3.30^{***}/0.53$	$2.91^{***}/0.58$	$4.15^{***}/-.97$	5.07^{*}	1.72	11.72^{***}	14.29^{***}
Antisocial influence	$2.99^{***}/2.35^{***}$	$2.45^{***}/3.13^{***}$	$3.52^{***}/1.99^{***}$	1.22	0.27	5.40^{*}	4.00^{*}
Social support	$-.09/9.30^{***}$	$-.10/7.97^{***}$	$-.17/8.99^{***}$	0.09	4.06^{*}	0.08	1.53
Parental monitoring	$-2.62^{***}/15.62^{***}$	$-2.26^{***}/12.86^{***}$	$-2.82^{***}/15.64^{***}$	0.13	1.98	1.22	11.98^{***}

NOTE: SRO = self-reports of offending.

$^{*}p < .05$.

$^{**}p < .01$.

$^{***}p < .001$.