

## Chapter 17

### Constructing Comparable Measures of Delinquency by Gender: A Theoretical and Empirical Inquiry

*Ann Marie Sorenson, David Brownfield and Gary F. Jensen*

Department of Sociology, University of Toronto

Department of Sociology, University of Toronto

Department of Sociology, Vanderbilt University

Several researchers have considered the question of using the same items to measure delinquency in different groups. Smith and Davidson (1986) focus on a latent variable indicating a propensity to violence for males and females. They use three items from the 1981 Monitoring the Future national sample of high school seniors: (1) How many times in the last 12 months have you gotten into a fight at school? (2) How many times in the last 12 months have you gotten into a fight with a group of people against another group? (3) How many times in the past 12 months have you hurt someone badly? They find that the second item, fighting in groups, is a less reliable indicator for white females compared with males and black females. The reliability of the third item, asking respondents about hurting someone badly, is lower for females than for males. While common factor loadings are found for black males and white males with respect to these items on self-reported violence, Smith and Davidson conclude that these measures do not reflect a single comparable latent variable for females. Weis (1976) reports a similar finding regarding measures of violence.

Hindelang, Hirschi, and Weis (1981) focus on a comparison of self-reported delinquency formale and female respondents. In their analysis of the Seattle Youth Study, Hindelang et al. (1981) find that nearly all of the reliability coefficients were quite substantial, ranging from .80 to .90; furthermore, these reliability levels did not vary appreciably between male and female respondents.

Hindelang et al. (1981) report the results of five types of coefficients or „links“ (e.g., among self-reported delinquency, official records, self-reported official contacts, and informant data) which assess the correlational validity of self-reports for males and females. They conclude that the various validity coefficients for males and females do not vary substantially. Based on this evidence, these researchers concluded that self-report data appear to be equally valid for males and females.

Huizinga and Elliott (1986) report that male and female respondents have similar levels of reliability for most scales of self-reported delinquency, including a UCR (Uniform Crime Reports) index scale and theft and assault scales. In terms of validity, they conclude that overreporting of behaviors too trivial to be charged or considered delinquent did not vary significantly by gender. The overall rates of underreporting of offenses officially recorded as arrests were also similar for males and females in the National Youth Survey. None of these tests of reliability and validity, however, provide direct evidence as to whether observed measures of delinquency reflect a single comparable latent variable for males and females.

## 1. Theoretical Issues

If we consider the literature summarized above, we can at least tentatively propose some hypotheses that would have been consistent with findings had they been specified in advance. For example, the items selected by Smith and Davidson (1986) to measure violence as a latent variable differ in content and wording in ways that could affect the correlation of responses by gender. One item asks about fights at school, another asks about group fights, and a third asks about harm done. There are several ways that such content and wording could affect responses. For example, the item asking about fights does not limit fights to physical altercations. It is taken for granted that respondents know that the referent is physical battles, but that assumption may apply more readily to males than females. It is conceivable that for females serious shouting matches and arguments could be included. If the female definition were more inclusive, then their responses would not as readily correlate with items asking about degree of physical harm.

The item asking about degree of harm is not phrased in terms of physical harm either. If males define „hurting someone badly“ in physical terms and females define it in other terms (e.g., emotional harm) or their conceptions of „badly“ differ, then the items may not be measuring the same phenomena.

Finally, since location is specified in one question and not in others we have no way of knowing what relevance that variable has for differences by gender. The items were not designed to test specific hypotheses about violence within a conceptual framework where variations are systematically introduced (i.e., location, severity, group context, etc.). Ideally, the wording should have been the same in each instance with the different aspects of „violence“ specified systematically and concretely.

Even if the items were found to have the same meaning we would not necessarily expect similar correlations within gender categories. Sampson (1985, p. 351) speculates that „offenses may cluster because they refer to a common object (e.g., school, store), reflect environmental restrictions or opportunities, or describe a single event that includes several distinct acts.“ We would add that if such circumstances vary among groups, „clusters“ may vary as well. For example, if male fights are more likely to result in harm than female fights, then we would predict that the number of fights would correlate better with the number of times someone was hurt badly for males than females. This lower correlation could be generated by an accurately and reliably measured difference in the outcome of fights for males and females. In fact, the correlations among items reported by Smith and Davidson (1986, p. 448) line up in the order that researchers would expect if fights varied in severity by gender. The strongest correlations between self-reports of fights and self-reports of harm are found for black males (.584, .589) and for white males (.504, .489), followed by black females (.394, .348) and white females (.285, .183).

As a tentative theoretical hypothesis we can propose that the more diverse and less concrete the wording of items intended to measure dimensions of criminality or delinquency, the greater the probability that they will not appear to measure the same latent variable in diverse groups. Moreover, if the items used are designed to measure a general phenomenon (e.g., fights) as well as some dimension of the general phenomenon that varies among groups (e.g., harm), then the probability that they will appear to measure the same characteristic is reduced.

While there is not sufficient space here to reanalyze these data or develop hypotheses about the specific circumstances generating different conclusions about the existence of latent variables, we can attempt a test of the general hypothesis stated above using items with properties that should overcome the problems discussed. Our intention is to examine a range of observed self-report items (including assault and serious crimes such as auto theft) from the Seattle data to further assess the latent variable of delinquent behavior by gender.

We will use the simultaneous latent class technique described in the introductory chapter. Such models involve fixing probabilities for groups (here, defined by gender), as well as allowing for equality constraints across groups. We estimate both heterogeneous and homogeneous simultaneous latent structure models.

## 2. Hirschi's Scale of Delinquency

We will begin by considering the six items included in Hirschi's (1969) scale of delinquent behavior. Three of those items measure theft and are identical in wording except for variation in the value of the items or money stolen. As a set they measure the full range of possible thefts and do not overlap. Thus, the items are designed, whether intentionally or not, to avoid the problems outlined above and should have a high probability of measuring the latent variable of theft.

The other items used in this popular index of delinquency include single questions about vandalism („Have you ever banged up something that did not belong to you on purpose?“), fights („Not counting fights you may have had with a brother or sister, have you ever beaten up on anyone or hurt anyone on purpose?“), and auto theft („Have you ever taken a car belonging to someone you didn't know for a ride without the owner's permission?“). All six items deal with distinct events that would be criminal offenses if committed by adults. In this paper, we examine data taken from the Seattle Youth Study, the same data set analyzed by Hindelang et al. (1981) and by Sampson (1985). In the Seattle study, more than one thousand ( $N = 1,089$ ) subjects provided complete responses in interviews and questionnaires. Approximately three-fourths of the sample are males (75.5%) and about one-fourth are females (24.5%).

## 3. Findings

In Table 1 we present the goodness-of-fit statistics for a series of simultaneous latent variable models. Each of these models represents a test of a substantive hypothesis, such as whether the observed measures form a similar latent variable by gender, or whether particular observed items such as assault are equally valid indicators of the latent variable for males and females.

First, we consider the model of independence (H1) for the cross-classification of the six observed measures of delinquency by sex. In the context of latent structure analysis, the model of independence stipulates that there is no common latent variable underlying the observed measures for males and for females. If the model of independence provides an adequate fit to the data, then we would fail to reject the null hypothesis that there is no association among the six observed measures for the two groups. As expected, however, the model of independence has a very poor fit.

<i>Model</i>	$L^2$	<i>df</i>	<i>p</i>	<i>BIC</i>
H1: one class	1046.45	122	<.01	193.30
H2: two classes sim.	173.12	117	<.05	-645.06
H3: three classes sim.	75.34	106	>.10	-665.92
H4: three classes hom.	188.74	124	<.05	-678.39
H5: partially hom.	151.71	121	<.05	-694.44
H6: partially hom.	101.23	118	>.10	-723.95
H7: Rasch model	429.36	115	<.01	-374.84

**Table 1:** Goodness of fit statistics for selected simultaneous latent variable models.

The second model (H2) summarized in Table 1 specifies a heterogeneous and unrestricted two class latent variable model for males and females. This model is equivalent to specifying an unrestricted two class model separately for males and females (Clogg and Goodman, 1986). While previous research (Brownfield and Sorenson, 1987) would suggest beginning with a three class latent variable model for these observed measures of delinquency, there are two reasons to begin with a two class model here. First, such a model is more parsimonious than a three class model; the specification of the model and the interpretation of results would be facilitated by fitting a two class model. Second, only males were included in the previous research of Brownfield and Sorenson (1987). It is conceivable that a two class model may adequately describe the patterns of female delinquency involvement, which may be thought to be more limited than the delinquency involvement of males.

Model H2 does provide a substantial improvement in fit over the model of independence. However, this two class latent variable model fails to adequately fit the data. We therefore examined the fit of a three class heterogeneous and unrestricted model (H3).

Model H3 does provide a reasonably good fit to the data. It allows us to conclude, tentatively at least, that latent variables containing three classes could be used to describe the association among the six observed measures of delinquency for both males and females. However, the nature of the latent variables for males and females in terms of class assignment and criteria for class assignment remains different under the heterogeneous and unrestricted simultaneous latent structure Model H3. Therefore, we consider the further possibility that the latent variable underlying the association among the observed items is identical for both males and females by specifying a homogeneous and unrestricted three class model (H4).

Model H4 specifies that the parameters used to assign respondents to the three classes of the latent variable are identical for males and females. This model allows us to assess whether or not we can create a scale or latent variable of „delinquency“ that has a completely comparable meaning for male and female respondents.

Unfortunately, Model H4 does not provide an adequate fit. Thus, we need to consider additional models which are partially homogeneous. These models will relax the constraints of complete equality imposed by Model H4. No within group restrictions were imposed on the models considered in this analysis.

Model H5, the first of the partially homogeneous models summarized in Table 1, drops the equality constraints for the conditional probabilities for the observed item measuring assault. Recall that Smith and Davidson (1986) and Weis (1976) found that measures of self-reported violence do not constitute a comparable latent variable for males and females. Model H5 tests

the hypothesis that the item on self-reported assault is not an equally valid indicator of latent class assignment for males and females. Although this model does not provide an adequate overall fit, it does appear to substantially improve the fit compared with the homogeneous Model H4.

We considered another partially homogeneous model (H6). Model H6 drops the equality constraints for the conditional probabilities for the observed items measuring assault and vandalism. This model retains identical parameters for latent class assignment for the four remaining observed items, all of which are measures of theft. Model H6 does provide an adequate fit to the data.

		Males			Females		
		Latent Class			Latent Class		
		I	II	III	I	II	III
Theft under \$2	No	.49	.03	.03	.49	.03	.03
	Yes	.51	.97	.97	.51	.97	.97
Theft \$2 - \$50	No	.96	.06	.34	.96	.06	.34
	Yes	.04	.94	.66	.04	.94	.66
Theft over \$50	No	.99	.08	.98	.99	.08	.98
	Yes	.00	.92	.02	.00	.92	.02
Vandalism	No	.70	.32	.49	.83	.68	.80
	Yes	.30	.68	.51	.17	.32	.20
Assault	No	.62	.24	.44	.83	.64	.65
	Yes	.38	.76	.56	.17	.36	.35
Auto Theft	No	.97	.55	.95	.97	.55	.95
	Yes	.03	.45	.05	.03	.45	.05

**Table 2:** Response distribution to observed measures by latent class and gender, under a three-class partially homogeneous model (H6)

In Table 2, we present the expected response distribution for Model H6 for each of the observed variables within each class of the latent variable of delinquency for both males and females. The response distribution is identical for the four theft items as specified by the partially homogeneous Model H6. The differences in the response distribution for the items measuring assault and vandalism vary somewhat in magnitude across latent classes. For example, the proportions of positive responses to these two items in the second latent class are substantially higher for males compared with females. It appears that the items measuring vandalism and assault do not help to distinguish latent class membership for females as well as for males. Among females, the proportion of positive responses to the assault item is nearly equivalent in the second and third latent classes.

We applied a dichotomous Rasch model (H7) to the observed measures of delinquency for males and females. In the introductory chapter, Rost and Langeheine present formulas for the dichotomous Rasch model, and they describe some of the main features of this model. This simple model is often difficult to fit adequately since it requires, for example, that all of the item response functions have the same slope. Reiser and Schuessler (1991) conclude that latent class models may be more widely applicable than the Rasch model, since Rasch models rarely fit for more than four observed items. Indeed, we find that the Rasch model fails to provide an adequate fit to these data.

The application of a mixed Rasch model to these data is another possible strategy to consider for future research. Such a model, described in the introductory chapter, would allow for a representation of differences both within and among classes. The ability parameter of the Rasch model could be used to describe differences among subjects within a latent class.

#### 4. Summary and Conclusions

We have found that a latent variable of delinquent behavior can be constructed for both male and female respondents based on a commonly used set of observed measures of delinquency. Consistent with prior research, we also find differences in the nature of this latent variable among males and females. While measures of theft seem to constitute a very comparable set of indicators in the construction of a common latent variable across gender, measures of assault and vandalism are less comparable indicators of the latent variable of delinquency across gender. Although measures of assault and vandalism can help form a latent variable scale for males and females separately, indices which include such measures are not identical for males and females.

One explanation for the lack of comparability of the assault and vandalism items may be ruled out. These acts are common enough among female respondents to the extent that they could potentially form part of a comparable scale to an index for male respondents. Over one-fourth (27.7%) of the females confess to assault and nearly one-fifth (19.9%) confess to an act of vandalism.

Future research involving comparisons of male and female delinquency should attend to the results of simultaneous latent variable analyses. Comparisons of groups require that the criterion variable for comparison have the same meaning across groups. We believe that simultaneous latent class analysis is a useful technique for developing such comparable measures. comparisons

#### References

- Brownfield, David and Ann Marie Sorenson. 1987. A Latent Structure Analysis of Delinquency. *Journal of Quantitative Criminology* 3, 103-124.
- Clogg, Clifford and Leo Goodman. 1986. On Scaling Models Applied to Data From Several Groups. *Psychometrika* 51, 123-135.
- Hindelang, Michael, Travis Hirschi, and Joseph Weis. 1981. *Measuring Delinquency*. Beverly Hills: Sage.
- Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley: University of California Press.
- Huizinga, David and Delbert Elliott. 1986. Reassessing the Reliability and Validity of Self-Report Delinquency Measures. *Journal of Quantitative Criminology* 2, 293-327.
- Reiser, Mark and Karl Schuessler. 1991. A Hierarchy for Some Latent Structure Models. *Sociological Methods and Research* 19, 419-465.
- Sampson, Robert. 1985. Sex Differences in Self-Reported Delinquency and Official Records: A Multiple-Group Structural Modeling Approach. *Journal of Quantitative Criminology* 1. 345-367.
- Smith, Douglas and Laura Davidson. 1986. Interfacing Indicators and Constructs in Criminological Research: A Note on the Comparability of Self-Report Violence Data for Race and Sex Groups. *Criminology* 24, 473-488.

Weis, Joseph. 1976. Liberation and Crime: The Invention of the New Female Criminal. Paper presented at the Annual Meeting of the Pacific Sociological Association.