Establishing the Reliability of Self-reported Delinquency Data

Paul K. Broder
Senior Staff Associate
National Center for State Courts
Ph.D., Psychology, 1973, Northwestern University

Joel Zimmerman
Research Director
National Center for State Courts
Ph.D., Psychology, 1973, Northwestern University

August 15, 1978

Abstract

The test-retest reliability of a 28-item self-reported delinquency
instrument was assessed with a sample of 161 junior high school students.
The subjects reported the frequency of their behavior two times, with an
intervening interval of three weeks. Responses to each item were recorded
in three different ways: number of times having engaged in a behavior
during the past year; number of times ever having engaged in it; and
whether or not the subject admitted the behavior at all, without regard to
frequency. Overall test-retest scale correlations were .89 or greater,
depending upon the type of data that was considered. On the average, a
total of 78% of the subjects reported exactly the same frequencies of
behavior on the second completion of the questionnaire as they had on the
first. It is concluded that subjects can provide reliable information
pertaining to their delinquent behavior, and that the information can be
interpreted reliably in any of several different ways.

Establishing the Reliability of Self-reported Delinquency Data[1]

Paul K. Broder

Joel Zimmerman

National Center for State Courts[2]

The use of self-report measures of delinquent behavior is well established in the literature. Their acceptability has grown largely because of deficiencies that have been noted in the research use·of legal categories and classifications of delinquency.

By its nature a dichotomous classification, the adjudication of a youth as delinquent does not distinguish well among the actual behaviors which led to adjudication and which frequently are of interest to social scientists.[3] It also is widely believed by social scientists that delinquent behavior occurs much more frequently and is more widespread than reported in official records.[4] ·The interpretation of official records often is made problematic by variations from jurisdiction to jurisdiction in the statutes pertaining to juvenile delinquency and in the treatment of juvenile offenders. Many cases are handled informally, and formal action may apply differentially to various subgroups within the population when it does occur.[5]

Studies of self-reported and unofficial, or "hidden," delinquency began appearing in the 1940s.[6] Even to this date, however, the use of self-report measures is not without its detractors, who contend that many of the apparently contradictory findings concerning juvenile delinquency may have resulted largely from inconsistencies in the measurement techniques employed in self-report studies.[7] Nevertheless, social scientists have turned increasingly to the use of self-reported delinquency measures in attempting to understand the causes and correlates of delinquent behavior.

Two principal considerations in the interpretation of self-reported delinquency data are the reliability and validity of the subjects' responses. A number of studies have addressed issues of validity, specifically how well subjects' responses match other sources of information about their behavior. Several types of checks have been proposed and used to investigate the veracity of subjects' responses, including comparing self-report responses to official police and court records, interviewing persons believed to be knowledgeable about subjects' behavior, comparing the responses of "known" groups of subjects to one another, using interlocking items or "lie" scales to detect inconsistencies in response patterns, and predicting other variables which are believed to be related to delinquent behavior from the self-report responses.[8] Although no one method is used standardly, the research on this topic suggests that subjects' responses are generally consistent with other available indicators of their behavior.[9]

Relatively few studies have addressed the issue of the reliability of self-reported delinquency data, particularly the retest stability of subjects' responses. This issue is particularly salient in any longitudinal research concerning the changes in delinquent behavior among particular groups of subjects. Dentler and Monroe[10] found that 92% of subjects' responses to a five-item scale remained the same over a two-week period. However, as Farrington[11] pointed out, the possibility of subjects having remembered their initial responses may have resulted in an underestimation of measurement error. Belson, Millerson and Didcott[12] reported 88% consistency in responses over a one-week period to 44 items concerning theft which required dichotomous responses. Elliott and Voss[13] reported stable patterns of concurrent validity over a three-year period, but did not relate it to consistency in individual subject's responses. On the other hand, Farrington reported that approximately one-quarter of his subjects' initial admissions turned into denials over a two- to three-year period.

The present research directly addresses the reliability of self-reported delinquency data, insofar as it is expressed in the retest stability of subjects' responses. In this study, a sample of junior high school students completed a questionnaire containing, among other items, a number of questions pertaining to delinquent behavior. After a three-week period, they completed another questionnaire in which the self-reported delinquency items were repeated. The comparability of responses to the two administrations of the questionnaires is the main focus of this report.

Method

Subjects. A total of 161 boys and girls in grades 7 through 9 of a public junior high school in an upper middle-class neighborhood in the Omaha area participated in this research. The youths participated in the study with the consent of their parents. (Parental consent was obtained from 75% of those from whom it was sought.)

The data reported in this paper were collected as part of a larger study being conducted in Baltimore, Indianapolis, and Phoenix, to explore the relationship between learning disabilities and juvenile delinquency. The main research sample was composed of a heterogeneous group of 1,030 male public school children who were between the ages of 12 and 16, and 665 males and females who were between the ages of 12 and 17 and who had been adjudicated delinquent by juvenile courts. The procedures of the study included the individual administration of a 25-minute interview, which included questions related to the subjects' socioeconomic status and family background, attitudes toward school and social desirability. The main portion of the interview was a set of items pertaining to self-reported delinquency.[14] The interview was administered in summer, 1978, as part of pretesting assessment procedures.

The Omaha sample was included in the research only for the purposes of investigating the interview materials generally, and the test-retest reliability of the self-report items in particular. The children, who were from 10 social studies classes, met as groups in their classrooms with the research staff. Each question was read aloud by the researcher.

The students were asked to read along silently and to write an answer to each question. The questionnaires first were administered in December, 1977. Three weeks later, in January, 1978, the self-report items were presented a second time. Following that, the children rated the serious-ness of each of the behaviors referred to in the self-report items according to a five-point scale that reflected a quasi-normal distri-bution with a mean of 3.0.[15] Each of the two sessions took approximately 35 to 40 minutes.

The self-report items. The 28 self-report items were adapted from previous research of the Institute for Juvenile Research (IJR).[16] The items were selected to include several different types of behaviors ranging widely in seriousness, from very low to moderately high. The items have been grouped conceptually into seven groups of four items each, each group representing a different type of behavior. The groups were formed judgmentally by considering the results of IJR's cluster analyses of their items, and the clusters of offenses suggested by the results of a survey of juvenile court personnel.[17] The clusters suggested by the two sets of data are quite similar; however, the empirical results of the IJR analyses were given preference when the two schemes placed a particular act into different categories.

The items to which the subjects responded are listed in Appendix A. Each item is identified with a code to facilitate easy item identi-fication. The code consists of the first two letters of the category to which the item belongs (i.e., status acts, miscellaneous delinquent acts, drug-related acts, automobile-related acts, alcohol-related acts,

criminal acts, and violent acts), and a number from 1 to 4. The ordinal position of the item within the set of self-report items in the questionnaire also is shown.

For each of the 28 items, the participants reported how many times they ever had engaged in the behavior, and how many times they had engaged in that behavior within the past year. Frequencies were recorded exactly as given, up to a maximum of 99. All responses of 100 or more, and responses to the effect of "all the time," "hundreds of times," and so on, were coded as 99. Responses such as "a few times," or "every now and then," were left blank and considered as missing data. Although confidentiality had been assured, a few children (no more than two for any item) still refused to provide some information. When information was refused, the data were treated as missing.

Results

Item characteristics. The responses to each self-report item for each of the two administrations of the questionnaire were recorded in three ways for purposes of analysis. First, the number of times a subject admitted ever having engaged in each act ("frequency-ever") was recorded. Second, the number of times a subject admitted having engaged in each act in the past year ("frequency-year") was recorded. Finally, the frequency-ever responses were converted to dichotomized responses. That is, a score of 1 was recorded if a subject admitted ever having engaged in an act one or more times, and a score of 0 was recorded if the subject denied ever having engaged in the act. In addition, the frequency data were transformed into logarithms, according to the following formula:

6

$$\text{Log Frequency} = \text{Logarithm}_{10} (\text{Raw Frequency} + 1)$$

This transformation into logarithmic equivalents reduces the statistical effects of very large numbers, which one could argue are less plausible and distort interpretation of the data. The effect of this transformation on particular scores is that a report of 0 is unchanged ($\text{Log}_{10} (0 + 1) = 0$); reported frequencies between 1 and 9 are transformed into a range of .30 to 1.00; and reported frequencies between 10 and 99 are transformed into a range from 1.04 to 2.00.

In sum, one might consider at least five types of scoring for each of the items: frequency-ever (raw or log); frequency-year (raw or log); and proportion of the sample admitting to the behavior. The raw frequencies of occurrence and the proportions of subjects admitting the behaviors for each self-report item on each questionnaire administration are presented in Appendices B through D, along with information about each item's test-retest stability. The overall mean frequencies of acts, per item of self-reported activities, are summarized in Table 1.

Insert Table 1 about here

The means reported in this table are mean frequencies per category per item. As expected, the mean frequencies of self-reported activities ever engaged in were greater than the means for the past year. On the average, considering both administrations of the questionnaire, each

youth reported having engaged in 2.7 acts per item ever, and 1.4 acts per item within the past year. The overall logarithmic means were .19 and .12 which, retransformed to raw score scale, were equivalent to averages of .53 acts ever engaged in, and .30 acts engaged in during the past year. There was little difference in the overall mean frequencies between the first and second administrations of the questionnaire.

The mean frequencies per item for acts ever engaged in ranged from near zero to 12.2. The comparable logarithmic scores were equivalent to a range of zero to 4.4 acts per youth. Reported frequencies were strongly negatively correlated with the rated seriousness of the items. The correlation between mean rated seriousness and mean reported frequency ever (raw scores) for the 28 self-report items was -.71 for the first questionnaire administration and -.67 for the second. The comparable correlations for log data were -.79 and -.76, respectively. All the correlations were significant ($p < .01$). This essentially supports the common-sense notion that more serious delinquent behaviors are engaged in less frequently.

With the dichotomized data, one can meaningfully consider the number of different types of delinquent behaviors a subject admitted having engaged in. To obtain an indication of overall self-reported involvement in delinquent behavior, each subject was given a score from 0 to 7, reflecting the number of delinquency categories in which he or she reported experience. Overall, subjects reported activity in a mean of 3.49 categories on the first administration of the questionnaire, and 3.43 on the second.

Test-retest item reliabilities. There are several indices that may be used to assess the test-retest reliability of the individual self-report items. Three of these measures are summarized in Table 2.

Insert Table 2 about here

Consider first the mean correlations between the frequencies reported for the items on the two test administrations. (Mean correlations were calculated on the basis of Fisher Z-scores, which then were converted to correlation equivalents.) The mean test-retest item correlation ranged from .65 ($p < .01$) for the frequency-year data, to .74 ($p < .01$) for the logarithmic frequency-ever data. Of the 28 items, 18 had test-retest correlations of .70 ($p < .01$) or greater, when the log frequency-ever data were considered.

Another way of looking at the retest stability of the self-report items is to consider the proportion of subjects who gave identical responses to the items on the two questionnaire administrations. As shown in Table 2, the mean proportion of subjects who gave identical responses to the items ranged from .78 for the frequency-ever data, to .91 for the dichotomous data. This relative relationship was reversed when the mean proportion of subjects who indicated lower frequencies of behavior on the second test than the first, which actually is an index of unreliability, was considered. Although the test-retest item correlations tend to indicate that the frequency-ever data are most reliable, responses are

most likely to remain the same when dichotomous data are considered. The lowest figure of .78, however, indicates that more than three-quarters of the responses are stable even when frequency data are considered.[18]

Categories of delinquent behavior. Although the reliability of individual self-report items is desirable, it·is important to determine whether the questionnaire as a whole, or significant groups of items within the questionnaire, are performing reliably. Table 3 shows the mean frequencies of response and the mean proportions of subjects admitting behavior in each of the seven categories of delinquent behavior.

Insert Table 3 about here

The means and proportions reflect the characteristics of the individual items within the categories. As expected, increased stability is gained by considering groups of items, resulting in test-retest reliabilties which are higher than those for the individual items. The mean test-retest reliabilities per category are shown in the top row of Table 4. The frequency-ever data evidenced the highest test-retest correlations, and the log transform enhanced the reliabilities of the frequency data somewhat.

Finally, consider the test-retest reliabilities of subjects' overall

scores, which are summarized in the bottom row of Table 4. For the

frequency data, each subject received a score that reflected the overall

mean frequency per category per self-report item. For the dichotomized

data, each subject received a score indicating the number of delinquent

categories in which he or she admitted behavior. The test-retest

reliabilities of these scores were very high, with correlations equal to

or greater than .89 ($p < .01$) for all the types of data. This indicates a

very high degree of test-retest stability of subjects' scores, based upon

the self-report instrument as a whole.

Split-sample correlations. The final means of assessing the

reliability of the self-report items that is considered here is a

comparison of the responses to the items that were made by randomly

constituted subgroups of the sample. This is somewhat analogous to the

calculation of split-half reliabilities, except that subgroups of

subjects are the focus of the analysis, rather than subgroups of items.

To perform this analysis, the sample was divided randomly into groups of

80 and 81 youths. Mean scores were calculated for each item for each

group, and the scores for the items for the two groups were correlated.

If the items were responded to reliably, one would expect the two groups

of subjects to yield comparable scores.

This was the case, as indicated in Table 5. The split-sample

Insert Table 5 about here

----

correlations for the two administrations of the questionnaire showed identical patterns. The correspondence of responses for the log frequency data and the dichotomized data was nearly perfect. The lowest correlation (.83, $p < .01$) resulted from the raw frequency-year data.

Discussion

The data presented here indicate that subjects can provide reliable information pertaining to their delinquent behavior, and that the information can be interpreted reliably in any of several different ways. Subjects are capable of providing reliable information about the frequencies of their behaviors, as well as simply whether or not they have engaged in particular behaviors. When the researcher's interest lies in the frequencies of particular types of delinquent behavior, it appears that asking the subject to report the frequencies with which he or she has ever engaged in particular behaviors may produce data that are somewhat more reliable than asking about the frequencies of behavior within a particular time period. It also appears that rational transformations of the data (e.g., conversion to logarithms) may enhance measurement reliability.

A conclusion concerning the reliability of these data would be supported further by a positive correlation between the reports of the Omaha youths and those of other public school students. Table 6 shows the correlations between the responses of the Omaha sample and those of the main research sample. Overall, there is a positive correlation between the delinquent behavior reported by the main research sample and that

----

Insert Table 6 about here

----

of the Omaha sample. These correlations are particularly strong when only the public school sample is considered, especially for the frequency-ever data and the dichotomized data. The correlations still are significant when only the delinquent sample is considered, but are lower, which would be expected from the differences in the samples' compositions.

The establishment of the reliability of self-reported delinquency data is necessary to the accurate interpretation of findings. It is particularly important in the collection of longitudinal data, where changes in behavior over time are of primary interest.

Footnotes

[1]This research was supported by Grants Numbers 76-NI-99-0133, 76-JN-99-
0022, 76-JN-99-0022S1, and 76-JN-99-0021 from the Office of Juvenile
Justice and Delinquency Prevention, Law Enforcement Assistance Administration,
U.S. Department of Justice. Points of view or opinions
in this paper are those of the authors and do not represent the official
positions or policies of the U.S. Department of Justice, Creighton
University, or the National Center for State Courts. Further information
and documentation pertaining to this research may be obtained from the
second author at the National Center for State Courts, 300 Newport
Avenue, Williamsburg, Virginia 23185.

[2]At the time this research was conducted, the authors were affiliated
with the Creighton University Institute for Business, Law and Social
Research, Omaha, Nebraska. The authors wish to acknowledge the contri-
butions made to this research by other members of the Creighton Institute
staff, particularly by Cassie Spohn, Betty Bullock and Caren Handleman,
who assisted in the data collection in Omaha.

[3]Many researchers have pointed out the distinctions between the legal
and behavioral considerations of delinquency. For example, see
A.S. Berger, "The Construction and Interpretation of a General Measure
of Delinquency" (Chicago: Institute for Juvenile Research, 1975,
unpublished draft cited with permission of the author and the Institute
for Juvenile Research); M. Gold, Delinquent Behavior in an American
City (Belmont, California: Brooks/Cole, 1970); T. Sellin & M.E. Wolfgang,
The Measurement of Delinquency (New York: Wiley, 1964).

[4]For example, see W.A. Belson, Juvenile Theft: The Causal Factors
(London: Harper & Row, 1975); M.L. Erickson & L.T. Empey, "Court
Records, Undetected Delinquency and Decision Making," Journal of
Criminal Law, Criminology and Police Science, 54 (1963), 456-469;
Gold (note 3); R.J. Rivera, Juvenile Delinquency in Illinois: Highlights
of the 1972 Adolescent Survey (Chicago: Institute for Juvenile
Research, 1972).

[5]R.D. Wirt & P.F. Briggs, "The Meaning of Delinquency," ed. H.C. Quay,
Juvenile Delinquency: Research and Theory (Princeton: Van
Norstrand, 1965).

[6]Among the first studies of self-reported delinquency was A.L. Porterfield,
Youth in Trouble: Studies in Delinquency and Despair with Plans
for Prevention (Fort Worth: Leo Potishman Foundation, 1946). Among the
early studies taking different approaches to assessing hidden delinquency,
for example, the use of respondents who were presumed to know about
delinquent behavior, was F.J. Murphy, M.M. Shirley & H.L. Witmer, "The
Incidence of Hidden Delinquency," American Journal of Orthopsychiatry,
16. (1946), 686-696.

[7]For example, see A.J. Reiss, Jr., "Settling the Frontiers of a Pioneer
in American Criminology: Henry McKay," ed. J.F. Short, Jr., Delinquency,
Crime and Society (Chicago: University of Chicago Press, 1976);

L.D. Savitz, M. Lalli, & L. Rosen, City Life and Delinquency: Victimization, Fear of Crime and Gang Membership (Washington: National Institute for Juvenile Justice and Delinquency Prevention, Office of Juvenile Justice and Delinquency Prevention, Law Enforcement Assistance Administration, U.S. Department of Justice, 1977).

[8] Measures of validity are discussed in a number of reports. For example, see D.S. Elliott & H.L. Voss, Delinquency and Dropout (Lexington, Mass.: Lexington Books, 1974); Gold (note 3); R.H. Hardt & S. Peterson-Hardt, "On Determining the Quality of the Delinquency Self-report Method," Journal of Research in Crime and Delinquency, 14 (1977), 247-261.

[9] Elliott and Voss (note 8), for example, found that their subjects reported about 80% of the offenses that were known to the police. Gold (note 3) had reports from informants about the behavior of 125 members of his sample. Of this group, 72% reported everything that the informants had concerning their delinquent behavior.

[10] R.A. Dentler & L.J. Monroe, "Social Correlates of Early Adolescent Theft," American Sociological Review, 26 (1961), 733-743.

[11] D.P. Farrington, "Self-reports of Deviant Behavior: Predictive and Stable?" Journal of Criminal Law and Criminology, 64 (1973), 99-110.

[12] W.A. Belson, G.L. Millerson & P.J. Didcott, The Development of a Procedure for Eliciting Information from Boys about the Nature and Extent of their Stealing (London: The Survey Research Centre, London School of Economics and Political Science, 1968).

[13] Elliott & Voss (note 8).

[14] The complete interview guide is available upon request from the authors at the cost of duplication and postage.

[15] A forced distribution scaling procedure was used. The subjects were instructed to pick the three most serious offenses in the list and assign them a score of 5. After that, the three least serious offenses were assigned a score of 1. Then, the five next-to-most serious offenses were rated 4, and the five next-to-least serious offenses were rated 2. Finally, the remaining 12 items were given a rating of 3.

[16] J.W.C. Johnstone, The Family and Delinquency: A Reappraisal (Chicago: Institute for Juvenile Research, 1976, unpublished draft cited with permission of the author and the Institute for Juvenile Research)

[17] L.E. Cohen, Pre-adjudicatory Detention in Three Juvenile Courts: An Empirical Analysis of the Factors Related to Detention Decision Outcome, Report SD-AR-8 (Washington: National Criminal Justice Information Statistics Service, Law Enforcement Assistance Administration, U.S. Department of Justice, 1975).

[18]One might suggest, as Farrington (note 11) does, that response recall has inflated the apparent stability of the responses. It is difficult to confirm or refute this hypothesis clearly within this set of data. One might argue that subjects would be better able to recall their responses to items that were more serious in nature. This is supported by a strong positive correlation (.80, $p < .01$) between items' mean rated seriousness and the proportion of subjects who reported identical responses (using the frequency ever data). On the other hand, subject's engaged in less serious behaviors less often, and may have been better able to recall their actual behaviors. The distinction between recall of previous behavior and recall of previous responses is a subtle one. It is noteworthy, however, that the .78 consistent responses was obtained with actual frequencies to 28 items which were embedded within a longer questionnaire. It would be difficult to argue that memory for this number of responses can account for this level of retest stability.

Table 1

Self-report Item Characteristics

	Frequency Ever		Frequency Past Year	
	Raw	Log	Raw	Log
Mean No. Acts Per Subject Test 1 (t_1)	2.70	0.19	1.42	0.12
Range: Lower Bound	0.01	*	0.00	0.00
Upper Bound	10.45	0.68	6.73	0.50
Mean No. Acts Per Subject Test 2 (t_2)	2.65	0.18	1.42	0.11
Range: Lower Bound	0.01	*	0.00	0.00
Upper Bound	12.24	0.73	6.43	0.53

*Less than 0.005

Table 2

Indices of Test-retest Item Reliabilities

	Frequency Ever		Frequency Past Year		
	Raw	Log	Raw	Log	Dichot.
Mean Test-retest Item Correlation (r_{12})	.70*	.74*	.65*	.65*	.69*
Range: Lower Bound	.02	.33	-.01	-.01	.40
Upper Bound	1.00	1.00	.99	.96	1.00
Mean Proportion of Subjects Giving Identical Responses in Each Test ($t_1 = t_2$)	.78	**	.83	**	.91
Range: Lower Bound	.28	**	.34	**	.76
Upper Bound	1.00	**	1.00	**	1.00
Mean Proportion of Subjects Indicating Greater Frequency in Test 1 than Test 2 ($t_1 > t_2$)	.13	**	.10	**	.05

Note: Table entries based on 28 items

* $p < .01$

** Data unaffected by log transform

Table 3

Summary Data Based on Item Clusters

Mean Frequency of Occurrence Ever

Item Cluster	Test 1 Mean	Test 1 SD	Test 2 Mean	Test 2 SD	$r_{12(raw)}$	$r_{12(log)}$
ST	4.47	8.08	3.86	8.62	.753	.871
MI	4.61	11.06	4.12	11.98	.780	.861
DR	1.29	4.93	1.51	7.13	.849	.924
AU	0.68	1.95	0.53	1.96	.356	.723
AL	6.30	12.71	6.94	13.98	.933	.923
CR	0.48	2.52	0.51	2.86	.649	.676
VI	1.17	4.27	1.05	4.18	.961	.891

Mean Frequency of Occurrences in Past Year

Item Cluster	Test 1 Mean	Test 1 SD	Test 2 Mean	Test 2 SD	$r_{12(raw)}$	$r_{12(log)}$
ST	1.18	2.36	1.27	3.50	.731	.795
MI	2.21	8.64	2.09	10.71	.780	.745
DR	0.91	3.94	1.04	5.09	.780	.881
AU	0.55	2.38	0.29	0.99	.323	.708
AL	4.07	9.82	4.14	11.76	.906	.894
CR	0.20	1.09	0.32	2.19	.438	.639
VI	0.79	3.69	0.78	3.93	.948	.847

Mean Proportion Admitting Occurrence Ever

Item Cluster	Test 1 Mean	Test 1 SD	Test 2 Mean	Test 2 SD	$r_{12(raw)}$
ST	.839	.369	.808	.396	.685
MI	.752	.434	.745	.437	.753
DR	.168	.375	.174	.380	.847
AU	.335	.474	.304	.462	.760
AL	.839	.369	.888	.316	.648
CR	.217	.414	.211	.409	.724
VI	.342	.476	.298	.459	.790

Note: Cluster scores based on 161 subjects; all correlations significant, $p < .01$

Table 4

Overall Reliabilities of Self-report Items

	Frequency Ever		Frequency Past Year		
	Raw	Log	Raw	Log	Dichot.
Mean Test-Retest Correlations Based on Cluster Scores*	.82	.86	.77	.80	.75
Test-Retest Correlations Based on Individual Subjects' Scores**	.90	.93	.92	.90	.89

*Based on 7 item clusters; all correlations significant, $p < .01$

**Based on 161 subjects; all correlations significant, $p < .01$

Table 5

Split-sample Correlations

	Frequency Ever		Frequency Past Year		
	Raw	Log	Raw	Log	Dichot.
Test 1	.88	.98	.86	.97	.99
Test 2	.88	.98	.80	.97	.98
Mean Correlations	.88	.98	.83	.97	.99

Note: Table entries based on mean responses to 28 items made by groups of 80 and 81 subjects; all correlations significant, $p < .01$

Table 6

Correlations of Responses to Items by the Omaha Sample and by the Main Research Sample

	Frequency Ever		Frequency Past Year		
	Raw	Log	Raw	Log	Dichot.
Entire Three-city Sample	.64**	.75**	.64**	.66**	.79**
Public School Subsample	.89**	.90**	.79**	.83**	.90**
Delinquent Subsample	.46*	.45*	.57**	.46*	.44*

Note: Table entries based on 28 items

* $p < .05$

** $p < .01$

Appendix A

Self-reported Delinquency Items

How many times have you ever (in the past year) . . . .

Code	Item	Position
ST1	Cheated on an exam in school or turned in work that was not your own?	1
ST2	Stayed away from school for at least part of the day just because you wanted to?	7
ST3	Been suspended from school?	8
ST4	Been thrown out of class by a teacher?	14
MI1	Taken things that don't cost much from home or school without permission?	4
MI2	Taken something small from a store?	5
MI3	Kept or used something that you knew had been stolen?	11
MI4	Deliberately damaged private or public property?	13
DR1	Used marijuana or hashish (grass, pot, hash)?	17
DR2	Used LSD, mescaline, or other psychedelics?	26
DR3	Used downers or barbiturates (without a prescription)?	27
DR4	Used methedrine (speed) or other uppers or amphetamines (without a prescription)?	28
AU1	Driven a car on the street by yourself?	10
AU2	Driven a car too fast or recklessly?	12
AU3	Ridden around in a car that was stolen for the ride?	23
AU4	Stripped someone else's car of parts to use or sell?	24
AL1	Drunk beer, wine, or liquor with your parent's permission?	2
AL2	Drunk beer, wine or liquor without your parent's permission?	3
AL3	Been drunk?	9
AL4	Bought beer, wine or liquor?	16
CR1	Taken at least $20 or something worth $20 that did not belong to you?	19
CR2	Broken into someone's home, or a store, or some other place in order to steal something?	21
CR3	Used force or threatened to use force to get money from another person?	22
CR4	Stolen a car?	25
VI1	Had a fist fight in which someone got hurt badly enough to go to a doctor or hospital?	6
VI2	Carried a weapon like a gun, knife, or razor in case you had to use it against another person?	15
VI3	Taken part in a gang fight?	18
VI4	Used a weapon like a brick, knife, or razor in a fight?	20

Appendix B

Mean Frequency of Occurrence Ever

Item	Test 1 (t_1)		Test 2 (t_2)		$r_{12(raw)}$	$r_{12(log)}$	Prop. $t_1 = t_2$	Prop. $t_1 > t_2$
	Mean	SD	Mean	SD				
ST2	2.65	9.21	2.75	11.99	.816	.729	.646	.186
ST3	0.02	0.14	0.04	0.27	.321	.425	.981	.006
ST4	4.88	14.80	3.20	10.21	.325	.826	.596	.255
MI1	5.18	14.27	4.28	12.50	.603	.674	.506	.256
MI2	4.44	11.75	4.44	12.75	.845	.832	.559	.193
MI3	3.65	11.59	2.93	11.55	.667	.616	.540	.217
MI4	5.19	14.88	4.84	14.42	.648	.748	.540	.248
DR1	4.78	18.20	4.70	18.25	.878	.923	.882	.099
DR2	0.03	0.26	0.02	0.18	.260	.336	.975	.012
DR3	0.22	2.38	0.66	7.81	.995	.950	.969	.006
DR4	0.1?	0.77	0.65	7.80	.524	.749	.969	.025
AU1	1.64	5.30	1.19	3.63	.488	.716	.752	.168
AU2	0.59	2.80	0.65	5.93	.017*	.330	.870	.087
AU3	0.06	0.49	0.02	0.19	.184*	.357	.975	.012
AU4	0.44	4.11	0.27	2.42	.435	.879	.975	.006
AL1	10.38	20.38	12.24	23.37	.733	.724	.277	.283
AL2	7.94	18.79	7.97	18.56	.764	.808	.522	.180
AL3	3.96	12.88	4.94	16.40	.922	.900	.706	.163
AL4	2.48	10.54	2.61	13.76	.811	.853	.857	.112
CR1	0.86	7.85	1.46	11.02	.701	.639	.901	.087
CR2	0.14	1.05	0.07	0.48	.181*	.486	.963	.025
CR3	0.91	4.56	0.50	2.02	.661	.702	.845	.093
CR4	0.01	0.08	0.01	0.08	1.000	1.000	1.000	.000
VI1	0.30	0.87	0.15	0.48	.553	.615	.894	.087
VI2	3.45	15.66	3.19	15.64	.988	.896	.863	.199
VI3	0.57	1.69	0.50	2.48	.648	.742	.783	.137
VI4	0.35	1.63	0.37	3.95	.215	.541	.932	.062

*All correlations except these significant, $\underline{p} < .05$

Appendix C

Mean Frequency of Occurrence in Past Year

Item	Test 1 (t_1) Mean	SD	Test 2 (t_2) Mean	SD	$r_{12(raw)}$	$r_{12(log)}$	Prop. $t_1=t_2$	Prop. $t_1>t_2$
ST1	3.01	7.20	4.19	13.07	.641	.769	.509	.230
ST2	0.68	2.07	0.40	1.02	.455	.591	.770	.143
ST3	0.00	0.00	0.01	0.08	--	--	.994	.000
ST4	1.02	3.03	0.47	1.12	.609	.588	.733	.174
MI1	2.99	11.99	2.34	11.27	.563	.579	.625	.206
MI2	1.99	9.57	1.96	11.12	.730	.717	.770	.118
MI3	1.69	8.39	1.62	9.78	.879	.595	.688	.182
MI4	2.18	8.60	2.45	11.46	.666	.597	.650	.219
DR1	3.45	15.31	3.56	16.37	.821	.894	.876	.062
DR2	0.02	0.14	0.02	0.18	.510	.481	.975	.006
DR3	0.1?	1.21	0.30	3.17	.993	.957	.969	.000
DR4	0.06	0.36	0.30	3.55	.031*	.235	.963	.031
AU1	1.52	8.21	0.78	2.79	.249	.639	.795	.143
AU2	0.27	0.96	0.28	2.03	.108*	.379	.882	.087
AU3	0.02	0.32	0.01	0.08	-.006*	-.006*	.988	.006
AU4	0.40	4.05	0.08	0.81	.224	.521	.969	.019
AL1	6.73	16.39	6.43	14.56	.548	.637	.338	.294
AL2	5.37	14.43	4.68	13.60	.899	.817	.559	.217
AL3	2.48	9.58	3.33	13.97	.834	.843	.752	.124
AL4	1.74	9.19	2.14	12.58	.833	.839	.863	.106
CR1	0.43	3.97	0.99	8.72	.440	.514	.925	.050
CR2	0.08	0.80	0.02	0.24	.091*	.251	.975	.019
CR3	0.30	1.10	0.27	1.17	.876	.800	.888	.075
CR4	0.00	0.00	0.00	0.00	--	--	1.000	.000
VI1	0.12	0.58	0.05	0.25	.441	.536	.944	.043
VI2	2.80	14.08	2.81	15.50	.954	.869	.882	.075
VI3	0.15	0.51	0.20	0.89	.533	.662	.894	.043
VI4	0.11	0.60	0.07	0.79	.023*	.129*	.963	.031

*All correlations except these significant, $p < .05$

Appendix D

Mean Proportion Admitting Act

Item	Test 1 (t_1) Mean	SD	Test 2 (t_1) Mean	SD	r_{12}	Prop. $t_1=t_2$	Prop. $t_1>t_2$
ST1	.689	.464	.677	.469	.598	.823	.093
ST2	.366	.483	.323	.469	.522	.783	.130
ST3	.019	.136	.025	.156	.568	.981	.006
ST4	.547	.499	.522	.501	.826	.913	.056
MI1	.509	.501	.478	.501	.542	.770	.130
MI2	.453	.499	.497	.502	.742	.870	.043
MI3	.391	.490	.398	.491	.493	.758	.118
MI4	.534	.500	.497	.502	.679	.839	.099
DR1	.168	.375	.174	.380	.847	.957	.019
DR2	.019	.136	.012	.111	.399	.981	.012
DR3	.025	.156	.037	.190	.601	.975	.006
DR4	.043	.205	.025	.156	.749	.981	.019
AU1	.280	.450	.236	.426	.697	.882	.081
AU2	.130	.338	.106	.308	.587	.913	.056
AU3	.025	.156	.019	.136	.568	.981	.012
AU4	.050	.218	.043	.205	.932	.994	.006
AL1	.783	.414	.826	.380	.553	.857	.050
AL2	.522	.501	.602	.491	.721	.857	.031
AL3	.385	.488	.342	.476	.856	.932	.056
AL4	.168	.375	.106	.308	.765	.938	.062
CR1	.093	.292	.099	.300	.608	.932	.031
CR2	.056	.230	.031	.174	.736	.975	.025
CR3	.130	.338	.112	.316	.565	.907	.056
CR4	.006	.079	.006	.079	1.000	1.000	.000
VI1	.161	.369	.106	.308	.673	.919	.068
VI2	.180	.385	.143	.351	.779	.938	.050
VI3	.211	.211	.174	.380	.646	.888	.075
VI4	.081	.273	.043	.205	.719	.963	.037

END