182324

75003

DOCUMENT RESUME

ED 182 321 TM 009 988

AUTHOR Blakely, Craig H.: And Others
TITLE Alternative Scaling Procedures for Constructing a Self-Report Delinquency Measure.
SPONS AGENCY National Inst. of Mental Health (DHEW), Bethesda, Md. Center for Studies of Crime and Delinquency.
PUB DATE [Sep 79]
GRANT MH-29160
NOTE 26p.; Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Delinquent Behavior: Elementary Secondary Education: Interviews: *Questionnaires: Self Evaluation: Test Reliability: Test Validity: *Weighted Scores
IDENTIFIERS Multitrait Multimethod Technique: Self Report Inventory

ABSTRACT

Ten different weighting schemes were applied to the responses to a self-report deliquency measure and the resulting scores were evaluated in a multitrait-multimethod analysis. Four subscores and a total score were derived from the 35-item questionnaire which was administered orally on four occasions during an 18-week deliquency prevention project. Results of the questionnaire were also compared with records from the probate court and other police records. The comparative advantages of official records and self-report inventories were discussed. The results of the analysis showed that neither weighting by seriousness of the deliquent behavior nor by frequency of occurence had any significant effect on either the reliability or the validity of any of the scale scores. (Author/CTM)

Alternative Scaling Procedures for Constructing a Self-Report Delinquency Measure

Craig H. Blakely, Martin G. Kushler,

Julia A. Parisian, and William S. Davidson II

Michigan State University

This paper was presented at the 87th annual convention of the American Psychological Association, New York, New York, 1979. This work was completed under a grant from the Center for the Studies of Crime and Delinquency from the National Institute of Mental Health #MH29160.

Abstract

This paper presented the methodology involved in the construction of a self-report instrument of delinquent behavior. Ten frequency and seriousness weighting schemes were compared. Reliability was assessed through internal consistency and test-retest methods. Validity was assessed through multitrait-multimethod analyses as well as by the self report data's relationship to official court and police data. The relative effect of the weighting schemes and their implications were discussed.

Alternative Scaling Procedures for Constructing a Self-Report Delinquency Measure

Those who have devoted their energies to delinquency research have usually been dependent upon inadequate archival records for typical outcome variables. The logical rationale behind the use of court contacts and dispositions as well as police contacts and referrals as primary outcome variables in delinquency research is obvious. However, the major drawbacks are twofold and perhaps equally clear. The most frequently mentioned criticism of the use of official archival data is that it is more a measure of police behavior than it is a measure of deviant behavior (Gold, 1966; Farrington, 1973; Williams & Gold, 1972). In fact, estimates of official detection of delinquents have ranged from three to twenty percent (Davidson, 1976; Krohn, Waldo & Chiricos, 1975; Williams & Gold, 1972). Early studies developed encompassing theories of juvenile delinquency incorporating official records as primary outcome data. These researchers tended to define delinquency in terms of socio-cultural antecedent conditions (Cloward & Ohlin, 1960; Cohen, 1955; Glueck & Glueck, 1950). These works are frequently seen as having been more "theoretical and speculative than empirical in their approach" (Peterson, Urban, & Vondracek, 1975). These studies frequently attributed or suggested a causal relationship between official delinquency and demographic variables such as sex, race, age and socio-economic status. Recently, these theories have come under fire due to their inherent dependency on official outcome data (Williams & Gold, 1972; Krohn, et al., 1975). The resultant concern

for the process of labeling still leaves the problem of how best to measure the concept of delinquency.

A second primary fault with the sole use of official record data as a dependent measure in delinquency research lies with the tendency for many current researchers to attempt to identify juveniles before they become a part of the official system. In short, the evaluation of prevention programs or treatments aimed at pre-delinquents has become a major focus of contemporary research. The problem becomes the credibility of evaluations of preventative programs using official recidivism as a major outcome variable when the target population has an extremely limited incidence or future probability of official contacts with the juvenile justice system.

These criticisms of official outcome data have led to the development of alternative dependent measures of delinquency; the most popular of which has been self report (Sellin & Wolfgang, 1963; Nye & Short, 1957). The methods by which one defines delinquency is not only crucial to the process of measurement, but as Hirschi and Selvin have noted, "How one defines delinquency determines in large part how one will explain delinquency" (Hirschi & Selvin, 1969). Indeed, the literature suggests a clear need for the inclusion of both official court and police records as well as self-report estimates of delinquent behavior in the area of research and development in juvenile delinquency (Gold, 1966; Erickson & Empey, 1963; Williams & Gold, 1972; Krohn et al., 1975; Hindelang & Hirschi, 1977).

The development of early self-report measures tended to be attempts at distinguishing a difference between "official delinquents", those currently incarcerated in institutional facilities, and non-delinquents, those without police records, most typically "normal high school students" (Kulik, Stein &

Sarbin, 1968. This procedure has been criticized in that those incarcerated tend to have few reasons to conceal delinquent acts, while those not currently under jurisdiction may react defensively regardless of all the experimenter's attempts at guaranteeing the anonymity of the collected information (Williams & Gold, 1972). Erickson and Empey have suggested that these studies even violate the definition of official delinquency (Erickson & Empey, 1963). Crucial sampling issues have been two frequently overlooked.

Once self-reported delinquency was accepted as an alternative measure of delinquency, traditional reliability and validity issues were examined. Various reliability estimates have been employed in past research. The methodology has included lie tests, test-retest, split-half, Guttman scaling, and internal consistency alphas. Although there have been inconsistencies in the literature, some positive evidence seems to be emerging. Self-report strategies have tended to be relatively stable over time (Belson, 1968) and fairly unidimensional (Farrington, 1973).

Validity issues have been dealt with by including such criterion variables as police records (Kulik, et al., 1968), court convictions (Erickson & Empey, 1963; Farrington, 1973; Blackmore, 1974), and informant records (Gold, 1970). The typical methodology has incorporated a predictive or concurrent validity framework.

Although inconsistencies in the reliability and validity of self-report instruments exist, they appear to be due to differences in the criterion measures, subject samples, or item content rather than an inherent weakness in the concept of self-reported delinquency estimates. Certainly at least as many criticisms could be leveled against the use of official archival data.

Additional developments in the area of measurement of self-reported estimates of delinquency have indicated that further refinements may be useful. In particular, the clarification of the utility of the application of various weighting schemes to self-reported estimates of delinquent behavior seems to be a crucial issue

that is unresolved to date. The basic self-report measure simply considers categorical information. That is, does the subject report having committed a certain defined behavior during a particular time interval? Weighting schemes have traditionally considered two dimensions, frequency and seriousness. Farrington (1973) found that the addition of weighting schemes for frequency and seriousness did not add significantly to the predictive validity of a self-report instrument designed with simple unit weights applied to categories of offenses reported. On the other hand, Gold (1966) found that the frequency of offenses committed was a critical determinant of police detection while the seriousness of the offense weighted heavily on dispositional factors. Erickson and Empey (1963) found that repeat offenders and non-official delinquents did not differ significantly in the proportion of categories of offenses committed, however, the two groups differed drastically in the frequency of offenses and the most serious offense committed. They also found that court dispositions tended to be for the most serious offenses committed by the repeat offender.

The application of weighting schemes to scaling procedures has been a topic of debate in many settings. In the area of self-reported delinquency research, clear inconsistencies have been noted in the literature (Erickson & Empey, 1963; Gold, 1966; Farrington, 1973). The assumption by many of the researchers in this area that their instruments are superior, or even compatible, suggests a need for a direct comparative evaluation of various weighting procedures.

The Current Research

In response to the research issues raised above, this paper will present data relevant to the development of a self-report delinquency measure. The instrument utilized in this research was developed by drawing items from or modifying items from the measures developed by Lincoln, Teillmann, Klein, and Labin (1977) and Gold (1970) as well as the inclusion of new items. The item set was designed to determine self-reported incidence of behaviors representing a wide spectrum of

frequently occurring delinquent activities. Thus, infrequently occurring behaviors such as serious crime against a person have been excluded. The particular behaviors selected are representative of typical categorizations of delinquent acts such as those compiled by Sellin and Wolfgang (1964) and Rossi (1974). In addition, the 35 item set includes five filler items depicting positive activities in an attempt to discourage response patterns or bias. Respondents were asked whether they had committed each behavior once, twice, more than twice or not at all during each of three time intervals (the last six weeks, the last year and ever).

Interviewing Procedures. The self-report instrument was administered by project staff at four points during each youth's involvement in a delinquency prevention project; at the point of referral, six weeks, twelve weeks, and eighteen weeks from referral (termination). The instrument was administered as part of a process interview package. These interviews were open ended and geared to gather information relevant to the youth's progress in the project (Kantrowitz, Davidson, Blakely & Kushler, 1978). The interview process was explained to the youth and his/her parents prior to their formally joining the project. The confidentiality and anonymity of the information gathered was stressed at this point as well as during each scheduled interview. Upon entering the project, process interviews were conducted at the specified intervals with the youth, a parent, and a peer nominated by the youth.

Interviews tended to last from one to one and one half hours. The self-report measure was administered immediately following the process interview. In order to avoid problems due to inconsistent reading abilities of the interviewees, the items were read aloud to the respondents and responses were recorded by the interviewer. The entire interview was audio recorded to minimize data loss.

Archival data. For use in validity estimates, official archival records were collected from the probate court, the county sheriff's office, and several local city police departments. In addition, records were collected from several

department stores and discount houses that frequently petition youth directly to the juvenile court.

Results

Complete data was available on 123 subjects gathered between March, 1977 and December, 1978. The data set, coded in straight frequencies, was first recoded into categorical information. Thus, item responses were available in both unit weights (whether or not the event occurred during the specified time interval) and frequency weights (whether the event occurred once, twice, more than twice, or not at all). These frequency weight alternatives were then crossed with five seriousness weighting schemes; those developed by Sellin and Wolfgang (1964); McEachern and Bauzer (1967); and Rossi (1974); as well as a weighting scheme developed locally. The fifth method applied no seriousness weightings. The seriousness weighting scheme developed locally was drawn from a larger sample of 806. Data was gathered directly from archival records in the county court files. Offenses were categorized and the probability of formal court proceedings was used as an index of offense seriousness. The result was a two by five matrix of available weighting schemes (see Table 1). The seriousness weights were simple multipliers applied to the basic frequency and unit weights. Following the removal of the five postively worded filler items the ten data sets were subjected to several analyses designed to determine their differential reliability and validity properties.

Reliability Analyses. Although, as noted earlier, self report measured tend to be unidimensional (that is, they tend not to break into reliable subscales) an initial factor solution was attempted. A few items were then transferred in order to make rational sense of the scales. The items causing the most intuitive problems tended to be loaded on more than a single factor. The resultant four subscales represented property crimes, crimes involving physical force, school related offenses, and offenses involving substance abuse. The subscales and the

Table 1

Matrix of Frequency and Seriousness Weighting Schemes Applied to the Raw Data Set

Frequency Weightings	No Seriousness Weights	Sellin and Wolfgang Seriousness Weights	McEachern Seriousness Weights	Rossi Seriousness Weights	Local Seriousness Weights	Seriousness Weights
Categorical Data						
Frequency Weighted Data						

total scale were then subjected to reliability analyses for each of the ten weighting schemes.

Internal consistency estimates were calculated on the one year time interval data sets for each of the scales and weighting schemes. Two additional items were deleted following these analyses resulting in a 28 item total scale. The resultant alpha's[1] for the ten seriousness schemes ranged from .77 and .89 with a mean of .84 (see Table 2). The alphas of the subscales .78, .70, .72, and .67 and their respective ranges are also included in Table 2. Though the subscales are obviously less reliable than the total scale, they do provide an adequate degree of internal consistency. The total scales are remarkably consistent. The only consistent difference can be found in the fact that the five frequency weighting schemes are more reliable than their categorical counterparts. However, the difference is not significant.

Test-retest calculations also yielded very consistent information. Across time correlations were computed for each subscale and the total scale within each weighting scheme. Correlations were calculated for the youth's reponses to the one year response interval at each of four data gathering points or waves. Therefore six test-retest intervals were available (time 1 with time 2, time 1 with time 3, etc.). The data presented in Table 3 are the averages of the six correlations for each scale within each weighting method. Again, it should be noted that there is very little variation between the various weighting schemes. Virtually all of the average test-retest correlations are in the sixties and seventies.

In summation, these analyses have suggested that the various scales consistently demonstrated similar reliability properties across weighting schemes. This notion is supported through both internal consistency and test-retest methodology. The

---

[1]Cronbach's alpha coefficient of internal consistency.

Table 2

Total Scale and Subscales mean and range of internal consistency alpha's over ten weighting schemes

	alpha	
	mean of ten weighting schemes	range of ten weighting schemes
Total Scale	.84	.77 to .89
Property Subscale	.78	.65 to .84
Force Subscale	.70	.60 to .75
School Subscale	.72[a]	.46[a] to .80
Substance Abuse Subscale	.67[b]	.35[b] to .79

[a] The .46 was the result of the local weighting schemes application to the school scale item set. Several of the items were zero weighted reducing this scale to a two item scale with a low alpha. Without including the local weighting scheme the mean corrleation would have been .74 with a range of .70 to .79.

[b] These figures were affected by the McEachern weighting scheme. Again, the scale is a two item scale. Without the inclusion of the McEachern scales the mean correlation was .74 with a range of .70 to .79.

Table 3

Average Test-Retest correlations for each scale within each weighting scheme

	Sellin & Wolfgang Categorical	Sellin & Wolfgang Frequencies	McEachern Categorical	McEachern Frequencies	Rossi Categorical	Rossi Frequencies	Local Categorical	Local Frequencies	Straight Categorical	Straight Frequencies
Total Scale	.61	.62	.75	.75	.76	.77	.73	.75	.59	.64
Property Subscale	.61	.65	.75	.70	.67	.62	.64	.63	.62	.64
Force Subscale	.63	.61	.65	.60	.60	.54	.63	.57	.59	.62
School Subscale	.60	.67	.63	.67	.64	.65	.64	.60	.60	.67
Substance Abuse Subscale	.71	.78	.57	.57	.64	.61	.61	.[illegible]	.71	.78

only significant departure appears in the substance abuse scale. On several weighting schemes, most notably the local weighting scheme as applied to the frequency weights, the substance abuse scale fluctuates somewhat. This is not surprising as it is a two item scale. The total scale again seemed to be the best measure available.

Validity analyses. Deleting the substance abuse subscale, the remaining subscales were subjected to multitrait-multimethod analysis (Campbell & Fiske, 1959). Though the correlation matrices do not neatly fit the ideal pattern demonstrating convergent and discriminant properties, the patterns are quite consistent across the various weighting schemes. For the most part, the monotrait-heteromethod correlations or those in the validity diagonals are in the upper sixties and lower eighties. Though these correlations support the notion of convergent validity, particularly in the light that they are generally higher than their respective monomethod-heterotrait correlations, there is little strong evidence to suggest support for the subscales discriminant abilities. Though the means of the monomethod-heterotrait triangle correlations are consistently lower than their respective validity diagonals, the absolute difference is not always that great. In addition, there generally occurred a large range of values in the correlations within the triangles resulting in some correlations being as high or higher than the corresponding validity diagonal correlation. An extraction or summary table from the multitrait multimethod matrix of the McEachern seriousness weights as applied to the frequency weighted data is depicted in Table 4. Clearly there is strong evidence for the convergent properties of the scales but a lack of discriminant evidence exists. These findings should not be taken in a negative light. Recall, as noted earlier, that self report measures tend to be unidimensional. The fact that the intersource correlations of the subscales suggest convergent properties in a multitrait-multimethod sense strengthens the scale as a whole.

Table 4

SED McEachern Frequency Weights

MTMM on Subscales

	Monotrait Hetero Method						Mono Method Hetero Trait							
	Youth with Youth Rating of Parent	Youth with Parent	Youth with Peer	Youth Rating of Parent with Parent	Youth Rating of Parent with Peer	Parent with Peer	Youth		Youth Rating of Parent		Parent		Peer	
							Mean	Range	Mean	Range	Mean	Range	Mean	Range
Property	.83	.62	.74	.76	.75	.82	.56	-.49 to .67	.55	-.50 to .59	.44	-.55 to .73	.57	-.45 to .71
Force	.74	.69	.75	.68	.73	.80	.49	-.41 to .67	.46	-.37 to .59	.56	-.38 to .73	.51	-.37 to .71
School	.83	.76	.70	.80	.80	.86	.56	-.41 to .73	.56	-.41 to .71	.47	-.38 to .55	.55	-.45 to .68

y=youth
yop=youth rating of parent
p=parent
f=peer

Total Scale Intercorrelations

	y	yop	p	f
y	—			
yop	.69	—		
p	.73	.78	—	
f	.67	.68	.82	—

Intersource correlations on the total scale are also included in Table 4. As an aside, it is interesting to note that the youth's rating of the parents response correlated with the parents actual response somewhat higher than the correlation of the youth's response directly with the parent's response. Also of interest is the fact that the friends' responses are more highly correlated with the parent's reponses than with those of the youth. The sources are all fairly highly correlated and the application of weighting schemes and the additional resultant variance within the correlation matrices does not change the relative size or patterns of the correlations.

Relationship with official outcome data. Official outcome data was collected as a normal operation of the overall project. Each youth was involved in the project for a period of 18 weeks. Outcome data consisted of the frequency of official contacts with area police departments or the county court during this 18 week period. Therefore, the six week, one year or ever referents built into the self report instrument do not chronologically match this 18 week outcome data interval. Though these time intervals have advantages during program evaluation, for the purposes of instrument refinement the six week data for the second, third, and fourth waves were summed to correspond exactly with the 18 week duration used to gather the official outcome data.

Though related, two methods of directly assessing the self-report instrument and its relationship to official court and police data are included. Table 5 includes the analyses of variance for each of the ten combined weighting schemes. Success/failure served as the independent variable. Success was defined as the lack of any additional contact with either the police or court during involvement in the project. The total scale and four subscales served as the dependent measures. As can be noted in the table, the resultant F values and probability levels are consistent across the weighting schemes. The same consistent pattern of figures can be seen in the correlations of success/failure with the self-report scales in Table 6.

Table 5

Self-Report Weighting Schemes Balanovas
on Success/Failure on Total and Four Subscales

		Straight Categorical Weights	Straight Frequencies Weights	Sellin & Wolfgang Categorical Weights	Sellin & Wolfgang Frequencies Weights	McEachern Categorical Weights	McEachern Frequencies Weights	Rossi Categorical Weights	Rossi Frequencies Weights	Local Categorical Weights	Local Frequencies Weights
Total	F	8.85	6.88	7.69	6.08	6.56	4.91	8.48	8.69	8.63	6.41
	p	.004	.01	.006	.02	.01	.03	.004	.004	.004	.01
Property	F	13.58	9.16	14.70	10.69	19.52	15.56	17.13	13.53	14.57	10.39
	p	.0005	.003	.0005	.001	.0005	.0005	.0005	.0005	.0005	.002
Force	F	.77	.90	.59	1.25	.25	.21	.81	1.10	.73	.84
	p	.38	.35	.32	.27	.62	.65	.37	.30	.39	.36
School	F	3.76	3.87	3.36	3.87	4.01	5.55	2.21	3.22	2.46	3.19
	p	.06	.05	.07	.05	.05	.02	.14	.08	.12	.08
Substance	F	3.56	3.30	3.47	3.30	4.76	4.68	3.97	2.78	3.47	3.30
	p	.06	.07	.07	.07	.03	.03	.05	.10	.07	.07

Table 6

Self-Report Weighting Schemes Total and Subscales
Correlation with Success/Failure

Year 2,3,4 w 2,3,4 Six Wee
Data Combined for Youth

Scales		Straight Categorical Weights	Straight Frequencies Weights	Sellin & Wolfgang Categorical Weights	Sellin & Wolfgang Frequencies Weights	McEachern Categorical Weights	McEachern Frequencies Weights	Rossi Categorical Weights	Rossi Frequencies Weights	Local Categorical Weights	Local Frequencies Weights
Total	r	.26	.23	.24	.22	.23	.20	.26	.26	.26	.22
	p	.002	.005	.003	.008	.006	.01	.002	.002	.002	.006
Property	r	.32	.27	.33	.28	.37	.34	.35	.32	.33	.28
	p	.001	.002	.001	.001	.001	.001	.001	.001	.001	.001
Force	r	.08	.09	.09	.10	.05	.04	.08	.09	.08	.05
	p	.19	.17	.16	.13	.31	.33	.19	.15	.20	.18
School	r	.17	.18	.16	.18	.18	.21	.13	.15	.14	.16
	p	.03	.03	.04	.03	.02	.01	.07	.04	.06	.04
Substance	r	.17	.16	.17	.16	.19	.19	.18	.15	.17	.16
	p	.03	.04	.03	.04	.02	.02	.02	.05	.03	.04

Though these relationships are all strong with the exception of the physical force scale, the absolute value of the correlation coefficients are not high. Again, this should not be taken in a negative light. There should be a significant relationship between self-reported estimates of illegal behavior and their official counterparts. On the other hand, the official data should in actuality be a subset of the self-reported estimates of actual behavior rates.

In summary, it appears that the self-report instrument designed by the authors has demonstrated consistent reliability both in terms of internal consistency figures and in terms of test-retest correlations. These points are particularly true of the total scale. Convergent validity was demonstrated through multitrait-multimethod procedures. While discriminant properties were not evidenced, the overall scale did demonstrate strong validity properties when related to official court and police data.

Of prime concern is the fact that these properties were demonstrated consistently across the ten weighting scheme combinations. It appears that adding seriousness weights in particular did not contribute appreciably to the differential ability of the instrument to detect official outcome data. The frequency weighting schemes did seem to add some variance to the overall data set when compared to their straight categorical counterparts. Though the increased variance was related to increased alphas in several cases, the differences were not of meaningful proportions nor were they consistent.

In short, the addition of weighting schemes to a self-reported delinquency instrument did not strengthen the instrument itself or its application. On the other hand, the inclusion of weighting schemes did not harm the instrument or interfere with its application. These results did not support conclusions drawn by others suggesting the need for seriousness weights in self-reported delinquency research. In fact, it should be noted that, when applied to these scales, the simple unit weights performed equally as well as the popular and more complex seriousness and frequency weighting schemes.

References

Belson, W. A. The Development of a Procedure for Eliciting Information from Boys About the Nature and Extent of Their Stealing. London: Survey Research Centre, London School of Economics and Political Science, 1968.

Campbell, D. T., & Fiske, D. J. Convergent and discriminant validation by the Multitrait-multimethod Matrix. Psychological Bulletin, 1959, 56, 81-105.

Cloward, R. A., & Ohlin, L. E. Delinquency and Opportunity. Glencoe, IL: Free Press, 1960.

Cohen, A. K. Delinquent Boys. Glencoe, IL: Free Press, 1955.

Davidson, W. S. The diversion of juvenile offenders: A comparison of the process and relative efficacy of behavioral contracting and child advocacy. Unpublished doctoral dissertation, University of Illinois, 1976.

Erickson, M. L., & Empey, L. T. Court records, undetected delinquency, and decision making. Journal of Criminal Law, Criminology and Police Science, 1963, 54, 456-469.

Farrington, D. Self-reports of deviant behavior: Predictive and Stable? Journal of Criminal Law, Criminology, and Police Science, 1973, 64, 99-110.

Glueck, S., & Glueck, E. Unraveling juvenile delinquency. Cambridge, Mass.: Harvard University Press, 1951.

Gold, M. Undetected delinquent behavior. Journal of Research in Crime and Delinquency, 1966, 13, 127-143.

Gold, M. Delinquent behavior in an American city. Belmont, CA: Brooks/Cole, 1970.

Hindelang, M. J., Hirschi, T., & Weis, J. G. HEW Grant Application #MH27778-03, 1977.

Hirschi, T., & Selvin, H. Delinquency research: An appraisal of analytic methods. Journal of Criminal Law and Criminology, 1969, 43, 177-200.

Kantrowitz, R. E., Davidson, W. S., Blakely, C. H., & Kushler, M. G. The effect of training supervision of nonprofessional intervention with delinquents. Paper presented at the 86th annual APA Convention, Toronto, Canada, 1978