Winter 1980

Measuring the Severity of Self-Reported Delinquency: The Development of a Seriousness Scale

Richard P. Kern

William D. Bales

Recommended Citation

MEASURING THE SEVERITY OF SELF-REPORTED DELINQUENCY: THE DEVELOPMENT OF A SERIOUSNESS SCALE*

RICHARD P. KERN** AND WILLIAM D. BALES***

INTRODUCTION

The inherent problems in utilizing official statistics to measure delinquency have been recognized for quite some time.[1] The many criticisms of official statistics have stimulated the development of several new methods of measurement. One of the most popular of these is the self-report technique.[2]

The reliability and validity of self-reports in measuring deviancy remains an open issue with considerable contradictory findings.[3] Although utilization of the method has increased dramatically, it has enjoyed little technical development.[4] A review of the delinquency indices[5] using informa-tion gathered from self-reports reveals a predominance of simplistic and primitive measures of deviant behavior. Many of these indices, although rectifying some of the deficiencies of official statistics, focus primarily upon the variety and frequency of delinquent activity and ignore important qualitative characteristics of deviant acts.[6]

Quantitative measures are quite useful in judging initial distinctions between individuals, but they are incapable of finely discriminating between individuals who are essentially different in the attribute under scrutiny. Thus, in order to arrive at more accurate measurement discriminations in deviant behavior researchers must address the qualitative dimension of offense severity. This article examines empirically the feasibility of developing a measure of offense severity readily applicable to a delinquency index of self-reported deviant behavior.

STATEMENT OF THE PROBLEM

In their pioneering work, *The Measurement of Delinquency*, Sellin and Wolfgang postulated that in measuring the qualitative nature of a criminal act exclusive reliance upon the generic label of an offense (e.g., "murder," "assault") is improper because it contains no relevant data on the degree of actual harm done to society.[7] Because most self-report questionnaires and interview schedules phrase their offense items with an emphasis exclusively upon the generic legal label of an offense, the direct applicability of the Sellin-Wolfgang

---

** Research Associate with the Office of the State Courts Administrator, Florida Supreme Court; Ph. D. Candidate, M.S., School of Criminology, Florida State University; B.A., King's College, Pennsylvania.

*** Research Associate with the Office of the State Courts Administrator, Florida Supreme Court; Ph.D. Candidate, M.S., B.S., School of Criminology, Florida State University.

[1] *See, e.g.,* Beattie, *Criminal Statistics in the United States,* 51 J. CRIM. L.C. & P.S. 49 (1960); Robison, *Critical View of the Uniform Crime Reports,* 64 MICH. L. REV. 1031 (1966); Wolfgang, *Uniform Crime Reports: A Critical Appraisal,* 111 U. PA. L. REV. 708 (1963).

[2] Farrington, *Self-Reports of Deviant Behavior: Predictive and Stable?,* 64 J. CRIM. L. & C. 99 (1973).

[3] W. Bales, Self-Report Measures: Valid and Reliable? (June 1978) (unpublished thesis in Florida State University Library).

[4] The primary concern of this research was to investigate the feasibility of increasing the reliability and validity of the results obtained from self-report measures. Although recognizing their importance, it is not within the scope of this article to address the other methodological issues which relate more specifically to the actual design and administration of self-report inventories. For a further discussion of these latter issues see R. HOODS & R. SPARKS, KEY ISSUES IN CRIMINOLOGY (1970); Farrington, *supra* note 2.

[5] An index measure combines several items or measures of an attribute or behavior. The result is a composite score for each respondent indicating the relative degree to which they possess the characteristic under measure, or engage in the behavior considered.

[6] Kelley & Winslow, *Seriousness of Delinquent Behaviour: An Alternative Perspective,* 10 BRIT. J. CRIM. 124 (1970).

[7] The system they developed for "scoring" criminal events requires that particular information on the crime be gathered (i.e., extent of injury, theft, or damage). Thus, in order for the Sellin-Wolfgang index to be accurately applied within any setting, the data collector must obtain with specificity the information needed to score a criminal event. This need for the congruence between the separate, but related, tasks of measurement—data collection and the scoring process (e.g., assigning of seriousness weights)—is a point which cannot be overly emphasized. T. SELLIN & M. WOLFGANG, THE MEASUREMENT OF DELINQUENCY (1964).

work to such instruments appears limited.[8] Examining the component of offense severity as explicated by Sellin and Wolfgang would require specifically phrasing each offense item to include varied aggravating circumstances. Greater specificity in offense items logically will result in fewer admissions because of the reduced range of behaviors encompassed within a particular offense description.[9] Furthermore, the additional probing by an interviewer for details surrounding criminal acts likely will create respondent suspicion concerning the researcher's intent which in turn may result in less veracity on future items. In cases of longitudinal delinquency studies, which require a time two measure of self-reported delinquency, the respondent's lack of perceived confidentiality decreases the probability of subjects taking part in the latter portions of the study. Finally, the time and effort required to collect such information from large samples of individuals renders employment of detailed offense descriptions impractical. Recognizing that present self-report instruments ignore an important ingredient of measurement,[10] the follow-

ing study seeks to construct a scale of offense seriousness enabling consideration of relative perceived seriousness of generically phrased delinquent acts.

METHODOLOGY

Certainly no objective criteria beyond personal judgment exists for evaluating the relative seriousness of crimes; yet, there are objective techniques of measurement capable of relating two different types of psychological scales which have been applied to such non-physical data as judged seriousness of crimes.[11] The development of the tool used in this study to measure the perceived seriousness of offenses originates in the psychophysical literature.[12]

Consistent with the work of Sellin and Wolfgang, this study utilized two types of scales to measure perceived offense seriousness—category and magnitude scales.[13] The category scale, used frequently in criminological research, has fixed lower and upper limits (e.g., one to eleven) with each division representing equal distance between points.[14] The magnitude estimation technique, on the other hand, contains no established or fixed upper limit and any value greater than zero is deemed a possible response.[15] As Figlio has explained, the category scale involves asking subjects to circle the number from one to eleven (least to most serious) best representing the perceived seriousness of the particular offense. The magnitude scale involves asking the subject to select any number adequately representing the seriousness of that particular offense description. Although easy to understand, the category scale is numerically constraining. The magnitude scale is more abstract than the category

[8] See, e.g., Erickson, The Changing Relationship Between Official and Self-Reported Measures of Delinquency. An Exploratory-Predictive Study, 63 J. CRIM. L.C. & P.S. 388 (1972); Hindelang, With a Little Help From Their Friends: Group Participation in Reported Delinquent Behaviour, 16 BRIT. J. CRIM. 109 (1976). Gold, in particular, found this to be the case when he recognized his inability to directly apply Sellin and Wolfgang's severity weights to offense information gathered in a manner unsuitable to their sophisticated scoring technique. Gold, Undetected Delinquent Behavior, 13 J. RESEARCH CRIME & DELINQUENCY 90 (1966).

[9] The existence of infrequent admissions on particular offense items becomes problematic when data analysis involves an examination of relationships between relevant variables and the incidence of particular delinquent acts. For example, cross-tabulations generated between independent variables and rarely admitted delinquency items are uninterpretable due to the low N sizes contained within some of the individual cells. Thus, the researcher is left with two undesirable options: 1) make cautious inferences and assertions from the results, or 2) confine the analysis to only those offenses which generated sufficient responses. The choice of this latter alternative might confine the inquiry to an analysis of admitted offense behaviors which are of a trivial nature.

[10] Though the desire for a seriousness of offense measure readily applicable to most self-reported delinquency indices has been oftentimes noted, the literature reveals relatively few efforts to attain such an end. Methodological problems, specifically the absence of empirical scaling devices, have served to cast severe doubts upon the reliability and validity of the offense seriousness values derived from these studies. See Christie, Andenaes & Skirbeck, A Study of Self-Reported Crime, in 1 SCANDINAVIAN STUDIES IN CRIMINOLOGY 86 (Karl O. Christianson ed. 1965); Erickson & Empey, Court Records, Undetected Delinquency and Decision-Making, 54 J. CRIM. L.C. & P.S. 467

(1963); Farrington, supra note 2, at 106; Heise, Norms and Individual Patterns in Student Deviancy, 16 SOC. PROB. 78 (1968); Hindelang, supra note 8, at 117-20.

[11] T. SELLIN & M. WOLFGANG, supra note 7, at 237.

[12] See, e.g., Galanter, The Direct Measurement of Utility and Subjective Probability, 75 AM. J. PSYCH. 208 (1962); Helm, Messick & Turner, Psychological Model for Relating Discrimination and Magnitude Estimation Scales, 68 PSYCH. REV. 167 (1961); Stevens, On the Psychophysical Law, 64 PSYCH. REV. 153 (1957); Stevens & Galanter, Ratio Scales and Category Scales for a Dozen Perceptual Continua, 54 J. EXPER. PSYCH. 377 (1957).

[13] T. SELLIN & M. WOLFGANG, supra note 7.

[14] Thurstone, The Method of Paired Comparisons for Social Values, 21 J. ABNORM. SOC. PSYCH. 384 (1927).

[15] The magnitude technique essentially amounts to a continuous ratio scale whereby subjects evaluate the magnitude of stimuli by assigning to such stimuli scores that reflect points on a psychological scale. The categorical technique of assigning scores to a stimulus involves the construction of a rank-ordered continuum of stimuli.

scale but it does not suffer from numerical constraints.[16] In order to use both scaling procedures, two types of questionnaires were constructed. The generically phrased offense items included within these instruments were those frequently found on self-report instruments administered by other researchers to teenagers, with additional items included to provide a continuum of offense types ranging from trivial to most serious. In establishing criteria for inclusion of an offense element within their inquiry, Sellin and Wolfgang selectively omitted consideration of both victimless and public order offenses primarily because of the low frequency with which such acts are reported to law enforcement agencies.[17] The nature of the self-report technique, however, permits to some extent gauging this "dark figure" of crime. Since this study seeks to create severity scale values for offense information collected from self-report questionnaires or interviews, it was felt that victimless and public order offenses should then be considered.

The brief, unambiguous description of each offense focused upon the commission of a deviant act (legal offense label).[18] Aside from a reference to the

offender as a juvenile, the descriptions did not mention any other extraneous elements associated with the act (i.e., offense and victim descriptions). To control for possible unknown response biases such as a queuing effect,[19] offense items were randomized into different rank orderings, which resulted in six questionnaires (three categorical and three magnitude) containing varied offense sequencing.

The selection of judges was guided by prior literature and convenience. Various empirical analyses investigating the utilization of disparate social groupings in rating the seriousness of offenses has yielded results which indicate that considerable consensus exists across such groupings as to what constitutes seriousness.[20] Many authors have considered these findings sound evidence that offense seriousness is a consensual phenomenon.[21] This inquiry used a sample of judges comprised of 172 undergraduate criminology students: $n1 = 85$ (categorical); $n2 = 87$ (magnitude). Although this study's sample selection is vulnerable to criticism on many grounds, previous research sustains the adequacy of the sample for the task to be performed.[22]

ANALYSIS

A principal objective of this research was to establish the feasibility of applying scaling devices, validated and successfully applied in many other research settings, to generically phrased criminal

---

[16] Figlio, *The Seriousness of Offenses, An Evaluation by Offenders and Nonoffenders*, 66 J. CRIM. L. & C. 189 (1975). Within the magnitude technique there are an infinite amount of discrimination points available for use in judging a phenomenon, with each judgment being a ratio of some standard item positioned at the outset. Any item may be assigned the position of standard item (i.e., "modulus") as well as any number since it is used merely as a base. *See* T. SELLIN & M. WOLFGANG, *supra* note 7, at 248.

[17] T. SELLIN & M. WOLFGANG, CONSTRUCTING AN INDEX OF DELINQUENCY: A MANUAL 67 (1963).

[18] Consistent with a belief in the necessary congruence of the manner in which the offense information is gathered and then scaled for offense severity, the offense descriptions presented to the scaling judges must be phrased in a generic fashion identical to the manner in which they are phrased within the self-report schedule. The offense descriptions presented to the respondents involved in this research effort include: "the offender (a juvenile) . . . 1) drinks liquor while under age, 2) steals or shoplifts something worth $100 to $1000, 3) damages someone else's property, 4) cheats in school, 5) takes part in an armed robbery, 6) uses drugs other than marijuana, 7) steals or shoplifts something worth more than $1000, 8) drives an automobile while under the influence of alcohol or drugs, 9) takes part in a rape, 10) smokes marijuana or hashish, 11) assaults someone with a weapon, 12) sells drugs, 13) runs away from home, 14) steals or shoplifts something worth $10 to $100, 15) commits a murder, 16) steals or shoplifts something worth less than $10, 17) breaks into a locked building (other than his/her own home), 18) takes part in an unarmed robbery, 19) skips school, 20) takes or uses someone else's car without permission (this offense was chosen as the modulus and assigned a value of 10 for those using the magnitude technique).

[19] A queuing effect is the possibility of bias that originates from a particular sequential ordering of questionnaire/interview items or events. The relative perceived seriousness evaluations of an offense item may be biased by a directly preceding offense description. For instance, seriousness ratings of the item "damage property" may vary depending upon whether it was immediately preceded by the offense item involving "murder" as opposed to the offense item involving "smoking marijuana."

[20] T. SELLIN & M. WOLFGANG, *supra* note 7; Figlio, *supra* note 16; Kelly & Winslow, *supra* note 6; Thomas, Cage & Foster, *Public Opinion on Criminal Law and Legal Sanctions: An Examination of Two Conceptual Models*, 67 J. CRIM. L. & C. 110 (1976).

[21] Lesieur and Lehman are two authors who provide opposite views on this matter. Lesieur & Lehman, *Remeasuring Delinquency: A Replication and Critique*, 15 BRIT. J. CRIM. 69 (1975).

[22] Criticism may be leveled at the choice of criminology students for a sample. Yet, as Wiatrowski has noted, these students would reflect in their estimations an objectivity which would not be obtained from other students who, because of their academic orientations, would not appreciate the scholarly intent of the study. M. Wiatrowski, The Measurement of Delinquency: A Replication Study and Test of the Assumption of Additivity 37 (unpublished thesis in Florida State University Library 1974).

TABLE I

MEAN RANK VALUES AND RANK ORDERING OF OFFENSE SERIOUSNESS EVALUATIONS (N = 162)

Rank	Offense Item	Mean Rank of Categorical Seriousness Values (n_1 = 81)	Rank	Offense Item	Mean Rank of Magnitude Seriousness Values (n_2 = 81)
1.	Murder	18.88	1.	Murder	19.63
2.	Rape	18.16	2.	Rape	18.41
3.	Armed Robbery	16.75	3.	Armed Robbery	16.83
4.	Assault/Weapon	16.33	4.	Assault/Weapon	16.44
5.	Larceny (>$1000)	14.35	5.	Unarmed Robbery	13.87
6.	Unarmed Robbery	13.26	6.	Larceny (>$1000)	13.70
7.	Sell Drugs	12.54	7.	Sell Drugs	12.93
8.	Other Drug Use	11.30	8.	Breaking & Entering	11.90
9.	Breaking & Entering	11.26	9.	Drive/Influence	11.10
10.	Damage Property	11.07	10.	Damage Property	10.43
11.	Drive/Influence	10.83	11.	Larceny ($100-$1000)	10.41
12.	Larceny ($100-$1000)	10.54	12.	Larceny ($10-$100)	10.25
13.	Larceny ($10-$100)	9.69	13.	Other Drug Use	9.69
14.	Steal Car	9.08	14.	Steal Car	8.22
15.	Larceny (<$10)	5.15	15.	Larceny (<$10)	6.68
16.	Marijuana Use	4.79	16.	Marijuana Use	4.58
17.	Cheat in School	4.70	17.	Runaway	4.18
18.	Runaway	4.51	18.	Cheat in School	4.08
19.	Drink Liquor	4.19	19.	Drink Liquor	3.98
20.	Skip School	2.62	20.	Skip School	2.70

Kendall's Coefficient of Concordance (W) = .727   Kendall's Coefficient of Concordance (W) = .761

Chi Square = 1119.13 (p < .001)   Chi Square = 1170.66 (p < .001)

offense items.[23] It was therefore necessary initially to assess the internal consistency of the sample's responses. To this end, cases containing missing data were eliminated as were cases in which it was clear that respondents had misunderstood the task (e.g., rated *all* offenses with values less than the standard score on the magnitude instruments). Likewise discarded were cases giving higher values to a larceny item with a lower dollar amount than to a larceny item with a higher dollar amount.[24]

The degree of internal consistency was measured by using Kendall's coefficient of concordance (W). Table I reveals that a significant W statistic is generated for both category and magnitude scaling techniques. These findings indicate the judges applied similar standards in the rank ordering of the offenses, and thus provide sound evidence for the internal consistency of the two scales.[25]

[23] Stevens, *A Metric for the Social Consensus*, 151 SCI. 530 (1966).

[24] These two initial reliability checks resulted in the elimination of ten respondents from the original sample.

[25] Sellin and Wolfgang plotted and compared scores of sample sub-groups to determine such consistency. This technique, under the circumstances, does not indicate the desired information. Sellin and Wolfgang's efforts to test for agreement *between* groups does not give any indication

An empirical analysis was then conducted to demonstrate both the nature of the measured phenomenon and the success of applying the scaling technique. Each of these issues was investigated with reference to a general psychophysical law relating subjective magnitude to stimulus magnitude. As Stevens stated, "the law is simply that equal stimulus ratios produce equal subjective ratios."[26] Numerous experimental findings support this law but limit its application to continua termed "prothetic."[27] This type of continua includes such phenomena as heaviness, loudness, and brightness. Discrimination between different intensities or levels of these phenomena is based on an additive process in which "excitation is added to excitation at the physiological·level."[28] In contrast,

whether there is agreement *within* each group of judges. Given that one does find consistency in ratings within groups of judges, there currently exist only exploratory and unestablished statistical techniques which can effectively test for evidence of rater agreement between groups of judges. T. SELLIN & M. WOLFGANG, *supra* note 7, at 268; Hollander & Sethuraman, Testing Agreement Between Two Groups of Judges (Air Force Office of Scientific Research, Technical Report No. 2 1977).

[26] Stevens, *supra* note 12, at 153.

[27] *Id.*

[28] Stevens & Galanter, *supra* note 12, at 377.

a substitutive process at the physiological level controls sensory discrimination in phenomena included in so-called "metathetic" continua, such as pitch or position. Stevens and Galanter note that "the distinction between class I (prothetic) and class II (metathetic) is something like the traditional distinction between sensory intensity and sensory quality . . . quantity vs. kind."[29]

The results of both the categorical and magnitude methods were directly compared in order to determine the type of continua (metathetic or prothetic) under study in this instance. Plotting the mean categorical scores against the geometric mean[30] magnitude values obtained from an evaluation of the severity of generically phrased offense items yields a concave downward trend which Stevens suggests is typical of prothetic continua

---

[29] *Id.* at 378.

[30] The geometric mean has been found to be an optimal method of averaging the magnitude scores since, despite the various numerical ranges utilized by each respondent, no normalizing of the data is needed. The geometric mean is equal to the nth root of the product of the scores on the magnitude scale;

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

This measure of central tendency, commonly used for ratio measures, can more easily be computed by summing the base 10 logarithms of each magnitude score, dividing by the number of scores, and then taking the base 10 antilog of this arithmetic mean. This formula is written as:

$$GM = Antilog \left[\frac{(\log x_1 + \log x_2 + \cdots \log x_n)}{n} \right]$$

There has been some confusion in the literature reporting attempts to replicate the study conducted by Sellin and Wolfgang in terms of how to compute the geometric mean of the magnitude scores, which partially explains why some efforts have led to results inconsistent with their findings. For example, Hsu claims the formula used to compute the geometric mean is: $GM = nX_1X_2X \ldots X_n$ where X represents the scores given by the raters. Not surprisingly, her mean magnitude scores differ by as much as 390 points (for murder) from those computed by Sellin and Wolfgang. Wiatrowski appears to have made a computational error in deriving a geometric mean value for each offense rated by his respondents. He computed the arithmetic mean of the magnitude measure for each offense description, determined the base 10 logarithm of each value, and then derived the geometric mean by taking the antilog of each logarithm value. The problem lies in the fact that he failed to take the logarithm of each score, compute an arithmetic mean, then compute the geometric mean by deriving the antilog of this value. Hsu, *Cultural and Sexual Differences in the Judgement of Criminal Offenses: A Replication Study of the Measurement of Delinquency*, 64 J. CRIM. L. & C. 348 (1973); Stevens, *supra* note 23, at 531; M. Wiatrowski, *supra* note 22, at 51–52.

(see Figure 1). Stevens writes: "on prothetic continua the partition scale (categorical) is practically always non-linear relative to the magnitude scale . . . the curvature of the category scale is usually intermediate between linear and logarithmic."[31]

Stevens notes that the true relationship between scales on prothetic continua as measured by both direct magnitude estimation and Thurstone's categorical technique is essentially logarithmic. Thus, "this relation provides a test that can be applied to nonmetric continua, such as seriousness of crimes, in order to determine the nature of the continua."[32] Should such a test verify the logarithmic relationship between scales created from nonmetric stimuli, added confidence attaches to the validity of the measurement outcome.

Figure 2 examines the relationship between these two scales when transformed into semilogarithmic coordinates. As Figure 2 demonstrates, this transformation effectively linearizes the relationship between the scales, and thereby confirms their logarithmic relationship. Similar findings led Sellin and Wolfgang to conclude that crime severity of offense "events" could be successfully measured psychophysically.[33]

Despite forewarnings that the employment of unspecific, generic crime descriptions for scaling offense severity would foster variable perceptions within each respondent (a factor which purportedly renders such a scaling effort futile), these findings reveal the operation of perceptual discrimination processes required of subjects in studies similar to Sellin and Wolfgang's.[34] Specifically, the relationship between the categorical and magnitude scales reveal that the manner in which respondents subjectively discriminated between the stimuli presented (generic offense items) is similar to the psychophysical process of evaluating different levels of intensities in physical objects such as heaviness. Thus, the results of this inquiry extend the applicability of psychophysical scaling properties to offense severity reflected in generic offense labels. The importance of this finding is that using crime seriousness values derived from this proce-

---

[31] Stevens, *supra* note 23, at 532.

[32] *Id.*

[33] T. SELLIN & M. WOLFGANG, *supra* note 7, at 261–62.

[34] These findings are contrary to those reported by Silvey who found a lack of consistency as to the relative offense seriousness evaluations when such behaviors were phrased in a general fashion. The study conducted by Silvey is reported in Rose, *Concerning the Measurement of Delinquency*, 6 BRIT. J. CRIM. 415 (1966).
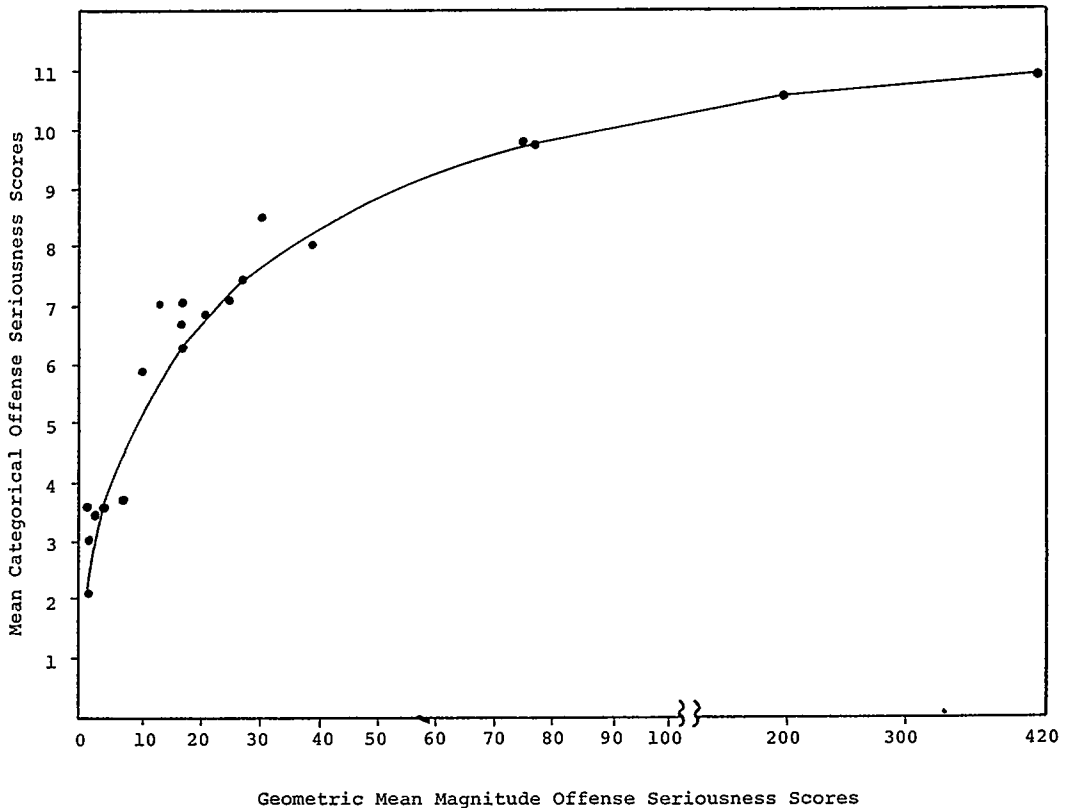
Geometric Mean Magnitude Offense Seriousness Scores

FIGURE 1 MEAN GEOMETRIC MAGNITUDE OFFENSE SERIOUSNESS SCORES PLOTTED AGAINST CORRESPONDING MEAN CATEGORICAL OFFENSE SERIOUSNESS SCORES

dure in self-reported delinquency measures will enhance the reliability of the results.

DISCUSSION

Table II presents both the mean categorical scores and geometric mean magnitude scores for the twenty generically phrased offense items. Despite the direct logarithmic relationship between these two sets of score values, it would appear that the magnitude scale values are the more appropriate offense severity scores for inclusion in a delinquency index.[35] This technique of deriving final offense seriousness scores differs from the procedure employed by Sellin and Wolfgang in the final version of their scale. These researchers used originally computed geometric mean magnitude values to derive an additive algorithm which provided

final scale values. Yet, as Turner has recently commented, "the most impressive part of Sellin and Wolfgang's work is the scoring derived from the basic silhouettes themselves and *not* the particular additive version that they impose on it."[36] Furthermore, Sellin and Wolfgang employed certain offense silhouettes specifying particular dollar amounts stolen (e.g., $5, $20, $100) to derive a power function for the amount of money stolen. They derived the score values for offenses involving monetary loss from this power function.[37]

It was inappropriate to develop a power function for money values in this study because no specific money values existed within the generic offense descriptions presented to the respondents. Consistent with the aforementioned belief that specificity in offense phrasing in self-report surveys may precipitate problems, the offense descriptions used in this inquiry contained phrases representing property values within monetary ranges such as

[35] This choice is justified because the scale values obtained from the magnitude technique reflect a product of the rater, rather than the experimenter (as is the case in the categorical method). Also, the magnitude technique provides raters with the availability of a greater latitude in numerical range with which to fix responses. *See* T. SELLIN & M. WOLFGANG, *supra* note 7, at 271-73.

[36] Turner, in T. SELLIN & M. WOLFGANG, THE MEASUREMENT OF DELINQUENCY ix, (2d ed. 1978).
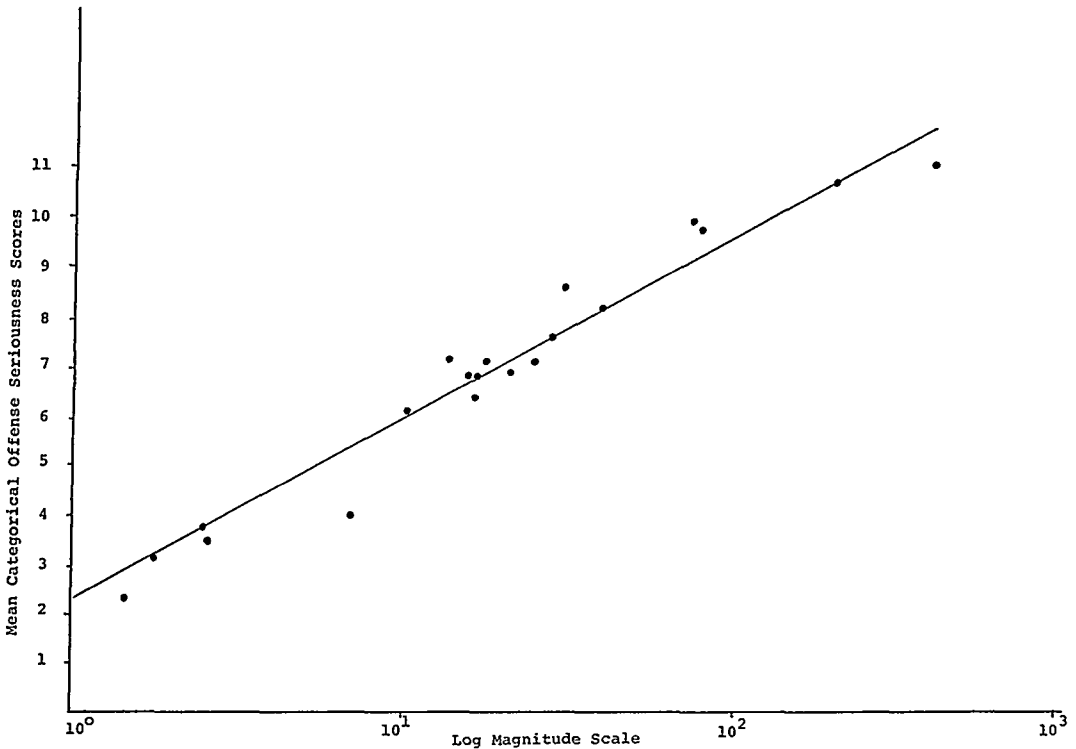[37] T. SELLIN & M. WOLFGANG, *supra* note 7, at 284-87.

FIGURE 2 PLOT OF MEAN CATEGORICAL VALUES WITH LOG MAGNITUDE SCALE

ten to one hundred dollars. As a consequence of this technique, it was impossible to derive a power function for monetary values.[38]

Initial inspection of the final geometric mean magnitude values presented in Table II provides evidence of the "face validity" of these findings.[39] For instance, subjects perceived all the violent personal crimes (murder, rape, etc.) as more serious than all property, public order, or victimless offenses. Of interest is the fact that weapon usage doubles the perceived seriousness rating of the offense, robbery.

As expected, the "status" offenses evaluated

[38] *Id.*

[39] For mathematical simplicity, Sellin and Wolfgang divided all derived values by a lowest score to arrive at final scale values. Such a procedure was deemed unnecessary herein since it would result in almost equivalent values to the original seriousness scale scores. This would occur because the divisor, the value for marijuana, is 1.03. Additionally, rather than apply an arbitrary rounding rule to the final scale values, we have allowed, as Rose suggests, the score values to take decimals. This procedure could possibly increase the sensitivity of any index created from such values. *See* Rose, *The Merits of an Index of Crime of the Kind Devised by Sellin and Wolfgang,* 7 THE INDEX OF CRIME—SOME FURTHER STUDIES—COLLECTED STUDIES IN CRIMINOLOGY 44 (Council of Europe 1970),

TABLE II

MEAN CATEGORICAL AND GEOMETRIC MEAN MAGNITUDE SERIOUSNESS VALUES FOR ALL OFFENSE ITEMS

(N = 162)

Offense Item	Mean Categorical Seriousness Values ($n_1 = 81$)	Geometric Mean Magnitude Seriousness Values ($n_2 = 81$)
Skip School	2.19	1.36
Cheat in School	3.54	2.43
Runaway	3.42	2.52
Larceny (<$10)	3.85	6.76
Larceny ($10–$100)	6.25	16.41
Larceny ($100–$1000)	6.73	16.56
Larceny (>$1000)	8.53	30.55
Liquor Use	3.03	1.75
Marijuana Use	3.58	1.03
Other Drug Use	7.02	13.58
Sell Drugs	7.54	27.67
Drive Under Influence	6.82	20.56
Steal a Car	5.88	10.00
Damage Property	7.03	17.91
Breaking & Entering	7.04	24.49
Unarmed Robbery	8.12	38.99
Armed Robbery	9.78	75.51
Assault with a Weapon	9.68	76.56
Rape	10.56	195.88
Murder	10.95	412.10

within the survey instrument are all perceived as being of minor relative importance. Examination of these scale values also indicates that marijuana use is perceived as being the least serious act. Given the increased usage and acceptance of marijuana, the finding of its perceived minor severity is no surprise. Marijuana use, however, is perceived as even less serious than delinquent acts such as skipping school, cheating in school, and running away from home. This result may be a function of the relatively high degree of exposure to, or use of, marijuana by the particular sample respondents utilized.

Although generally the scale values of offenses involving monetary loss exhibit a discrimination between severity of acts based upon the monetary amounts involved, subjects failed to finely discriminate between larceny "$10 to $100" and larceny "$100 to $1000." The latter item's lack of a relatively specific monetary frame of reference may account for this lack of discrimination. In contrast to items such as larceny "less than $10" and larceny "$10 to $100," the item larceny of "$100 to $1000" would seem to provide a greater opportunity for respondents to create an arbitrary reference point when evaluating its perceived severity.

Finally, results for status offenses of skipping school and running away from home mirror those of Galvin who found the latter offense to be perceived as more serious.[40] This finding initially would appear inconsequential considering the low relative scale values of these acts; however, since they are among those acts most often admitted on self-report inventories, their relative weights may accrue added significance in the construction of a delinquency index.

The derived seriousness scale (magnitude) can now be applied to a self-reported delinquency index in any number of ways. Historically, the admissions on a number of self-report items have been combined by several means to obtain index scores reflecting each respondent's involvement in illegal behavior. Summing the number of different offenses admitted, regardless of the number of times the behavior occurred, can indicate the variety of delinquent acts committed. A serious problem with such a variety index is the inability to differentiate individuals who only admit to trivial behaviors from those who only admit to serious behaviors. The application of the derived seriousness weights to each admitted offense would en-

hance the discriminatory capability of such a delinquency index (i.e., a "variety-serious" index).

If the data collection instrument allows, an index comprised of the total number of admitted behaviors can be computed by summing the frequency of admissions over all items. The existence of an extremely large number of admissions to very minor violations, however, undermines the reliability of this technique by creating misleading variations in derived index scores.[41] Such a "variety-frequency" delinquency index would assign a higher status of delinquency to individuals involved in a large number of trivial behaviors than to those youths admitting only more serious acts but with less persistence. The application of an appropriate seriousness weight to each admitted offense would enhance the reliability of this index measure by providing the discrimination necessary to distinguish among individuals possessing conceptually distinct delinquent pasts.

The problematic effects of extreme distortions in index scores also can be approached by collapsing the frequencies of each offense item into a quartile distribution. A value (ranging from 0 to 4) indicative of one's relative involvement in a delinquent activity would be assigned to each admitted act based upon the quartile which contains their frequency count for a given item. Therefore, the value of zero would indicate no admissions, while a score of one would be assigned to individuals whose involvement in a delinquent activity places them in the lower twenty-fifth percentile of the total distribution of their peers' involvement in the behavior. Under this scoring scheme, the seriousness weight of each admitted act could be multiplied by the appropriate frequency quartile value, thereby providing this measure with the capacity

---

[40] Galvin, *The Seriousness of Offenses: An Evaluation of Children and Adolescents*, in QUANTITATIVE STUDIES IN CRIMINOLOGY 48 (Wellford ed. 1978).

[41] Extreme variations in index scores can be attributed to the use of a recall period (on the self-report instrument) so broad that respondents seriously overestimate their involvement in particular delinquencies. To address this difficulty, a researcher can administer a self-report inventory designed to gather information only on activities from one's recent past (e.g., the last six months). The temporal specificity provided by a limited recall period would avoid huge distortions in index scores as well as increase the relative accuracy of the derived measure. R. HOOD & R. SPARKS, *supra* note 4, at 69; Farrington, *supra* note 2. It might also be argued that variable recall periods be provided on a self-report inventory on the assumption that a respondent's memory would be more accurate with regard to more serious offenses. Therefore, a recall period ranging from three to six months might be used for offenses of a less serious nature while a longer period of one year could be employed for the acts perceived to be of a more serious nature. The seriousness scale presented herein would aid one in the design of such an instrument.

to make qualitative distinctions among delinquent youth.

In sum, there are a number of ways to incorporate the derived seriousness weights into a numerical index to account for variations in the qualitative nature of self-reported delinquency.

SUMMARY & CONCLUSION

Measurement problems continue to plague criminologists in their pursuit of performing sound methodological inquiries. The self-report inventory provides a viable alternative to official statistics for measuring delinquency. Yet, because the self-report technique frequently fails to collect information pertaining to qualitative dimensions of the deviant act, delinquency indices created from such data are often inaccurate.

Because of the specificity it requires, the research effort of Sellin and Wolfgang,[42] designed to gauge qualitative aspects of offense severity, is almost useless to the researcher confronted with self-report data collected from generically phrased offense items.

This study adopted scaling techniques developed within the field of psychophysics and applied them to the evaluation of relative perceived seriousness of generically phrased offense items. The study established empirically that the offense severity scale values were obtained by discrimination processes similar to those occurring in sensory psychophysics. Such an outcome would appear to highlight the validity of the scale results. Because the derived severity scale values are readily applicable to data collected from most self-report instruments, this study's findings extend to the very core of delinquency research.

Several areas of concern for the direction of future research deserve close scrutiny in developing more valid and reliable self-reported delinquency indices. One particular area deserving attention is the selection of offense items for inclusion within the severity scale. Along these lines, efforts can be made to refine the monetary values presented within particular offense descriptions. The present study underscores the importance of monetary range selections. A range of money values too large may create ambiguity and inhibit the discriminatory capabilities of judges. A monetary range too narrow causes the problems that specificity in described offense acts generates with self-report data.

Further, a considerable amount of literature discusses the type of sample to be utilized in the task of judging offense severity.[43] The choice of judges employed within this inquiry was premised more on practical than theoretical concerns. Where feasible, future research should test the results of this study with varied samples reflecting different cross-sections of the populace. A suggested methodological innovation would be using judges who both respond to a self-report inventory and evaluate the relative seriousness of all acts appearing on the instrument. This procedure would enable empirical determination of whether a respondent's perceived seriousness evaluations of particular items are influenced by his admitted involvement in the act.[44]

Finally, as previously noted, there are many ways to apply these severity scores within an index of delinquent behavior. The qualitative attribute offered within the offense severity scores may present no better a measurement alternative than a simple variety count of admitted offenses. Researchers should strive for the technique best measuring delinquency and should seek to develop procedures for judging the overall measurement adequacies of delinquency indices.

These efforts may proceed by investigating the self-reported delinquency of various groups of youths who can be assumed to differ both quantitatively and qualitatively in their delinquent behavior (e.g., institutionalized youth, community-diverted youth, and high school youth). After constructing several indices of self-reported delinquent behavior (taking into consideration the many combinations of variety, frequency, and severity), an analytic tool, such as discriminant function analysis, may be applied to the data in order to statistically gauge the ability of each measurement index to correctly classify the affiliation of group members. Presumably, the index which most accurately predicts group membership would be most valid. One may recognize this technique of index choice as a form of concurrent validity ("known-group validation").

---

[43] Wolfgang, *On Devising a Crime Index*, 7 THE INDEX OF CRIME—SOME FURTHER STUDIES—COLLECTED STUDIES IN CRIMINOLOGY 55 (Council of Europe 1970).

[44] Hepburn adopted such a procedure in an attempt to gauge the relative seriousness of self-reported delinquency; however, he made no effort to determine the validity or reliability of the derived seriousness scale scores. In addition, no assessment was made regarding the extent of possible variance in seriousness evaluations which might be attributed to the respondent's relative involvement in past delinquency. Hepburn, *Testing Alternative Models of Delinquency Causation*, 67 J. CRIM. L. & C. 450 (1976).

[42] T. SELLIN & M. WOLFGANG, *supra* note 7.