

Summer 1979

Scale of Sentence Severity

Deborah Buchner

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

Deborah Buchner, Scale of Sentence Severity, 70 J. Crim. L. & Criminology 182 (1979)

This Criminology is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in Journal of Criminal Law and Criminology by an authorized editor of Northwestern University School of Law Scholarly Commons.

SCALE OF SENTENCE SEVERITY

DEBORAH BUCHNER*

INTRODUCTION

The purpose of this study was to create an interval scale measuring the comparative severity of types of criminal sentences. It is assumed that the scale could be used as a methodological tool for studies in which comparisons are made by the degree of severity of the sentences. The problem addressed in the present study is that sentences contain disparate elements which cannot easily be compared and this in turn raises several important questions. For example, is ten years probation less severe than one month in jail? Or do the possible sorts of restrictions involved in probation or the character of the prison affect the severity of the sentence?

This study utilized Thurstone's Law of Comparative Judgments¹ to obtain empirically the values on a scale of severity for a sample of sentences. Regressions were run on the empirically determined values to obtain estimates of the importance to severity of the theoretically defined components of the sentence. By analyzing the effects on severity of various elements found in sentences, any sentence can be placed on an interval scale.

No successful attempt to place sentences on an interval scale has been made thus far. This situation is remarkable in view of the fact that no sophisticated or reliable comparisons based on severity of sentence can be made without the use of such a tool.

*Research Associate with the Institute for Law and Social Research, Washington, D.C.; B.A. Sociology, University of Wisconsin, 1972; M.A. Criminology, University of Pennsylvania, 1973; Ph.D. Sociology, University of Pennsylvania, 1978.

This research was conducted as part of a dissertation entitled "Effects of Prosecutorial Decisionmaking on Sentence Severity." The project was supported by grant number 77-NI-99-0027, awarded by the Law Enforcement Assistance Administration. The author wishes to thank Dr. Marvin E. Wolfgang and Dr. Robert Figlio of the University of Pennsylvania and Dr. Peter Murrell of the University of Maryland for their help, and especially to thank the judges who generously gave of their time and experience.

¹ A. EDWARDS, *TECHNIQUES OF ATTITUDE SCALE CONSTRUCTION*, ch. 2 (1957); W. TORGERSON, *THEORY AND METHODS OF SCALING*, chs. 8, 9 (1958).

One attempt by Lehtinen and Smith² adopted an ordinal scale constructed by a one-time chief of the California Bureau of Criminal Statistics. There were no claims by the authors that this ordering had any basis in reality. To have any validity, the ordering of severity should be rooted in either a consensus of community feeling, in the perceptions of those who sentence or in the perceptions of those who are sentenced; otherwise it is an arbitrary and therefore suspect ordering. The authors took this unsubstantiated ordering and treated it as a ratio scale with an absolute zero point.³ Lehtinen and Smith made no attempt to discover actual (true) distances between sentences on the continuum; therefore, one can judge that the sentences have been arbitrarily placed.

Ultimately, the authors' "technique" resulted in a coincidence on the scale of very different sentences that intuitively seems incorrect. For example, a sentence of ten years to life in prison was equal to about five years in jail.⁴ Thirty years on probation was equated to fifteen years in prison. The criticism, however, is not that the placing of the sentences on the scale does not make intuitive sense, but rather that there is no justification for the placement.

The present study has attempted to avoid making unwarranted assumptions like those existing in the Lehtinen and Smith study. Criminal court judges were interviewed to order their perceptions of the degree of severity of a sample of sentences. The sentences were generally of a familiar type to the judges. The aim was to discover which was considered a more serious and which a less serious sentence. By averaging the judges' perceptions, calculations were made to place sentences on a scale of severity.

² Lehtinen & Smith, *The Relative Effectiveness of Public Defenders and Private Attorneys*, 32 NAT'L LEGAL AID & DEFENDERS' ASS'N BRIEFCASE 13 (1974).

³ *Id.* at 15. This is a questionable practice because this assumption has no empirical basis. The authors acknowledged that this practice adds a "curious distortion" which must be guarded against in interpreting the study's results. *Id.*

⁴ By "jail" it is assumed that the authors mean a county institution, and that "prison" refers to a state institution.

The judges' collective responses provided an ordering for a sample of sentences so that each sentence would have a fixed point on a continuum according to its severity. Thus, the relation of each sentence to all others would be known. The scale reflects the perceptions of judges in a particular community at a particular time. For this reason, the concept of severity was not defined for the judges. Rather, it was hoped that their intuitive understanding would provide a more accurate reflection of their actions.

As a pilot study, the present research provides an example that a scale measuring the severity of sentences can be created. However, a larger research project employing more stimuli and a greater number of comparisons would provide greater accuracy and a greater number of scale scores from which to form the scale. Nevertheless, we do not know whether, like attitudes about crimes, attitudes about the severity of sentences can be generalized. A larger study would have to reflect geographic as well as demographic differences.

THEORETICAL FACTORS

In order to place sentences on meaningful points on a continuum, specific theoretical questions about the sentences have to be answered first. The present theoretical question concerns the effect on the degree of severity of the various components that make up the sentence. The question is whether factors other than the number of years add to the severity of the sentence. Specifically, does mandatory treatment attached to probation result in a more severe sentence than regular probation? What is the relation of the minimum and the maximum to sentence severity? Does the type of prison, in this case the county prison or the state correctional institution, add to the severity of the sentence? Finally, is it correct to assume that all prison sentences are more severe sentences than all probation sentences?

Thurstone's Law of Comparative Judgments was employed as the technique to scale a sample of sentences which, in turn, would provide answers to the theoretical questions posited. Once these theoretical questions are answered, any sentence can be placed on the scale.

SCHEDULE CONSTRUCTION

In selecting a sample of sentences that reflected sentencing patterns under study, examples of some of the most frequent sentences given during a

specific time period were chosen and listed. The aim was to locate these sentences, which encompass the above theoretical concerns, on a severity continuum according to the degree of severity each possessed.

Theoretically, the technique requires the comparison of each stimulus (sentence) with all others in the sample. However, this process would lead to an inordinate number of comparisons

$$\left(\frac{n(n-1)}{2} \right)$$

in an experimental context where n is the number of stimuli. Methods of reducing experimental labor have been derived⁵ and the following one was used in the present study. The stimuli were divided into three subgroups having similar sentences with respect to the components of the sentence. This procedure meant that the relative severity of the three subgroups fell on three distinct parts on the continuum. Comparisons of all possible pairs were set up internally within each subgroup.

The groups of stimuli were arranged such that it could be assumed that all the stimuli in one group would be more severe than in another. However, to test this assumption, comparisons that linked the groups had to be included. The assumed least severe sentence in one group was paired with the sentence assumed to be the most severe in the group theoretically below it on the continuum. In this way, all three groups were given an assumed ordering on the severity continuum.

The following sixteen sentences were chosen as the stimuli and were grouped as shown. Each sentence was paired with all others in its group and linking pairs were formed between groups:⁶

Group I	3½ years probation
	3 years medical probation
	4 years probation
	* 9 years medical probation
	* 10 years probation
	* 1-2 months county prison
Group II	6-12 months county prison
	* 3-23 months county prison
	3-20 months state correctional institution
	** 3-23 months state correctional institution

⁵ W. TORGERSON, *supra* note 1, at 191-92.

⁶ The sentences in Groups I-III preceded by an asterisk (*) were not only compared to all other sentences in their group, but also to the sentences with the same number of asterisks in the adjacent group.

		4-22 months state correctional institution
Group III	**	1-10 years state correctional institution
	**	2-4 years state correctional institution
		2-10 years state correctional institution
		3-9 years state correctional institution
		4-8 years state correctional institution

The pairs created from the above three groups produced the comparisons needed to allow ordering. The forty pairs were listed and then given a random ordering on the questionnaire. The order within a pair was also randomly assigned. The purpose of randomization was to avoid biasing patterns.⁷ Each judge was asked to make comparisons for all forty pairs.

DATA COLLECTION

The judges of the Common Pleas Court of a large metropolitan city were asked to fill out the schedule. The Common Pleas Court has jurisdiction over cases for which a penalty of five years or more in prison has been legislated. It also encompasses several subdivisions such as the Trial Division, Family Court and Orphans Court. The judges rotate among these divisions so that any judge who has had Trial Division work within the past five years was included.

Of the total number of judges from the Common Pleas Court, eighty percent were available for interview purposes. Availability was defined as those judges who were not ill or on vacation during the period of data collection and those who have had criminal trial court experience. Of the judges who were available, eighty-three per cent completed the schedule and the interview. Two schedules were discarded because they were incomplete or improperly done. The sample was composed of fifty-eight judges.

The judges were asked to choose in each of the paired comparisons the sentence they perceived to be the less severe of the two. When no difference was perceived, the judge had the option of circling "same." This was done to give the judges a greater number of options. It was felt that it was better not to force the judges to make a choice between two options they believed to be the same, thereby reducing the negative reactions to the schedule itself.

Answers that were circled "same" were later assigned randomly to either part of the pair. The decision to assign the "same" answers randomly was based on the assumption that if they truly are the same and the judges were forced to choose between them, then the answer with respect to which sentence was less severe would have been found to be randomly assigned by the judges themselves.⁸

ANALYSIS

The purpose of the analysis⁹ was to create one continuous scale so that each sentence could be compared to all others. To achieve this, the methodology required two stages. First, it was necessary to discover empirically the scale values on a scale of sentence severity of a sample of sentences. Second, regressions were run on these scale values to discover the relative importance of the theoretically defined components of the sentence to the severity of any sentence.

The sentence comparisons of the judges provided the data for the first step. The paired comparisons were analyzed using Thurstone's Law of Comparative Judgments,¹⁰ a scaling technique designed to produce interval level comparison between discrete factors. The analysis resulted in a set of "scale scores" or "scale values." Thus, each of the sentences in the sample holds a unique position on the scale.

The analysis produced a scale score for each sentence stimulus. However, an insufficient number of comparisons made it impossible to learn the relative positions of two subsets of sentences. The result was two separate scales comprising two different segments on the same sentence severity continuum. The problem was overcome in the regression analysis in stage two.

The second part of the analysis entailed the calculation of the values of the various weights for the different components of a sentence. Regressions were run to obtain coefficient weights for the important elements of the sentence. The beta weights of the regression equation provided the weighting factors against which the numerical values of components of the sentence were multiplied.

The sentences that composed the two separate scales together contained all the theoretically de-

⁸ *Id.*

⁹ The following discussion of the analysis is in summary form. For a more complete explanation of the statistical manipulations, please refer to the Discussion section, *infra*.

¹⁰ See note 1 and accompanying text *supra*.

⁷ W. TORGERSON, *supra* note 1, at 168.

finer components that might be included in a sentence. Thus, when separate regressions were run on the two scales, weights for all the components were obtained.

However, because the sentence components are part of two separate scales, the relationship between the two subsets of components was not known. A technique was employed to combine the two scales so as to be able to calculate the relationship between the two scales and thus the coefficient weights.

DISCUSSION

Explaining the study in greater detail, computer analysis of the data was calculated according to Thurstone's Case 5 Scale Score Program.¹¹ The computations carried out by the program were those described by Edwards for Thurstone's Law of Comparative Judgments in Case 5 for incomplete data.¹² The Case 5 program produced "scale scores" or "scale values" for the sixteen original stimuli (sentences). The scale values were as follows:

<i>Stimuli</i>	<i>Scale Scores</i>
3 years medical probation	-0.1232
3 1/2 years probation	0.0
4 years probation	0.0424
1-2 months county prison	0.3866
3-23 months county prison	1.1775
9 years medical probation	1.4162
10 years probation	1.4936
3-20 months state correctional institution (SCI)	1.9354
4-22 months SCI	2.4652
3-23 months SCI	2.6400
6-12 months county prison	2.7070
1-10 years SCI	2.7070
2-4 years SCI	2.8973
2-10 years SCI	3.1612
4-8 years SCI	3.2412
3-9 years SCI	3.4089

The above scores fell into two distinct groups which can be viewed as two separate scales defined over different regions of the same continuum. The

¹¹ UNI-COLL CORPORATION, UNI-COLL PROGRAM USAGE GUIDE (1975). See also A. EDWARDS, *supra* note 1, at 27-28. Thurstone's Law of Comparative Judgments assumes that judges' perceptions vary with respect to the exact position of a sentence on the scale. In order to estimate a sentence's position on the final scale, additional assumptions have to be made. Case 5 refers to one of the additional assumptions. It assumes that the amount of variation in judges' perceptions of the position of a stimulus is equal for all stimuli.

¹² A. EDWARDS, *supra* note 1, at 40-46.

split appears between the two stimuli with repetitious scale scores (2.7070). An insufficient number of overlapping stimuli made comparisons between the two stimuli with the repetitious scale scores impossible, resulting in a lack of information about the relative positions of the two stimuli. The fact that the scale scores had the same numerical value is a function of the lack of information. The computer program without the necessary information arbitrarily assigned the scores the same value. Therefore, the computations resulted in two scales with different intercepts and thus no direct comparisons could be made between them.¹³

Bearing in mind that the analysis resulted in two separate scales, the original ordering of stimuli for the purposes of computation gave satisfactory scale scores for stimuli one through fourteen.¹⁴ However, it was found that by placing stimuli fifteen and sixteen in the numeric order of their calculated scale scores and thus using this ordering to calculate a new set of scale values, preferable scale scores resulted.¹⁵ The recalculated scale scores gave regression coefficients with higher significance levels. The multiple correlation squared, or the amount of variation explained, was also higher for the recalculated scale scores. Therefore, the recalculated scale scores were used. They are: four to eight years, SCI = 3.5114; three to nine years, SCI = 3.6791.

The object of the study was to obtain numerical weights for the various components of a sentence

¹³ For the case of incomplete data, a matrix of Z-scored proportions was calculated. From this matrix, a matrix of differences between adjacent columns of the Z-scored proportions is formed. The mean of the differences of adjacent columns gives the distances between adjacent stimuli on the continuum. When there is incomplete data, it is possible that two adjacent columns lack comparable proportions. Therefore, differences cannot be obtained. Consequently, the difference between the stimuli, and thus the distance, cannot be found. See also A. EDWARDS, *supra* note 1, at 42-43.

¹⁴ The original ordering of the stimuli for purposes of computation was based on a frequency matrix of the number of times stimulus j (column) was preferred (considered to be less severe) over stimulus i (row). The column sum of frequencies determined the order for calculation of the scale scores. Edwards did not propose a specified theoretical ordering; therefore, we may assume the order utilized by the program to be only one of several possible. An alternative is the order based on the calculated scale scores derived from the frequency matrix. This process would require a second calculation of the scale scores.

¹⁵ Preference for the new scale scores was determined when regressions were run to find the coefficient weights for the theoretically defined components of the sentence. The regressions will be discussed in greater detail below.

such that any sentence had a defined place on the severity scale. Regressions were run to obtain coefficient weights for the important elements of the sentence. The scale scores were regressed against the components of the stimuli which are the elements of the sentence such that the elements formed the independent variables. In other words, the scale scores acted as the data for the dependent variable while the sixteen stimuli formed the data for the independent variables. The beta weights of the regression equation, when calculated, formed the coefficients against which the numerical values of components of the sentence were multiplied.

The two sets of scale scores on which the regressions were calculated consisted of eleven stimuli and five stimuli. The eleven-point scale contained data points for all sentence elements and thus, at first glance, appeared adequate: probation was significant at the .007 level; probation with mandatory treatment was significant at the .007 level; the minimum sentence in the county prison was significant at the .002 level; the maximum sentence in the county prison was not significant; the minimum sentence at the state correctional institution was not significant; and the maximum sentence at the state correctional institution was significant at the .066 level.

The coefficient for maximum in the county prison was not significant. However, information obtained from the judges in conversation led to the author's conclusion that this is an appropriate result. According to the judges, prisoners sentenced to the county prison are expected to serve out their minimum sentence there. However, after the minimum is served, parole is left to the sentencing judge's discretion. A majority of judges related that their practice is to parole after the minimum. Thus, as shown by the data, the minimum county prison coefficient is a significant measure, whereas the maximum county prison coefficient is not.

The four coefficients for probation, probation with mandatory treatment, and minimum and maximum county prison were deemed to be valid measures of the degree of severity they contribute to a sentence. However, it was decided that the coefficients for minimum and maximum SCI on the eleven-point scale were not satisfactory weights for two reasons.

First, there was a high degree of intercorrelation between the two variables (.986). For purposes of constructing the schedule, the stimuli were set up to obtain barely discernible differences in order to yield maximum information with which to form the scale. These stimuli were also such that they varied together. The multicollinearity in the regres-

sion resulted in the placing of all of the explained variance on one of the two variables, specifically maximum SCI. Based on conversations with the judges, the author was led to question this result. Of the two elements, minimum SCI was given the greater importance by a majority of the judges.

Similarly, the second reason results from the nature of the stimuli—three to twenty months SCI, four to twenty-two months SCI and three to twenty-three months SCI. As one can observe, the minima SCI in the eleven-point scale are approximately equal. In interviews with the judges, it was learned that it can take up to six months of admissions procedures to be incarcerated in SCI. Therefore, the judges would tend to perceive the minima of these three sentences as being equal and thus disregard them. Maximum SCI is considered important by the judges because parole from SCI is a function of the State Parole Board and is consequently out of their purview. Both of these attitudes are reflected in the ordering of the SCI scale values in the eleven-point scale. If the stimuli are ordered according to the numeric values of the scale scores, then the minima appear at random, whereas the maxima appear according to the length of time (see list of stimuli above).

For these two reasons the coefficients for the minimum and maximum SCI of the eleven-point scale were considered unreliable. Thus, it was decided that their coefficients should not be included in the final weights. Another set of coefficients would have to be obtained.

The coefficients for minimum and maximum SCI derived from the five-stimuli scale proved to be a satisfactory alternative for the following reasons. The variables representing the SCI elements were not affected by multicollinearity. Second, the stimuli that make up the five-point scale are a better representation of the type of sentence imposed for an SCI incarceration. Therefore, the coefficients that represent SCI on the five-point scale were judged to be more reliable.

The solution would be to employ the coefficients of the five-point scale instead of the SCI coefficients of the eleven-point scale. One technique to remove multicollinearity is to substitute the ratio of two coefficients (in this case the SCI weights from the five-point scale which are not highly intercorrelated) in place of the original two SCI coefficients of the eleven-point scale.¹⁶ By using this technique two goals were accomplished. The intercorrelation

¹⁶ A. GOLDBERGER, *ECONOMETRIC THEORY* 192-94 (1964).

was removed and a set of reliable SCI coefficients utilized. Second, a regression could be calculated on the eleven-point scale yielding coefficients for all elements. Once calculated, the coefficients would enable one to obtain reliable estimates of the scale values of one scale continuum.¹⁷

The relationship between the two scales can be seen in terms of the structural equation which is as follows:

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_7$$

where,

Y = the scale value

x_1 = the months of medical probation

x_2 = the months of probation

x_3 = the months of minimum county prison

x_4 = the months of maximum county prison

x_7 = a combination term for minimum and maximum SCI

thus,

$$b_5x_7 = b_5 \left[x_5 + \frac{b_6}{b_5} (x_6) \right]$$

where,

x_5 = the months of minimum SCI

x_6 = the months of maximum SCI

and where,

b_6/b_5 = the ratio of the coefficients of maximum/minimum estimated from the regression coefficients of the five-point scale.¹⁸

The combination term for minimum and maximum SCI in effect adjusts the relationship of the minimum and maximum SCI of the eleven-point scale to the ratio of the coefficients of the five-point scale. The coefficient of the combination terms provides the b_5 weight and the coefficients of the combination term multiplied by the ratio of the coefficient of the five-point scale provides the b_6 weight.

RESULTS

A final regression was run to obtain estimates of the coefficient weights for the combined scale.¹⁹

¹⁷ The solution is predicated on the necessity of assuming linearity. The final R^2 (.962) of the regression gave support for the validity of this assumption.

¹⁸ The calculated ratio equals 1/7.444.

¹⁹ The R^2 of this final regression is 0.962. The F statistic is 25.17 with 5 and 5 degrees of freedom and a .002 level of significance.

The following is a list of the weights and their level of significance:

Medical Probation	0.0175	(.005)
Probation	0.0160	(.006)
Minimum County Prison	0.5450	(.002)
Maximum County Prison	0.0038	over (.500)
Minimum SCI	0.4633	} under (.001)
Maximum SCI	0.0622	

The coefficient maximum county prison was not expected to be significant. Interviews with the judges revealed that parole from the county prison was at the discretion of the sentencing judge and that it is common practice to parole after the minimum sentence is served. Thus, it is the minimum, not the maximum, which is the significant factor in the county prison sentence. The minimum and the maximum SCI share a significance level as a result of the technique used to combine the two scale subsets. A combination term and thus a combination significance level was calculated in order to judge the relationship of all sentence components to each other.

The estimates of the coefficient weights are estimates of the degree of severity that the judges attribute to the various components of the sentence which were then multiplied by the respective number of months in the component. By multiplying the components of any sentence by the appropriate weights, one can obtain the position of that sentence on the scale of sentence severity.²⁰

In concrete terms, the weights allowed us to make comparisons between the severity of sentences according to the perceptions of the judges. For example, 3 to 23 months in the county prisons (1.722) is approximately equal to 9 years probation (1.728) or 8 1/4 years medical probation (1.732); 11 1/2 to 23 1/2 months in the county prison plus 3 years probation (6.93) is approximately equal to 1 to 2 years SCI (7.05); 2 to 5 years SCI (14.85) is approximately equal to 1 to 12 years SCI (14.51).

The judges took into account many factors in their determination of the degree of severity in a sentence. Incorporated in these determinations were philosophies and understandings of the harshness of prison versus probation, the realities of the institutions to which they sentenced defendants, and the psychological and physical burdens that the various sentences entailed.

²⁰ For example, for a sentence of 3-23 months in the county prison, one would calculate as follows: $3(0.5450) + 23(0.0038) = 1.7224$.