

Spring 1932

Reliability of Factors Used in Predicting Success or Failure in Parole

Clark Tibbitts

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

Clark Tibbitts, Reliability of Factors Used in Predicting Success or Failure in Parole, 22 *Am. Inst. Crim. L. & Criminology* 844 (1931-1932)

This Article 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.

RELIABILITY OF FACTORS USED IN PREDICTING SUCCESS OR FAIL- URE IN PAROLE*

CLARK TIBBITTS

Attempts to predict the conduct of prisoners following release from penal institutions have developed quite rapidly since the appearance of Professor E. W. Burgess' study of parolees in Illinois.¹ In 1930 Sheldon and Eleanor Glueck² presented a study of 500 criminal careers in Massachusetts. This was about the time that George B. Vold³ was completing a study of parolees from the Minnesota State Prison. These studies have all followed the general plan of basing predictions upon the relationship of various factors affecting the individual, such as the *Nature of the Offense, Previous Criminal Record, Age*, etc., to success or failure on parole.

Each of these studies is evidence of an increasing interest on the part of the investigators in basing the predictions, not upon a multiplicity of factors, but rather, on a group of selected factors. Selections have been suggested or made on the basis of degree of association between particular factors and parole outcome, on inter-relationship among factors, and on reliability of classification, as well as on a priori considerations. Thus Professor Burgess collected data on nearly every available social factor, omitting from final consideration two which showed virtually no relation to success or failure on parole. The present writer in a study of 3,000 parolees from the Illinois State Reformatory employed a scheme of prediction which neglected every category under the several factors which did not show a definite relationship to parole outcome.⁴

The Gluecks assembled data on fifty different items characterizing each individual. By means of the coefficient of contingency the several factors were ranked according to the extent of association with parole and post-parole conduct. The different types of predic-

*Studies from the Institute for Juvenile Research, Chicago. Series C, No. 197.

¹Bruce, Harno, Landesco, *The Workings of the Indeterminate Sentence Law and of Parole in the State of Illinois*. Illinois Criminal Justice Survey, Part IV, 1928.

²*Five Hundred Criminal Careers* (New York: Alfred Knopf, 1930).

³*Prediction Methods and Parole* (Hanover, New Hampshire: The Sociological Press, 1931).

⁴Clark Tibbitts, "Success or Failure on Parole Can Be Predicted," *JOURNAL OF CRIMINAL LAW AND CRIMINOLOGY*, May, 1931.

tion tables were based on from six to fifteen factors selected on the basis of degree of association with outcome.⁵ The advantage of working with a small number of significant factors over the use of a large number of slightly significant ones is apparent.

Dr. Vold, however, turned his attention to testing the reliability of the classifications in which the cases were placed.⁶ He wished to know whether the same or a second investigator reclassifying a group of cases would place them in the same categories under each factor as was done in the original classification.

The importance of this point is readily comprehended. Classifications can hardly be valid unless they are somewhat rigidly defined. Moreover any scheme such as these must be constructed in such a manner that it can be placed in the hands of prison officials for use. Therefore the various categories must necessarily be well defined and not open to any large amount of difference in interpretation.

The present study is concerned with measuring the reliability of the classifications in which were placed 907 cases of young men paroled from the Illinois State Reformatory. Each case was classified by the writer during the summer of 1927, and again, one year later. The second study, made in 1928, was not undertaken with the purpose of testing the rigidity of classifications, but to enlarge upon the experience gained in the first, and certain definitions were changed. The definitions under two of the earlier factors, *Social Type* and *Residential Neighborhood*, were altered to such an extent that comparison was rendered impossible. A third factor, the *Psychiatrists' Prognosis*, has also been omitted from present consideration because of the removal of many of the psychiatrists' reports from the records during the years 1927 and 1928.

A further purpose of this second study was to discover whether, with a larger number of cases, the categories under the several factors would maintain their significance or lack of it for predicting parole outcome. That the data should lend themselves to a study of reliability classification was purely incidental, and not thought of at the time.

Had the original study been concerned with testing reliability it would not have been possible to make arbitrary changes in the classifications. These changes were made in the light of experience gained in Professor Burgess' original study, so that the classifications might better fit the data. For example, it was discovered during the progress of the 1927 study that a number of parolees had not

⁵Ibid., Chapters XVI, XVII, and XVIII.

⁶Ibid., Chapter V.

been previously committed to any institution, but that they were characters known to the police. Since their status was somewhat different from that of boys who had had no contact with the police, a new category "police character," was added under *Previous Criminal Record*, in the 1928 outline. Thus certain cases were arbitrarily changed from the 1927 classification of "no previous record," to "police character." In this case the changes did not influence the results of the present study because it was possible to combine the "police character" group with the "no previous record" group. No correction was possible for the three factors mentioned in the text above, and they were omitted.

There remain, however, sixteen factors, which were employed in the two years with no, or very few, changes upon which to base comparisons. The tests for agreement in classification have been made in a number of different ways.

CORRELATION OF PREDICTION SCALES

In the Burgess scheme of prediction the classifications under each of the twenty-one different factors were rated as either favorable or unfavorable toward successful outcome on parole, according to whether the proportion of violators in any particular category was greater or less than the proportion for the entire group. Thus each case was rated according to the number of classifications demonstrating favorable relationship to successful parole outcome. The cases were next grouped according to the number of favorable factors possessed, and the prediction scale was constructed by computing the proportion of parole failures in each of these groups. Accordingly a prospective parolee with sixteen or more favorable points was 98.5 per cent certain of making a successful parole, while one with fewer than five favorable points was 24 per cent certain of failing.⁷

The first check employed in this study was the construction of similar prediction scales for the 907 cases under scrutiny according to the 1927 and 1928 classifications, and then to correlate the number of unfavorable factors accorded each case in 1927 with the number given to identical cases in the second year. The correlation might have been expected to be nearly perfect had the data for making the classifications been complete, had the classifications been rigidly defined and completely free from arbitrary changes, and had the clerical work been without error. The relationship between the number of unfavorable factors possessed by each case in the two separate years was found to be expressed by $r = +.763 \pm .009$, low

⁷Ibid., p. 248.

enough, despite the omission of three factors (*Neighborhood, Social Type, and Prognosis*), to give evidence of serious limitations at some point or points in the study. Since the factors retained for this study were almost wholly devoid of arbitrary changes, the low correlation must be explained either by clerical errors or by difficulty of classification. That it is very largely the latter, and not clerical mistakes, will be seen in the analysis of the factors one at a time.

Before proceeding to the analysis of individual factors, it may be well to introduce a table showing the proportion of changes from one prognosis group to another in the 1927 and the 1928 classifications. Table I shows the number of cases for both years in each of five prognosis-groups, and the number and per cent distribution of the 1927 cases into the 1928 categories.

TABLE I
TABLE SHOWING THE NUMBER AND PER CENT OF CHANGES IN PROGNOSIS FOR
OUTCOME ON PAROLE AMONG 907 IDENTICAL CASES CLASSIFIED
IN 1927 AND AGAIN IN 1928

		1928 STUDY					
Number of Unfavorable Factors	Per Cent of Successes on Parole	1-4	5-6	7-9	10-11	12-15	
		Number of Cases	174	312	292	95	34
1-4	93.9	115 100%	81 70%	27 24%	7 6%		
5-8	86.9	464 100%	90 19%	241 52%	122 26%	10 2%	
9-10	68.6	191 100%	2 1%	40 21%	112 59%	33 17%	4 2%
11-12	48.1	108 100%	1 1%	4 4%	47 44%	40 37%	16 15%
13-15	34.5	29 100%			4 14%	12 41%	13 45%

It is seen, for example, that of the group of 115 cases of which 93.9 per cent were counted successful in the earlier year, 70.4 per cent fell into the corresponding group of the 1928 study, 23.5 per cent into the next group, and 6.1 per cent into the group two steps removed. Similarly, the other categories show a great many changes with the majority moving up or down one step. It is evident from this table, as well as from the correlation between the number of unfavorable factors in 1927 and 1928, that these studies leave much to be desired at the point where the data are classified.

CHANGES AMONG INDIVIDUAL FACTORS

The 907 cards in each set were matched and the classifications compared to discover the number of changes made in 1928 from the 1927 ratings for each factor and for each category under each factor.

The factors are the various aspects of the cases under which the data are classified. Thus, each case is classified according to the factor, *Offense Named in the Indictment*, as Larceny, Robbery, Burglary, Fraud, Sex, Homicide, or Other. The sixteen factors may be briefly characterized according to the numbers in Table I, as follows: (1) nature of sentence—whether indeterminate sentences of 1 to 10 years, 10 years to life, and flat sentences as 20 years; (2) offenses named in indictment; (3) age of parolee when he was released from the reformatory; (4) country of birth or race of the father of the inmate; (5) whether or not the parolee was permitted to plead guilty to a lesser charge than that named in the indictment at the time of the trial; (6) length of time spent in the institution; (7) severity of previous incarcerations, probations, etc.; (8) number of punishments while in the reformatory; (9) diagnosis as egocentric, inadequate, sensational personality, or psychotic; (10) statement of the circumstances of the case by the prosecuting attorney; whether he merely states the facts, recommends leniency, or protests early release; (11) whether first, occasional, habitual, or professional offender; (12) size of city in which the youth resided at the time of the crime; (13) whether or not he was a resident or a transient in the place where the crime was committed; (14) number of associates in the crime; (15) mental age; (16) whether regular skilled worker, casual unskilled, etc.

From the following table it is at once seen that *Nature of Sentence*, *Nature of Offense*, *Age at Time of Parole*, *Nationality of Father*, *Acceptance of a Lesser Plea*, and *Length of Time Served*, show the smallest proportion of changes in classification and have, in general, the highest coefficients of contingency, and hence are the most reliable factors, on the basis of adaptability, for securing classification. With these should be included *Previous Criminal Record*, and *Punishment Record*, which would have demonstrated greater reliability had it not been for certain minor arbitrary changes. The high degree of accuracy of these factors is explained by their extreme objectivity. The classifications are made from almost wholly reliable, complete, and accurate data with virtually no necessity for inference or interpretation. With rigidly defined categories and extremely careful consideration of the data at the time of classification, the uncertainty under these factors could be practically eliminated. On the basis of re-

TABLE II

FACTORS USED IN PREDICTING OUTCOME ON PAROLE SHOWING THE PER CENT OF CHANGES IN CLASSIFICATIONS BETWEEN 1927 AND 1928 AND THE RATIO OF THE COEFFICIENTS OF CONTINGENCY TO THE HIGHEST POSSIBLE COEFFICIENTS

Factors	Per Cent of Changes 1927 to 1928	Coefficient of Contingency		
		Computed C	Highest Possible †	Col. 2
				Col. 3
		$\frac{\text{Col. 2}}{\text{Col. 3}} \times 100$		
	1	2	3	4
1. Nature of Sentence.....	1.0	.890	.894	99.6
2. Offense Named in Indictment...	2.9	.887	.894	99.2
3. Age at Time of Parole.....	3.8	*	—	—
4. Nationality of Father.....	4.1	*	—	—
5. Acceptance of Lesser Plea.....	6.6	.755	.816	92.5
6. Length of Time Served.....	6.9	*	—	—
7. Previous Criminal Record.....	10.7	.861	.894	96.3
8. Punishment Record	11.4	.739	.816	90.6
9. Psychiatric Personality Type....	14.9	.827	.894	92.5
10. Statement of the Prosecuting Attorney	16.8	.614	.866	70.9
11. Type of Offender.....	19.6	.667	.816	81.7
12. Size of Residential Area.....	22.2	.746	.894	83.4
13. Mobility	22.4	.428	.816	52.5
14. Number of Associates.....	22.5	.813	.894	90.9
15. Mental Rating.....	27.1	.851	.913	93.2
16. Work Record	38.7	.454	.894	50.8

†Yule. *An Introduction to the Theory of Statistics*. 8th Edition, 1927, p. 66.

*In view of the arduous labor involved, and because of the certainty of their being exceptionally high, the coefficients for these factors were not computed.

liability, then, these eight factors are highly satisfactory for use in predicting outcome.

Another group of factors, *Statement of the Prosecuting Attorney, Psychiatric Personality Type, Type of Offender, Size of Residential Area, Mobility* (whether resident or a transient in the community), *Number of Associates*, and *Mental Rating* show changes ranging from 14.9 per cent to 27.1 per cent and ratios of the coefficients of contingency to the highest possible coefficients ranging from 95.2 to 52.5. The reader will undoubtedly notice that there is not an exceptionally high correlation between the percentages of change and the coefficients of contingency. This is explained by the fact that the two are not measures of quite the same thing. The first measures simply the number of changes, while the coefficient measures, in part, the degree of concentration or dispersion.⁸ Hence it is necessary to strike

⁸The two factors, *Statement of Prosecuting Attorney* and *Number of Associates* are cases in point. The former has a comparatively small number of changes, but a low coefficient, while the latter has a much higher proportion of changes and yet a high coefficient. The changes are more widely dispersed and fewer in number than those under *Associates*. The wide dispersion

some sort of interpretive balance between the two measures. It will be seen from Table II, that three (*Personality Type, Associates, and Mental Rating*) of the seven factors just now under consideration (numbers 9 through 15) might, according to their coefficients, be placed with the preceding group of more reliable factors. The reason for the high coefficients is that while there were many changes they of changes, on the one hand, and the fairly high concentration, on the other, are shown in the accompanying contingency tables.

TABLE III
STATEMENT OF PROSECUTING ATTORNEY
THE 1927 CLASSIFICATION DISTRIBUTED AS IN 1928

	No Record	Factual	Recommends	Protests	Total
No Record	2	9	0	0	11
Factual	21	650	36	20	727
Recommends	2	28	80	4	114
Protests	5	23	4	23	55
Total	30	710	120	47	907

$C = .614.$

Per Cent of Changes = 16.8.

TABLE IV
NUMBER OF ASSOCIATES—1928 CLASSIFICATION

1927 Classi- fications	None	One	Two	Three	Four	Total
None	266	48	14	6	0	334
One	29	260	30	17	2	338
Two	9	22	115	4	1	151
Three	0	5	9	42	5	61
Four	0	1	2	0	20	23
Total	304	336	170	69	28	907

$C = .813.$

Per Cent of Changes = 22.5.

The size of the coefficient is to some extent a function of the size of the table. The *Number of Associates* is shown on a five-fold table (Table IV) in which the highest possible C is .894, while the *Statements of Prosecuting Attorneys* are shown on a four-fold table in which the highest possible C is .866. When the same data on number of associates is compressed into a three-fold table the coefficient becomes approximately .7 and the number of changes is 18.0 per cent. The " C " is approximately 86 per cent of the theoretical highest possible value. To some extent, then, the discrepancy in the values indicating reliability in these two cases is due to the size of the table, i. e., to the number of classifications employed. The remainder of the discrepancy is due to extent of dispersion and concentration. Since it is this that the contingency coefficient measures and not simply the number of changes, some discrepancy must be expected between the two measures. The discrepancies are even useful in that they warn the investigator of the necessity of examining the tables, and not to depend upon the " C " alone. Concentration or dispersion of changes may have different significance for the problem.

were so concentrated as to affect the coefficients favorably. They properly belong in the less reliable group.

The unreliability among these seven factors (*Statement of the Attorney, Personality Type, Type of Offender, Size of Residential Area, Mobility, Associates, and Mental Rating*) is explainable on a number of bases, the most important of which are incomplete data, lack of certainty or definitive character of the data, and the necessity of interpretation. The psychiatrist, for example, frequently wavers, of necessity, in assigning a case to a particular personality type, making it exceptionally difficult for an investigator to decide which of two, or perhaps more, carried the stronger weight. In many cases insufficient data are present to permit certainty in determining the *Size of Residential Area* in which the criminal resided or whether he actually resided there or elsewhere. Thus, while the definitions of the categories under size of *Residential Area* and *Mobility* are rigid, the factors are unreliable because of the lack of completeness of the data.

The *Number of Associates* in the crime should be a rather easy matter to determine, yet the number of individuals brought to trial in a particular case does not always represent the number involved in the crime, nor does the number sentenced necessarily represent the number tried or involved in the crime. The various documents in the prisoner's information file very frequently fail to agree on the number of associates. *Mental Rating* would have shown fewer changes had the various prison records been in agreement and had the various ways of reporting it been more uniform. This study is concerned, of course, only with the reliability in classifying the data found, and not with the methods or results of the mental tests.

Type of offender is a factor involving a relatively large amount of interpretation from data which must necessarily remain somewhat meager. The classification of the cases was based on previous criminal record, work habits, and whatever information was available, on home and community life. While these data can undoubtedly be improved, it is, of course, truly impossible to know the number of crimes committed without apprehension; yet this is perhaps the most important single factor in arriving at a classification of type of offender.

In general, the factors in this group deserve greater reliability, although to attain it will involve better record-taking in the institutions and more complete investigations of community factors. If these factors show a high correlation with outcome on parole and are

otherwise satisfactory, the collection of additional material becomes indispensable for accurate prediction.

The factor, *Work Record* before admission to the reformatory, demonstrates by far the greatest lack of reliability. Next to *Social Type*, which was omitted from this study, *Work Record* is perhaps the most subjective of all the factors with which the investigator deals first hand. It was necessary to classify each case as a "casual," "irregular," or "regular" worker, which would have been no easy matter even had the data been complete rather than sketchy and unreliable. It seems doubtful whether it is possible to make rigid definitions for data of this sort, especially if the definitions are to be used by a second investigator.

Examination of the individual factors, then, reveals wide discrepancies in the certainty of classification. The discovery of these discrepancies by the methods here employed or by any other methods is of great significance in setting up any scheme for predicting outcome of parole. The classifications must be reliable or they are valueless.

Through measuring the extent of their unreliability and through a knowledge of the limitations of the data, the investigator is able to set about improving the data, reworking the definitions, or, in some cases, eliminating the factors from consideration. Thus the eight factors showing the highest reliability among those considered here could be given positive certainty through adherence to rigid definitions. Improvement of the seven next most reliable factors depends largely on the availability of more complete and reliable data. This would include the factor, *Neighborhood*, omitted from this study.

As would be expected, the most reliable categories are those which require little more than tabulation of specific and complete data. Errors begin to enter when the data are incomplete or uncertain, when the various records do not agree, and when interpretation is required.⁹ When categories are not clearly defined, when the data do not conform to the definitions, and when entries must be in any

⁹When any scheme such as this is put to practical use it must necessarily fall into the hands of many different persons. For this reason, the present study would have been a more severe test of reliability had two different persons, rather than the same one, classified the data. Perhaps the importance of this is somewhat lessened in that the study is designed to test differential reliability among the factors rather than among investigators. While the reliability of the factors would, undoubtedly, be different had there been two investigators, the rank of the differentials would probably have remained the same.

way forced, it is difficult for the same or different individuals to classify the same data with any considerable agreement.

Improvement among virtually all of the factors is certain to come with the adoption of scientific methods of parole prediction. While the first student of the subject must accept the data as he finds them, the permanent investigator, knowing the extent of unreliability among the various factors, will set about to improve the bases on which they rest. Increased reliability will be correlative with the development of prediction. But the improvement of technique will not necessarily wait on interest in the problem of parole, for similar methods are being applied to outcome of probation, rearing of children, marriage, and so on.

SUMMARY

The coefficient measuring the correlation between the number of factors unfavorably related to success on parole for identical cases in 1927 and in 1928, and the table showing the changes among prognosis-classes, indicate that there are many uncertainties of classification. Table II, showing the percentage of changes under each factor and the coefficients of contingency between the two separate classifications, indicates where the greatest unreliability is to be found. The factors ranked according to reliability show the investigator at once where he must begin in order to improve his data and his definitions. The bases of selection like the studies themselves are still developmental. At the present time it seems that there are at least three bases for choosing the factors to be employed. First, the factors must be related to the outcomes they are expected to predict. Secondly, they must be subject to reliable classification or else the predictions will be uncertain and of little utility. Thirdly, it is questionable whether factors should be used which show interrelationship. In all of the existing studies of parole there is reason to suspect that considerable overlapping if not duplication exists among factors. The question of whether or not it is desirable or correct to include overlapping factors is just now being attacked.¹⁰ It is obvious, however, from this paper that greater accuracy of prediction can be gained through the application of some sort of test for reliability of classification and one may feel quite certain that as interest in this and in similar problems grows, both the data and the method of handling them will be improved.

¹⁰See S. A. Stouffer and C. Tibbitts, "An Application of the Method of Expected Cases to Research in Sociology." Paper read at the Annual Meeting of the American Sociological Society at Washington, D. C., December 30, 1931.