Increasing Reliability and Validity in Juvenile Delinquency Research

Balogh, Joseph K.
INCREASING RELIABILITY AND VALIDITY IN JUVENILE DELINQUENCY RESEARCH

JOSEPH K. BALOGH

Department of Sociology, Bowling Green State University, Bowling Green, Ohio

"In recent years there has been a great deal of interest created concerning the behavior of our teen-age population. While theories of punishment relative to human behavior on the associational level are very old, nevertheless, scientific application and critical objective studies are still conspicuously absent. The lack of scientific studies to predict anti-social behavior is one of the great academic challenges of our time ..." (Balogh, 1958). Professor Carr's classic statement lends considerable support to the thesis that contemporary society is grossly misinformed about the entire problem of juvenile delinquency (Balogh, 1958):

Now the interesting thing is that millions of people like you don't yet see the crucial difference between the way they handle their motor cars and the way they handle their children. The crucial difference is the difference between fact-mindedness and myth-mindedness; the difference between a cause-and-effect approach to a difficult and a traditional, what-do-I-think-ought-to-be-done approach. Like the famous character who didn't know he was talking prose, the average man is blissfully unaware that he has three totally different ways of dealing with reality.

What is juvenile delinquency? It is a well known fact that the concept of juvenile delinquency is not readily identifiable. If society could come up with a fixed definition as to what juvenile delinquency is, many problems relating to legal jurisdiction, therapy, corrections or institutionalization, and research would be closer to solution. As a case in point, McDavid and McCandless (1962) state that the definition of juvenile delinquency is significantly related to these criteria: legal, religious or ethical, sociological, psychological, legal conviction and sentencing, court appearances or records of arrests, dependent and neglected children, judgmental ratings of behavior by teachers, social workers, and specific kinds of behavior. "Still other research workers may base their definition of delinquency on certain assumptions about the attitudes and values held by the adolescent, with empirical and normative research called upon to determine the nature of attitudes and values normally held by socially accepted adolescents" (McDavid and McCandless, 1962). Kahn (1965) is of the opinion that ... "delinquency is not a disease but rather an administrative category which joins together many behaviors, circumstances, and statuses and which reflects certain societal assessments and strategies for coping with deviance." Carr (Balogh, 1958) feels that "delinquent behavior is not an invariant certainty but only a stated degree of probability."

It can be readily argued that all of these definitions of juvenile delinquency have some validity of purpose. At the same time, however, these many and varied definitions make for greater confusion in juvenile delinquency research. "All of these definitions can be useful for research purposes, but serious problems arise in the integration of research findings from various sources. One cannot superficially combine evidence that delinquency (by one definition) is related, for...

---

1Presidential address, delivered on April 19, 1967, at Annual Meeting of The Ohio Academy of Science in Bowling Green, Ohio.

example, to intelligence, with evidence that delinquency (by some other definition) is related to social class, without taking careful account of each of the definitions employed" (McDavid and McCandless, 1962).

Why is contemporary society concerned about the problems of juvenile delinquency? Obviously, most of the explanations relate to legal and moral aspects of the problem. The contention can be made, however, that basically juvenile delinquency is an acceptable way of life among many children peer groups. Of course, society cannot accept this interpretation, because it, too, is very defensive, attitudinally.

Finally, whereas delinquency in the past was frequently a function of variable community attitudes and standards, the growing tendency towards a middle-class sense of normalcy and uniformity have tended to make us far more sensitive to the manifestations of such youthful disorders than we were, very likely, in the past. This sensitivity registers itself in higher statistical roles of delinquent disorders (Bloch, 1958).

The problem of probable future delinquents is also a very serious matter. This observation is particularly accentuated by present-day research, as this research relates to juvenile delinquency proneness. A statement by the Council of the Society for the Psychological Study of Social Issues is particularly appropriate:

Unless the utmost caution and care are taken, children who are 'identified' and labeled as probable future delinquents are likely to be treated and isolated as 'bad' children by teachers and others who are now subjected to the virtually hysterical climate of opinion concerning juvenile delinquency. Such treatment is likely to increase the child's sense of social alienation, and thereby, increase the probability of his becoming delinquent or of developing other forms of psychological maladjustment (Kahn, 1965).

The important problems of contemporary juvenile delinquency research are many and varied. Today's researchers have not yet found a way to implement theory and methodology. More importantly, the researcher finds it exceedingly difficult to establish a point of departure for any research. Balogh and Rumage (1956) add these supportable data:

The big task facing the student of juvenile delinquency is the creation of a research tool which will have predictive value, and, thus, emphasize prevention rather than purely and simply emphasizing the matter of rehabilitation ... It is the contention of these writers that there is sufficient theory of a valid nature available for theoretical generalizations and that what is now needed is a research tool which can predict with some degree of reliability the many manifestations of deviant behavior. Studies dealing with juvenile delinquency are numerous, but these emphasize a theoretical rationale only and are conspicuously devoid of predictive research techniques indicative of delinquency proneness.

More importantly, some of the questions and problems raised by Empey (1967) are also noteworthy:

... we need a series of related studies which would, first, identify a representative population of adolescents, their class positions, their value-beliefs and commitments, various measures of delinquent acts (self-reported and official), their symptoms of disability, and their group affiliations; and, second, follow these adolescents through the institutional paths—educational, economic, or correctional—along which they are routed by officials. Which juveniles are processed legally and on what criteria? In what ways are they the same or different from nonprocessed juveniles in terms of values, class position, group affiliations, actual delinquent acts, and so on.

Given such research we might then be in a better position to know not only what the consequences are for those who are apprehended and processed by legal and correctional institutions but also what the consequences are for those who are not processed. This would most certainly apply to middle-class as well as lower-class juveniles. Hopefully, we might gain better insight into the total mosaic composed of delinquent values, actual behavior, and official reaction. Are delinquent values widely shared and is delinquent behavior common? Does legal or semilegal processing contribute to the solidification of delinquent groups? Is there differential treatment of juveniles based not on actual behavioral or value differences but on other identifying characteristics?

The nature of the prediction problem involves an individual's performance
or his expected behavior. "The prediction problem, viewed in this way, involves two independent assessments of persons; the two assessments are separated in time. On the basis of one assessment, 'predictor categories' are established (by various means). These are items of information believed to be helpful in prediction; they are the 'predictors'. The second assessment establishes 'criterion categories'; these are the classifications of performance to be predicted" (Gottfredson, 1967). It is obvious to the researcher that all requirements for reliability and validity are essential for both "criterion" classifications as well as "predictors."

It is a fact that the elements of reliability and validity are very often absent in juvenile delinquency research. Aside from tests of operational validity and reliability, we need experimental designs "which will give functional validity" (Balogh, 1965). As a further illustration, "The crux of the matter is not so much how consistently ... psychiatrists, for example, agree with one another as it is: are meaningful categories being used to agree upon. And by meaningful, one means how successful is the outcome in comparison with other possible agreements" (Balogh, 1965). In other words, validity is very important.

What does one mean by reliability? "Reliability refers to the consistency or stability of repeated observations, scores, or other classifications. If a procedure is reliable, then repetitions of the procedure lead to similar classifications" (Gottfredson, 1967). One cannot predict, for example, when measures are completely unreliable. It is now apparent that a foolproof technique of prediction is not a virtual certainty. We can think only in terms of degree of probability. For instance, "A prediction method has as many validities as there are criterion measures to be predicted" (Gottfredson, 1967).

What are some factors influencing predictions? Gottfredson (1967) states that the factors of selection ratio, base rate, method used to combine predictor variables, representativeness of source, and the number of predictor variables are all important in prediction. Mannheim and Wilkins (1962) add that prediction tables should be characterized by simplicity, efficiency, repeatability or reliability, and validity. It is the contention of this writer that the representativeness of the sample is a basic ingredient of juvenile delinquency research (Balogh, 1965).

We can now raise the question as to what are the actual observable differences that differentiate juvenile delinquents from non-juvenile delinquents? Balogh (1965) states that "the answer to this question is related to the sampling process first, and the subsequent evolving of the delinquent population (experimental group) and the non-delinquents (control groups) ... These subjects must be truly representative of the universe from which they have been drawn ... Further, the probability of the differences being attributable to chance factors must be ascertained by proper statistical steps." It should be apparent even to the neophyte that many serious difficulties are encountered in obtaining random samples. Where and how the samples are obtained impose many limitations on juvenile delinquency research. Let us examine this problem, albeit briefly.

Slum subjects tend to be criminogenic. There is a significant difference between institutionalized offenders and non-institutionalized offenders. Rubin (1958) states that institutional life may contribute "to the greater display among delinquents of social assertiveness, defiance, ambivalence toward others, hostility, suspicion, destructiveness" and less submissiveness to authority. Many samples are based on court cases. "Frequently, records tend to assume that a child is delinquent behaviorally, even though he may not have been officially adjudicated as delinquent. Inferential analysis concerning juvenile delinquency in general are also vital factors" (Balogh, 1965). The selection of subjects from a guidance center or a clinic poses some methodological problems. It can also be argued at this time that an increase in the level of experience of psychiatric personnel may also contribute to an increase in reliability. A researcher must frequently assume that his particular sample is not universally applicable. Perhaps his universe
may remain undefined and nebulous. The researcher may not be in a position to know very much about the characteristics of his sample. Significantly, the technique of cross-validation is frequently omitted in juvenile delinquency research.

The sample has other very important research implications. A refinement of sample data is highly relevant to both reliability and validity. Many important questions can be raised concerning the characteristics of the subjects comprising the sample. It is very important, for example, "to recognize what is 'typical in the group'." It is deemed important to show individual variances in the group. One might also show how the subjects are distributed with respect to the variable being measured. Also, what is the relation of the different variables in the data to one another. A researcher may want to describe differences between two or more groups of individuals (Balogh and Rumage, 1956).

What part can measuring instruments play in juvenile delinquency prediction? It is readily admitted that our ability to predict juvenile delinquency is imperfect. Although many actuarial measures are used, it is quite apparent that these, in themselves, are not sufficiently reliable in predicting juvenile delinquency. Kahn (1965) makes the observation that "the need for caution is heightened by the fact that the prediction is actuarial: it may tell us that a boy is in a group whose members have an 85 per cent chance of being delinquent. His chances do not thereby become 85 out of 100, since special things may operate for him."

Some researchers (Reckless and Dinitz, 1967) are of the opinion that what is needed is an effective self-concept measure:

Undoubtedly, there is need for the development of an effective self-concept measure which can assess the direction toward or away from delinquency or deviant behavior generally. There is need also to develop measures of other self factors which control directionality. When such factors are uncovered and when they are effectively measured, then it should be possible to chart workable programs to prevent delinquency and to re-enforce the components of self which enable the youth to be an effective conformer.

This writer is impressed with Gross' interdisciplinary theory on human growth and development. He conceptualizes a theory of numerical patterns of human association wherein certain types of inferences may be associated with certain types of processes. It is conceivable that these numerical patterns of association can be incorporated into a schematic pattern in juvenile delinquency research. Gross (1967) lists these inferences and processes:

A. Inferences from one characteristic to one characteristic (unilinear process)
B. Inferences from one characteristic to many characteristics (divergent process)
C. Inferences from many characteristics to one characteristic (convergent process)
D. Inferences from many characteristics to many characteristics (multilinear process)

Gross has also conceptualized categories of sameness and differences among characteristics. This writer feels that these conceptualizations have a direct bearing on the reliability and validity of measuring instruments in juvenile delinquency research. Furthermore, Gross' theory is directly applicable to developing a rationale for interdisciplinary research in juvenile delinquency. Gross (1967) postulates the following categories of sameness and difference among characteristics:

1. Inferences within the same kinds of characteristics or
2. between different kinds of characteristics
3. Inferences within the same population group or
4. between different population groups
5. Inferences within the same period of time or
6. between different periods of time
7. Inferences within the same level of analysis or
8. between different levels of analysis
9. Inferences within the same scientific disciplines or
10. between scientific disciplines

Gross (1967) places great emphasis on the conceptual dichotomies “one” and “many,” “same” and “different,” and is of the opinion that these conceptual dichotomies “are but a step or two removed from what is, perhaps, the most basic ingredient in the structuring of empirical generalizations—that of simple discrimination.” The researcher must involve resemblances and differences in empirical observations (Gross, 1967).

Turk in his attempt to construct a theory of delinquency makes some very important observations concerning the reliability and validity of variables. He states that verification studies are needed to stress “particularly (1) the need to establish the reliability and validity of variables in relation to theoretical constructs, and (2) the fact, sometimes neglected in practice, that ‘failures’ in verification studies may be attributable to methodological deficiencies as much as, or even more than, theoretical shortcomings” (Turk, 1964). There are undoubtedly more significant omissions in the methodology of juvenile delinquency research than there are theoretical deficiencies. It is also true that we have not been able to construct predictive scales and predictive tables with reasonable certainty. Many so-called studies of verification have been, in the main, dishonest intellectual attempts on the part of the graduate students of some of the “great researchers,” who have themselves developed unreliable and methodologically unsound measuring instruments. Intellectual dishonesty has no place in any branch of research, let alone juvenile delinquency. It is time that these pseudo-behavioral scientists are reminded of some simple rules of statistical procedure in statistical research (Balogh and Rumage, 1956):

1. Never have preconceived ideas as to what the figures are to prove.
2. Never reject a number that seems contrary to what you might expect merely because it departs a good deal from the apparent average.
3. Be careful to weigh and record all the possible causes of an event and do not attribute to one what is really the result of the combination of several.
4. Never compare data which are not fully comparable.

Statistical interpretations are also highly relevant to prediction, reliability, and validity. Jahoda, et al., (1951) make these poignant observations:

1. To characterize what is ‘typical in the group.’
2. To indicate how widely individuals in the group vary.
3. To show how the individuals are distributed with respect to the variable being measured.
4. To show the relation of the different variables in the data to one another.
5. To describe the differences between two or more groups of individuals.

Reliability and validity are the significant concomitance of scale construction. Variations of scores are affected by countless variables. Jahoda, et al., (1951) state that certain conditions are instrumental in affecting the variable factors involved in score differentiation:

1. True differences in the enduring characteristic which one is attempting to measure.
2. True differences in other enduring characteristics of the individual which affect his score.
3. Differences due to transient personal factors.
4. Differences due to situational factors.
5. Differences due to inadequate sampling of items.
6. Differences due to lack of clarity of the measuring instrument.
7. Differences due to variations in administration.
8. Differences due to mechanical factors.
9. Differences due to factors in the analysis.

Before the behavioral scientist can really predict juvenile delinquent behavior,
he will have to construct a measuring instrument that will incorporate all of the methodological criteria postulated by Jahoda, et al. (1951).

Pertinent questions can also be raised concerning the use of certain statistical techniques. Much of contemporary juvenile delinquency research is predicated on so-called sophisticated statistical techniques when, in reality, more refined and less complex statistical techniques would be just as appropriate. It is equally to be deplored that researchers in the behavioral sciences tend to over-elaborate results that are statistically significant (particularly, if these results are statistically significant beyond the P<.001 confidence level), but, at the same time, completely ignore those findings that have not attained statistical significance. It can be argued with a great deal of validity that it is equally important to publish data which have not attained statistical significance. It is conceivable that data lacking statistical significance can be maximized in later research. Furthermore, it is not within the academic or intellectual province of the researcher to completely disregard data lacking statistical significance. Wilkins (1962) states that "the real test of statistical methods is not whether a logical case can be made out for them, but whether they work and whether they work better than any other method."

Scaling techniques have come a long way in the behavioral sciences. Present-day researchers (particularly those in social psychology and sociology) have developed some excellent predictive scales. At the same time, however, behavioral scientists must readily admit that predicting juvenile delinquency is not a virtual certainty, considering the present level of scale construction. Data gathering must improve. In the construction of tables, involved factors will have to be objectified. Glueck (1962), in referring to prediction tables, states that researchers should "remain open-minded about the uses of these and similar devices and to look upon predictive devices or research tools in a search for the etiology of delinquency . . ." It is frequently argued that the behavioral scientist could accomplish more by developing a diagnostic measuring instrument. Along these lines, Chwast (1958) makes these observations:

The study seems to substantiate the value of understanding deviant groups by comparisons of multi-level measurements of the same attitudinal variables. This technique holds considerable promise in more objectively assessing such theoretical constructs as self-concept, social role perception, stereotypy, repression, dissociation and the 'malevolent transformation,' to mention but a few. Although deeper-level techniques appear superior to more direct methods in evaluating basic feelings and attitudes, a determination of the direction and degree of distortion in upper-level perceptions can vastly enrich diagnostic and therapeutic research and practice.

Some researchers are of the opinion that what is now needed is a sociological predictive instrument. In support of this idea, Amos and Wellford (1967) state that "... we would then indicate that a sociological prediction has yet to be developed and that these instruments now in operation will continue to fail because their objective has a low probability of occurrence. The thrust of the position taken . . . is toward the development of a causally relevant, sociological, statistical (though not necessarily of the traditional form) instrument for prediction and evaluation."

Although Amos and Wellford argue well, the simple fact is that the sociologist does not have an academic monopoly on researching juvenile delinquency. As a matter of fact, sociology could come under academic indictment for presuming too much. Juvenile delinquency cannot be delimited solely to sociological analysis. From a methodological standpoint, it might be more advisable to harness the many contributions of other highly related disciplines. One of the great gaps in the understanding of this problem has been the fact that many behavioral scientists are interested in protecting their own vested academic interests instead of developing an interdisciplinary perspective. The clinical psychologist, psychiatrist, social
anthropologist, social psychologist, and representatives from the field of education can also contribute significantly in developing a predictive instrument. Westie and Turk (1965) claim that "if the strategy is to bear fruit, atomistic, individualistic research must be de-emphasized in favor of large-scale coordinated research programs involving many researchers with varied skills and interests."

A serious question can also be raised concerning the monotonous regularity with which we stress the fact that any measuring device must be a statistical instrument. Of course, the behavioral scientists (and this is especially true of the sociologists) have been on an empirical "binge" for at least the last decade. It would seem that the only way to achieve academic respectability is to develop an empirical orientation. Although objectivity has a definite place in sociology, it can also be argued that the analysis of human behavior is not always specifically limited to the objective. The fact remains (and many behavioral scientists concur) that the subjective approach in the understanding of human behavior is also very important. The subjective approach may be viewed as a complementary process. It might, therefore, be of some importance at this time to discuss the reliability and validity of interdisciplinary psychiatric diagnosis in juvenile delinquency research.

It is a fundamental fact that there is no single approach to studying juvenile delinquency. The utilization of a control-oriented approach, however, is sound. Measuring inter-relationship criteria such as interpsychic, constitutional, and group interaction serves a useful purpose. Ecological factors can also be studied in this context. The idea of correlating sociological variables with clinical data is excellent, particularly when these data are ultimately integrated into the dispositional process. The use of a table of random numbers to select the study sample, the training and orientation program of the interdisciplinary team members in the study, the use of only well-qualified and experienced child psychiatrists for data gathering, and the use of both psychiatric data as well as rating schedules that permit machine tabulation and statistical manipulation of the data, are all excellent research techniques. Testing the reliability of the items of the schedule against that of psychiatrists' ratings by statistical means appears to be a satisfactory attempt at attaining intra-test reliability. With respect to sampling bias, one must remember that bias is further enhanced by the greater rather than the decreased awareness of court workers of intrapsychic problems, because only the most advanced courts have guidance centers. It must also be pointed out that the psychiatric approach, in and of itself, is not beyond scientific scrutiny. The clinician, who is obviously a patient-oriented person, can see nothing wrong with focusing on the actor rather than on the act. Focusing on the act is meaningful only if it is subordinated to the actor and is interpreted in the light of his emotional make-up. Although a rating scale is highly useful, and even desirable, it does not deepen the understanding of the person at all. On the contrary, it schematizes it. This is not always wrong, but we need to face the fact that we cannot encompass all aspects of the person's needs in a scale. It is of dubious mathematical validity to use an ordinal scale as though it were cardinal.

The knowledge of early developmental data is of paramount importance in arriving at a psychiatric or a social diagnosis. We need to be aware of the fact that most habitual delinquents, or recidivists (repeaters), have developed certain characterological traits which, at our present state of knowledge, can be identified to best advantage from etiological factors. Retroactive recall may be only a slight disadvantage in the light of this need. It is outweighed by the attendant exploitation of current parental attitudes and the clinical impression of the patient of his behavior. Furthermore, it would also depend on the personality and needs of the psychiatrist as to whether he would consider juvenile delinquents more disturbed and impulse-ridden than they actually are. Presumably, a psychiatrist would take under advisement the factor of control over these impulses rather than
the impulses themselves. It is conceivable that a person with a neurosis, psychosis, or characterological problem may have equally strong impulses and not act them out, but develop symptoms which are detrimental to his own personality rather than to society. McDavid and McCandless (1962) state that "because juvenile delinquency represents a diffuse and pervasive syndrome within the personality structure of the adolescent, actual psychological therapy may be a necessary technique for the control of delinquency."

Let us improvise a hypothetical situation where a team of clinicians, psychiatrists, and other behavioral scientists are interested in devising a measuring instrument or a scale for the purpose of predicting juvenile delinquency. What methodological questions might be raised concerning validation and reliability? These questions may be relevant to the validation of the scale (Balogh, 1965):

1. How was the scale validated?
2. What scaling techniques were used?
3. Are the limits of the continuum known?
4. Were there tests of unidimensionality given?
5. Were the items arbitrarily assigned a position, or were they fixed by test?

These questions may be relevant to the reliability of the scale (Balogh, 1965):

1. How much of the effect is due to uniform training in the training sessions?
2. How much of the effect is due to a compressed scale with relatively few points, thereby limiting the possibility of deviating?
3. How much of the effect is due to clumping of the scale—were certain scale values used more frequently than others?
4. How much of the effect was due to the limited number of psychiatric categories of higher abstraction?

Although not all of the above questions may be universally applicable in the construction of all scales and measuring instruments, the simple fact is that, in constructing a scale or a measuring instrument, there are many dangerous pitfalls of a methodological nature. Jahoda, et al., (1951) make these observations:

It is important to remember that any validation which attempts to predict behavior must rely on both a logical as well as empirical definition even though both the logical as well as the empirical rely on one and the same thing, namely, prediction . . . There are instances in which one could assert that a measuring technique is logically valid, in the sense that it is an appropriate empirical definition of a theoretical construct, but empirically invalid in the sense that the construct is not relevant to the phenomenon one wishes to predict . . . Similarly, there are instances when a measure, though empirically valid, would not be considered conceptually valid.

What conclusions, if any, can be drawn from present-day research in juvenile delinquency? When will we approach the time when theory development can keep pace with the methodology of juvenile delinquency research? Westie and Turk (1965) may very well have the reasons for the lag between theory and methodology in these observations:

1. Failure of the proponents of particular theories to subject their theories to empirical test.
2. Confusion of theoretical interpretations with empirical generalizations.
3. The selection of particular theoretical interpretations from the array of possible interpretations without systematic consideration of alternative, and frequently contradictory, interpretations.

We must continue to improve our research techniques. We can no longer find justification for separating basic research from applied research. It is my personal belief that both basic research and applied research are most vital to contemporary juvenile delinquency research. Some specific areas that are in need of additional sophistication are as follows:
1. Collection of reliable data
2. Longitudinal studies
3. Continuous validation of predictive instruments
4. Early identification of juvenile delinquents
5. Studies in probation success and failure
6. Validation of community programs
7. Development of predictive instruments predicated on interdisciplinary perspectives
8. Need for more cross-validation studies
9. Improvement of clinical research
10. Decision-making process as related to correctional rehabilitation.

I would like to conclude this paper with some appropriate observations made by McDavid and McCandless (1962) concerning the status of present-day research in juvenile delinquency:

It is doubtful if juvenile delinquency has ever before received the thunderous attention it is now getting. From sources ranging from pulpits to 'slick' magazines come opinions, exhortations, and counsel, but less frequently simple presentations of the facts on the issue. Professional people are subject to great (but justifiable) pressure to do research in both the etiology and therapy of juvenile criminality, and the number of man-hours invested in such research has increased in response to these demands. Unfortunately, however, it does not follow that the quality of the product parallels the quantity. An honest and self-critical examination of the net results of these many research efforts leaves the evaluator disappointed. The bulk of research in juvenile delinquency treats at a superficial level what many already knew or suspected. Critical questions, such as those about the developmental processes that result in delinquency, the 'triggering' of a given delinquent act, the mediation of social forces so that the result is delinquent behavior, have not only been unanswered, but rarely even asked.

LITERATURE CITED


