THE DESIGN AND EVALUATION OF EXPERIMENTAL COURT REFORMS

John B. Jennings

October 1971

P-4709

ABSTRACT

As new ideas are generated (and old ideas regenerated) for dealing with court congestion and other criminal justice problems, those that appear to be the most promising must eventually be tested in practice. However, it is generally far from straightforward to determine whether there is in fact any resulting improvement. Accordingly, this paper is addressed to considerations of experimental design, the selection of evaluation criteria, the measurement of court performance, and the assessment of the results of such a test. The principles introduced are discussed in terms of an ongoing evaluation of a major internal reorganization of the New York City Criminal court.

THE DESIGN AND EVALUATION OF EXPERIMENTAL COURT REFORMS

John B. Jennings*

The New York City-Rand Institute
New York, New York 10022

I. INTRODUCTION

Recently, a great deal of public attention has been focused on problems afflicting our systems of criminal justice in general and our criminal courts in particular. Matters of special and continuing concern to judges, lawyers, court administrators, and laymen include the congestion and delay often seen to strangle the judicial process, the challenge of assuring equal treatment to poor as well as rich, and the failure of sentences to rehabilitate those who are found guilty.

To cope with these problems, many reforms have been suggested:

- o administrative reforms, such as the automation of court records and information systems, the division of a court into units which perform specialized functions, or conversely, the elimination of specialization;
- o procedural reforms, such as the expansion of conditions under which a defendant may be released on his own recognizance, the encouragement of vigorous and continuing representation of indigent defendants by individual publicly provided attorneys;
- o sentencing reforms, such as narcotic addiction treatment programs, work-release programs, and the probationary release of convicted persons in the custody of community leaders or agencies.

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The New York City-Rand Institute or of the City of New York. However, the evaluation study referred to herein was supported by the City of New York. Papers are reproduced by The Rand Corporation as a courtesy to members of its staff.

This paper is based on a presentation made at the 12th American Meeting of the Institute of Management Sciences in Detroit, Michigan, October 2, 1971.

It has frequently been said that the wheels of justice turn slowly, and I think it is fair to say that the wheels of court reform turn even more slowly. In view of the complexity and gravity of the judicial process, this is often appropriate. However, as a result, many of the potentially most effective reforms remain untried. At the same time, many of those that are tried are implemented in such a way and under such conditions that it is not possible to determine whether they have been successful or not.

One means of alleviating this two-sided problem of separating the promising from the not-so-promising reforms is the use of carefully controlled and carefully evaluated experiments. This technique, which forms the heart of the scientific method, is used all too rarely in our attempts to deal with our social problems. The advantages of experimentation in the context of court reform are several:

- o innovative approaches to the administration of justice can be tested without the risk of upsetting the complex and delicately balanced operations of the court as a whole;
- o reforms which are found to be unsuccessful in practice can be easily terminated;
- o changes made on a provisional and small-scale basis tend to generate less automatic opposition on the part of special-interest groups, as well as on the part of the bureaucracy itself;
- o successful experimental programs form a natural starting point for the phasing-in of complex procedural changes.

One serious question which should be raised in connection with experimentation in courts of law is the extent to which such experimentation may conflict with the Constitution's guarantee of equal protection to all. This question has been considered at some length by Zeisel.* In essence,

---

*Zeisel, H., "The New York Expert Testimony Project: Some Reflections on Legal Experiments," Stanford Law Review, Vol. 8, No. 4, July 1956.

his conclusions are that, within substantial limits (including the requirements that any discrimination between groups of persons be made on a random basis and that existing rights not be abridged) experimentation should be permissible because (a) any discrimination imposed by the experiment is temporary and applied at random, and (b) the ultimate aim of the experiment is to eliminate any unfair discrimination by discontinuing either the original or the experimental procedures. Elsewhere, Zeisel* has reviewed past uses of experiments in connection with the reform of systems of criminal as well as civil justice.

In the remainder of the present paper, we consider the design and evaluation of such experiments, with reference to a particular experimental court reform of the administrative type.

---

*Zeisel, H., "The Law," Chapter 4 in The Uses of Sociology, edited by P. Lazarsfeld, W. Sewell, and H. Wilensky, Basic Books, Inc., 1967.

II. BACKGROUND INFORMATION

The New York City Criminal Court is a City-wide court with a branch in each of the five counties, or boroughs, of the City. The Court has jurisdiction over all adults and youths* charged with misdemeanors or lesser offenses, unless such cases are removed to the New York State Supreme Court via grand jury indictment. In addition, accused felons are preliminarily arraigned and granted preliminary hearings in the Criminal Court pending grand jury action.

The portion of this Court with which we shall be concerned here is the system of 16 Court "parts"** which process all adults arrested on felony or misdemeanor charges in Manhattan and whose cases are not disposed of at the initial ("arraignment") appearance in Court. During the first six months of 1971, approximately 30,000 such cases*** entered this portion of the Court.

Prior to February 1, 1971, this portion of the Court consisted of a fragmented set of specialized parts. Hearings were held in one group of parts and trials in another. Within each group, the parts were further specialized according to the level of the crime charged (felony or misdemeanor) and whether or not the defendant was being detained in jail. It was not uncommon for a single case to appear in as many as half a dozen distinct parts, each time before a different judge, and with as many Assistant District Attorneys and Legal Aid Society lawyers.****

---

*A youth is defined as an individual between the ages of 16 and 19 years; an adult is an individual 19 years of age or older. Younger persons charged with criminal acts are under the jurisdiction of the Family Court.

**A Court "part" is defined as a courtroom and all the staff necessary to conduct the business of the Court.

***When used in connection with Court statistics, "cases" will refer to counts of Court docket numbers, a unique docket number being assigned to each defendant-offense combination.

****In New York City, indigent defendants are provided with attorneys by the Legal Aid Society, a private organization deriving support from the City as well as independent sources.

For some time it had been clear that the fragmentation of the Court embodied in this system served neither the ends of justice nor of efficiency. There was no clearly defined responsibility for each case--whether on the part of the Court, the District Attorney's Office, or the Legal Aid Society. As a result, it was relatively simple for defense and prosecuting attorneys to delay, as well as to "shop" for a suitable judge, by having a case transferred from one part to another. As a further result, the number of Court appearances and the length of time required to dispose of cases had become unduly large.

Gradually, there developed fairly widespread agreement among judges, Court administrators, and lawyers that some form of "all-purpose" parts, in which the two stages of post-arraignment processing (hearings and trials) would be combined, and in which all types of (felony and misdemeanor) cases would be mixed, was desirable. Thus, it was generally agreed that the trend towards specialization in search of efficiency had run its course and that the time had come to reverse directions.

In spite of this agreement on direction, however, there was considerable disagreement concerning the precise structure into which the "all-purpose parts" should be organized. The two primary alternatives under consideration were a "simple" All-Purpose Part System, and a Master Calendar System.

In theory, both systems are based on the all-purpose concept of Court operation. In the "simple" All-Purpose Part plan, a single part performs all the functions required in the disposition of its cases. In a variant of this plan, a pair of "All-Purpose Parts" act in concert. One (the "calendar part") handles administrative matters, the control of the daily calendar, and the disposition of cases by dismissal or guilty plea; the other (the "backup part") conducts hearings and (non-jury) trials required for cases called in the calendar part (jury trials are conducted in separate parts).

In the Master Calendar plan, the calendar part/backup part plan described above is modified as follows: First, a variety of activities which cause the judge in the calendar part to be idle or to act only in an administrative or clerical capacity are removed from the courtroom

(or eliminated altogether), thereby freeing a substantial amount of his time for increased judicial activity. These activities include conferences between the defense and prosecution, the assembly of the various parties required in each case, and the granting of adjournments (continuances) in specified situations where both sides consent. Second, the freed time is utilized by increasing the number of cases calendared for this part each day. Third, additional backup parts (to give a total of four) are added to handle the additional hearings and non-jury trials and to provide jury trials.

A grant from the Law Enforcement Assistance Administration for the purpose of operating a Master Calendar System on an experimental basis in the Manhattan Criminal Court was applied for and received, and the Master Calendar operation (usually referred to as the MAP--for Master All-Purpose--system) was begun on February 1. At the same time, however, the Court administration, feeling certain that the simple All-Purpose System represented a significant improvement over the existing system, converted the remainder of the Manhattan branch of the Court, as well as the four other branches, into simple All-Purpose Parts. The New York City-Rand Institute was selected to conduct an independent evaluation of the Master Calendar Project.

Since the focus of this paper is not on the precise nature of Court operations but, rather, on the design and evaluation of the experiment, we now turn our attention in that direction.

III. SOME CONSIDERATIONS OF EXPERIMENTAL DESIGN

Since the purpose of an experiment is to test one or more hypotheses, the experiment should be designed from the start in such a way as to facilitate the evaluation of the results. Specifically, several steps must be followed:

- o The hypothesis(es) to be tested must be formulated explicitly.
- o The criteria by which one is to accept or reject the hypothesis(es) must be specified.
- o The experimental operation of the Court must, of course, put the hypothesis(es) to a test.
- o If more than one hypothesis is to be tested, the experiment must be arranged in such a way that all the factors involved are systematically varied so as to enable the separation of their combined effects. Similarly, the experiment must be designed so as to facilitate the separation of the "true" effects of the experimental procedures from transient and extraneous effects.
- o Means of acquiring and analyzing the data necessary to determine the performance of the Court according to the selected criteria must be established.

In terms of the above-listed stages of proper experimental design, the reform of the New York City Criminal Court was a poorly designed experiment: the hypotheses or objectives of the experiment had not been explicitly stated; the criteria by which the experiment was to be evaluated had not been identified; a wide range of changes were made in Court operations with no provision for testing their effects independently; there was no unchanged portion of the Court which could serve as an experimental control to assist in the isolation of transient and extraneous effects; independent evaluation of the experiment was not begun until half of the initial five-month test period had elapsed.

As a result of these deficiencies, the evaluation of the experiment had to be limited largely to an overall comparison of the two currently operating systems: the Master All-Purpose System and the simple All-Purpose Parts. Determination of the specific contributions of individual component reforms embodied in these two alternatives could not be made with any confidence.

It should be pointed out that the experimental design deficiencies noted above are by no means unusual in the context of the reform of social institutions in general and courts in particular. However, they point to the importance of bringing the evaluator into the planning stages of a reform so that the limitations imposed by the design of the experiment can be minimized.

IV. EVALUATION OF THE MASTER ALL-PURPOSE AND SIMPLE ALL-PURPOSE SYSTEMS

Selection of Performance Criteria

Although the selection of criteria for the evaluation of an experiment is normally guided by the objectives of the experiment, several factors dictated that, in the present case, the criteria be broad enough to reflect most aspects of Court performance: First, the objectives of the experiment had been specified in such general terms as "making more effective use of the Court's resources." Second, the administrative changes introduced were of such a variety as to affect many aspects of Court operations. Finally, since the pre-February 1 system had been totally eliminated in favor of several simultaneous experiments, it was difficult to identify a narrow set of performance measures which would reflect the most important changes.

The 22 criteria which were selected are presented in Table 1, divided into the following areas:

- o Factors affecting the quality of justice in the Court.
- o Efficiency with which cases are processed.
- o Burdens placed on participants in Court proceedings.
- o Peripheral criteria.

These criteria were selected after consultation with members of the Court's administration and other interested parties. We believe that they represent a fairly complete compendium of the most important measures for evaluating the effects of administrative court reforms.*

---

*Note that while a number of the "efficiency" criteria reflect Court operating costs indirectly, an explicit consideration of such costs was consciously excluded from the list of criteria by prior agreement with the Court's administration.

Table 1

CRITERIA FOR EVALUATING THE MANHATTAN MASTER CALENDAR PROJECT

Quality of Justice

- o Amount of individual attention given to each case
  - Number of cases calendared per part per day
  - Number of cases handled per Assistant District Attorney and per Legal Aid attorney per day
  - Time spent per case-appearance
- o Nature of the dispositions rendered
- o Speed with which cases are processed
  - Number of appearances per case, duration of cases, and time between successive appearances
- o Quality of prosecution and defense representation
  - Continuity of defense representation
  - Subjective ratings of the effects of any changes in the tactics or procedures of counsel and judges
  - Subjective ratings of the quality of representation and the quality of justice
- o Balance of power in the courtroom
- o Understandability of Court proceedings

Processing Efficiency

- o Utilization of judicial time
- o Rate at which cases are disposed of
  - Number of cases disposed of per part per day
  - Number of cases disposed of per Assistant District Attorney and per Legal Aid attorney per day
  - Fraction of calendared cases disposed of
- o Loss of defendants through bail- and parole-jumping
  - Fraction of cases in which bench warrants are issued and executed
- o Speed with which cases are processed
  - Number of appearances per case and duration of cases
- o Causes of adjournments and delay
  - Reasons for adjournments
  - Fraction of cases requiring multiple calendar calls and the reasons for multiple calendar calls
  - Subjective assessments of efficiency and wasted time

Table 1 (continued)

Burden on Participants

- o Number of Court appearances per case
- o Burden of Court proceedings on the Department of Correction
- o Subjective assessments of burdens and conveniences

Peripheral Criteria

- o Job satisfaction of Court personnel
- o Dignity and decorum of the courtroom

Collection of Data

In order to obtain the information required to determine the effects of the Court reorganization on the selected performance criteria, three primary sources of data were utilized:

- Case-history information obtained from Court records: A complete history of Court appearances was collected for samples of cases processed through the Master All-Purpose, the simple All-Purpose, and the pre-February 1 systems (approximately 2,700 cases in all).
- Activity data obtained from Court calendars and statistical reports: These data were used to determine such measures as the numbers of cases calendared and disposed of per part per day, the numbers of bench warrants issued and executed per part per day, the number of cases transferred per day between Court parts.
- Observation, interviews and questionnaires. Those kinds of information which were not available from any existing records were obtained through direct observation of more than 140 part-days of courtroom activity, and by interviewing and administering questionnaires to selected judges, Assistant District Attorneys, defense attorneys, defendants, and Court employees (more than 40 in all*).

For most of these sources of information, the analysis required to convert the raw data into the appropriate performance measure was straightforward. Two criteria which cause some difficulty, however, are the average number of appearances required to dispose of a case and the average duration of cases processed through the portion of the Court

*The original design of the evaluation included plans for much more extensive interviewing of those participants in the judicial process who are rarely consulted in the design of court reforms--e.g., defendants, police officers, correction officers, witnesses, jurors, the "public," and probation personnel. Unfortunately, a major portion of this effort had to be diverted to the collection of quantitative data which was to have been collected and provided by the Court itself.

under consideration. The difficulty arises from the fact that the focus of the evaluation was the five-month period from February 1 through June 30. Clearly, many cases which began during that period had not been completed by the cutoff date, and their complete appearance-history could therefore not be observed. The method developed to calculate the required performance measures--based on the principle of "maximum likelihood estimation"--is described in the Appendix to this paper.

Summary of Findings

The list of evaluation criteria presented earlier is in many respects quite similar to the range of factors with which judges and court administrators have traditionally concerned themselves in considering court reforms. However, their sources of information are typically informal and anecdotal in nature. Never before has there been available as systematic, reliable, and comprehensive a basis for evaluating court performance.

Although many of our conclusions are still tentative, it appears from the first five months of operations that the central hypotheses of the Master Calendar experiment are valid ones: that many tasks now performed by judges can be transferred to administrative and clerical personnel, that much of the activity which wastes the time of judges can be removed from the courtroom, and that the effectiveness and efficiency of the Court can thereby be improved.

For example, it was found that by preliminarily assembling the various participants in each case before the case is called in the courtroom, and by requiring that the defense and prosecuting attorneys confer with each other outside the courtroom, rather than before the judge--as had been the practice--approximately 20 percent of a judge's available courtroom hours could be saved (relative to the All-Purpose Parts). This saving, together with such factors as the availability of backup parts capable of providing immediate jury trials, has made it possible to increase the capacity of each judge in the system to dispose of cases by

some 25 percent. As a result, the backlog of cases is being rapidly reduced, the length of time required for cases to move through this portion of the Court is 15 percent shorter than that in the simple All-Purpose portion of the Court, and the associated detention population in City jails has been correspondingly reduced. At the same time, Assistant District Attorneys and Legal Aid Society attorneys are able to devote greater and more continued attention to each case.

The primary cost associated with these gains is the increase in personnel required to staff the Master Calendar System--additional Assistant District Attorneys, Legal Aid Society attorneys, and administrative personnel. It is still too early to make a detailed comparison of all the costs and benefits. However, based on the results of the continuing experimental phase of operations, the Court administration will eventually decide whether to extend the Master Calendar plan throughout the Court or to revise or abandon it.

V. SUMMARY

The theme of this paper has been that court administrators, in seeking to improve the operations of their courts, should make much greater use of carefully controlled experiments. The advantages of this approach include the following:

- o innovative approaches to the administration of justice can be tested without the risk of upsetting the complex and delicately balanced operations of the court as a whole;
- o reforms which are found to be unsuccessful in practice can be easily terminated;
- o changes made on a provisional and small-scale basis tend to generate less automatic opposition on the part of special-interest groups, as well as on the part of the bureaucracy itself;
- o successful experimental programs form a natural starting point for the phasing-in of complex procedural changes.

This approach has been illustrated with a major experimental effort now underway in the New York City Criminal Court. This project, though imperfect from the point of view of experimental design, is yielding much useful information to the administration of this court.

Appendix

METHOD OF ANALYSIS OF THE NUMBER OF APPEARANCES PER CASE AND THE DURATION OF CASES

Because the focus of this evaluation was on the February-through-June period of this year, the case-history data collected for the purpose of determining the post-arraignment number of appearances per case and the post-arraignment duration of cases had to be cut off at June 30. Cases which were still pending on that date could be classified only as "incomplete." Since it was desired to include data on cases which had begun throughout the period, the "incompletes" would include some cases which had had little if any chance to be disposed of.

Clearly, if one had simply discarded all cases which were incomplete, one would have biased the results in favor of the shorter cases. The alternative of assuming (conservatively) that all incomplete cases would have made one additional appearance -- which has been used in a number of studies -- would also have biased the results, particularly if one were to compare sets of cases containing different fractions of incomplete cases. A third alternative of taking only cases which had begun early in the five-month period would have reduced the problem but would not have solved it, and would have made it impossible to utilize data pertaining to that period of operation which reflected the effects of refinements made in the course of the project.*

A final alternative developed for this evaluation avoids all of these difficulties: it utilizes information pertaining to the entire period of operation and yields the frequency distribution for the number of appearances or the case-duration which is most likely to have produced the pattern of appearances or case duration observed for both complete and incomplete cases. In addition, it provides a method of making statistical comparisons between two sets of cases.

The method, which is an application of what is known as maximum likelihood estimation, is illustrated below for the number of appearances per case.

---

*An alternative at the other extreme -- that of looking only at incomplete cases and estimating total appearances, etc., from the number already made -- has its own difficulties and also would have required the discarding of information concerning the early operation of the project.

Suppose one has a sample of 100 cases, of which 90 had been completed by June 30, while 10 were still incomplete. Further, suppose that the numbers of post-arraignment appearances made by the 90 complete cases are as follows: 50 cases were disposed of in one appearance, 24 took two appearances, 10 took three appearances, 4 took four appearances, 1 took five appearances, and 1 took seven appearances. Of the 10 incomplete cases, suppose that 2 had made one appearance prior to June 30, 6 had made two appearances, 1 had made four appearances, and 1 had made five.

While the maximum likelihood formulas must be determined mathematically, they turn out to have an intuitive interpretation: Of the 100 cases in question, 50 were completed in one appearance, while the remaining 50 are known to have taken longer (for the complete cases we know exactly how many appearances were made, but for the incomplete cases we know only that they required more than some number). Thus, the most likely estimate of the fraction of all cases which would be completed in exactly one appearance would be 50/100, or 50 percent. We may now discard the 50 cases which took one appearance and the two which were incomplete after one -- they can tell us no more. Of the 48 remaining cases, 24 are known to have required exactly two appearances, while 24 are known to have taken more. Therefore, <u>of the unassigned 50 percent</u> of the frequency distribution being derived, the most likely fraction to require exactly two appearances is 24/48, or 50 percent. Thus, 50 percent of 50 percent, or 25 percent of all cases are estimated to require two appearances for disposition. This procedure is then repeated until the entire distribution is determined.* One may then calculate the estimated average number of appearances per case -- 1.97 in this instance.

The distribution produced in this fashion may then be subjected to statistical comparisons with other distributions in order to determine whether any

---

*When the sample under consideration includes an incomplete case which has made as many appearances as or more than the largest number observed for a complete case, there will remain a (small) unallocated portion of the estimated distribution. This eventuality does not affect the statistical comparison of the distribution with another by means of the likelihood ratio test. However, in order to obtain an average value, one must arbitrarily assign this remainder to some number of appearances. We have chosen to assign it to the next higher number above that already observed. Note, however, that, if this procedure is required at all, it applies to only a fraction of the incomplete cases and therefore introduces much less distortion than would the practice of simply assuming all incomplete cases would be completed in one additional appearance.

differences are significant. The vehicle for such comparisons is known as the likelihood ratio test.* This method, with only slight modifications, may also be used for case-durations.

---

*This test is similar to the more common chi-square test of significance which cannot be applied here because of the need to distinguish between complete and incomplete cases. The likelihood ratio test is described in most books on mathematical statistics: for example, Rao, C.R., Linear Statistical Inference and its Applications, New York, John Wiley, 1965, pp. 349 ff.