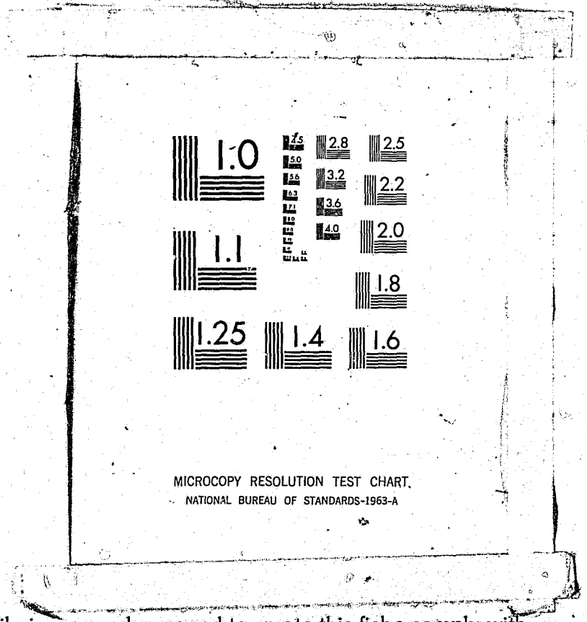

MF-1

The Measurement of Judge-Time and The Evaluation of Judicial Performance: Reducing the Discrepancies

Samuel Krislov

University of Minnesota

This study was undertaken with support of a grant from the ...onal Institute of Justice, 79-NI-AX-0077. I am also ...ebted to the Guggenheim Foundation for support during ...-80 for a larger project in which this is embedded. ...view expressed are solely my own. Linda Archer assisted in ...research.

19498

I. BACKGROUND

The strong tradition of judicial independence in American society has been an impediment to clear understanding of the judge's role. The legitimate (and praiseworthy) reluctance of our judges to go beyond the written opinions in justification of their actions has (at its extreme) shaded into fierce unwillingness to be accountable for performance or even courtroom manner and demeanor. The truth is that any form of accountability can be abused to diminish independence. The boundary between accountability for workload and the substantive outcome is not a precise, well-built Maginot Line of fortresses but is rather a straggly indistinct no-man's land. Judges could once rely on this to create an autonomous life-style. The public and its representatives have relied on peer-group pressure to deal with most problems of effort or conduct, preferring the risks of occasional abuse by judges who are independent over the perils of abuse of the judges' autonomy by other authorities. That decision has meant that much information available in other areas of political life is (to use Theodore Becker's phrase) hidden behind the "velvet curtain" of the Judiciary.

Significant changes in the environment of courts have altered that balance. Attitudes are in transition on those questions, almost as much on the part of those on the bench as off it. As judge-time becomes more-and-more a scarce social resource, especially relative to demand, its allocation is

no longer seen as something best handled by the sovereign individual judge. With priorities, and even deadlines, on litigation now commonly mandated by the legislature, judges feel external pressure on their disposition of time. They may actually seek planning as a counter-weight to legislative control as well as a rational means of dealing with the central problem of overload.

As techniques have been developed to study and aid executives doing non-repetitive and creative tasks, some of the stigma of using efficiency control methods disappears. Techniques evocative of assembly-line control are regarded as demeaning, but computer-product flow chart models have a comforting and even ego-enhancing aura. As the judges have made their peace with management technology, they have also grasped some basic facts about it. They too, know that ambiguity abounds and seemingly precise results rest ultimately on arguable assumptions. This lack of true precision actually makes those management tools less threatening, since the conclusions drawn can be easily contested. The central importance of methods used and assumptions made is such that it is wise not merely to accept but rather to pre-empt these approaches, housing the technicians firmly under one's own control. The increased attention of the public to judicial matters has been also the result of their more active role. Their increased importance in social life has made judges newsmakers and celebrities.

What they do and how they do it assumes social importance beyond that of the past. (This, too, has increased the pressure for court courtesy and limits imperious use of the judges' awesome power. As with monarchs, so with judges. "Redress of grievances precedes supply.")

As Courts have found it necessary to seek augmented resources on an accelerating scale, justifications have required better and more impressive data. Attention to wise use of available resources has been a prerequisite for successful additional aid. The Courts have therefore developed, and made available information formerly denied not only to the public but even to other members of the judiciary.

The records made available are purposive and teleological. The methods reflect the fact that the data are instruments of argumentation. Still, as we shall argue, they provide some rough guides to reality. Their accuracy can be tested against other more limited findings. And in turn the available material permits other more precisely informal inquiries.

II. MEASURING JUDICIAL EFFORT

The disposition of judicial time is both of theoretical and practical significance well beyond the interminable studies designed to identify the optimal allocation of judges. On the theoretical level, it bears upon Friedman's strongly argued conclusion that Courts increasingly are ratifiers of decisions and thus less and less relevant adjudicators of meaningful

disputes (Friedman and Percival, 1976). But this conclusion rests largely on examination of raw case numbers. The question of what time-fraction is allocated to such matters also affects the question of what the trend toward the increase in absolute numbers of such functionless functions portends. The argument to be developed here is that these matters are increasingly crammed into lesser and lesser [illegible] spans of time--and less valuable and skilled judge-time, at that. They have become both superfluous and superabundant. Our fundamental understanding of what is happening to American courts, what their social role is, depends in large part, on our understanding of the pattern of expenditure of judge-time.

Such findings have practical implications as well. An example is the current effort to increase judges' bench time. Critics suggest bench time is low largely because of lack of diligence. It has been shown, for example, that correlations of time spent on the bench with esteem by the bar are strikingly positive. The more time the better the evaluations. It is argued, therefore, that greater effort can be asked of all judges without risk of deterioration of judicial performance or quality of effort. (It is, of course, questionable whether the finding that quality and effort are positively correlated is a universal phenomenon or one that holds at a particular time for a particular court. But certainly a relationship in a voluntary situation will be altered by time

required of judges, and service imposed upon them.) It has been suggested that the fact that judges in lower courts generally have higher bench time, and that judges temporarily assigned or promoted to higher courts initially amass more decision time, indicates that indifference and laziness set in. However, there are alternative explanations of this behavior. Temporary persistance of inappropriate work patterns until adjustments take place, is another explanation.

What is the source of evidence on these matters? The most accessible data are found in the position-allocation studies already referred to. There are also studies of other sorts. Thus Ryan et al., (1980) sent direct questionnaires to judges. A third type of study is in-court observation and recording of time by Court Watchers. Each method has its advantages and costs.

Simple mail questionnaires tend to produce idealized results, probably not because of deception on the part of respondents so much as; (a) self-selection and, (b) self-serving recall. Ryan et al. judges (p. 26) report they work about eight hours a day (8.7-9.2 including lunch), a figure roughly one hour (or about 15%) more than any workload study ever completed.

Workload studies are in some senses overly influenced by the objectives. By and large the effort is to magnify burden and to suggest the need for additional help. These studies generally elicit good cooperation, as the purpose is seen by judges as possibly contributing to their collective good.

This cooperation may well produce some reliable information on other, more neutral, aspects of the material, particularly the distribution of judge-time across categories. In most instances, no obvious advantage would inhere to any individual court sufficient to affect the data. The methods of recording and observing are sufficiently well-defined that distortion requires almost willful deception, as opposed to roseate recall.

Finally third party recording, as with Court Watchers, is probably the most precise court-time measure of all. (Staff recording of time introduces some observor error due to press of duties, and, perhaps, even bias in directions the staff might expect the judges to desire.) The weakness of third-party observors is their inability to observe non-court service and even the danger of missing low-visibility court functions (e.g. signing of court papers during a recess, or court conference.)

Yet these studies tend to show judge work days at or about the same level. As might be anticipated broad questions about "typical" days draw out the most generous recall (8.7-9.2 hours inclusive of lunch), Caseload studies diminish these claims. In Washington the Superior Court judges averaged a shade less than 7 hours (6.7 hours plus 1.2 hour lunch). (Washington, 1977, p. 16). California figures are higher but not remarkably different from this (7.7 hours of work in Superior Court). (California, not dated p. 6). New York Court Watchers indicate sharply less bench time (under four hours, and less than 6 in court time). The study is of criminal courts, however, and is

influenced by some highly extreme examples. (Three judges averaged one hour-and-a-half per day in the study, for example). (Economic Development Council, 1976, pp. 5, 9).

The desire of most official calculators-usually paid consultants-to make the judges look good is found in the casual assumptions of most studies as to the number of days worked in a year. In Washington the work year is calculated at 221 days, (based upon 250 possible workdays minus 20 days of vacation, 5 of illness and 4 for conferences, etc.) In California, the municipal courts judges are assumed by the statisticians to work 216 days, (assuming 11 holidays, 21 vacation and 8 sick days and 5 conference days). This figure is at times also calculated as 215 days. In Kentucky, 215 days are hypothesized, and the report even suggests that the high number of part-time judges makes empirical results inapplicable. (Kentucky, 1976, p. 14).

In point of fact, the number of reported work days is quite different in California and closely approximates the District of Columbia's informal assumption of 200 days of bench time. Specifically in a 1976-77 Superior Courts study, vacation and sick leave averaged 28 days, while conferences and civic duties averaged 22 days. (In 1973, it was only 17.5 days for vacation and illness, and 6.5 for conferences. However 17 percent of all judge time went to "non-case" and "administration" compared to less than ten percent in 1976-77, suggesting total

activities stayed generally at the same level, but were simply classified differently. Municipal court judges were found to work only a bit more in 1977. They averaged 25.8 days for vacation and sick time and 20.8 days for conferences. (Earlier surveys used non-comparable categories but it is likely these figures represent a slight reduction in non-bench time). In short, California judges actually worked about 200 days or 10 percent less than assumed. (California, 1977, Tables 5, 6).

A major loss of judge time results from the surprisingly high vacancy rate. New Jersey found 5,235 total days lost due to vacancies in 1978-79 (or 9% of all available judge time) but even this pales beside its fourteen percent loss in 1977-78. (New Jersey, 1979, supplement, XXVIII). Whether New Jersey is typical is difficult to assess. However some clues as to frequency of turnover can be gleaned from Ryan et al., . They characterize judicial selection in 29 states as "purely" elective in nature, 8 as appointive by the governor and the remainder as legislative or hybrids. In point of fact, however, 45.1% of sitting judges were initially appointed by the governor and only 43.6 of their sample elected. (In California, technically an elective state, fully 88 percent were initially appointed) (Ryan, et al., 1980, p. 124). With that huge a turnover rate, even a small delay in the appointive rate would produce large losses in available judge-time.

Of the seven or more working hours most sitting judges averaged, roughly half are usually bench time. Washington

Superior Court judges average 739 hours a year. California loads were first estimated at 750 and later at 1000 hours and New Jersey judges have consistently averaged about 900, (slightly higher in 1978-79, less in 1977-78). The bottom-line is that judges spend perhaps four hours a day on the bench, though the variation is considerable from judge to judge and system to system.

Many factors contribute to such differences. One of them, of course, is the attitude toward what is desired and emphasized by the system. An apocryphal story of New Jersey, which has for years required accounting of judge time in 15 minute intervals, [illegible] epitomizes concerns about possible negative effects of accountability. The tale is of a judge who at 11 a.m. has before him a defendant who wishes to plead guilty. Rather than report a full hour of non-bench time, the judge insists on empaneling a jury, and in engaging in other busy work connected with a trial. Finally allowing the lawyers to approach the bench, he accepts the plea on the final stroke of twelve-wasting an hour for a score of people including himself, but recording more bench time.

In California where actual trial time and later total bench time had been the basis of all their case weights, public criticism by a "hot dog" governor for their "lack of effort" stimulated use of the more subjective "all case-related time" as the standard for computation. This is also theoretically a more

reasonable definition (though more difficult to measure) and has permitted continuous increases in the theoretical work year. The increase that has been found certainly is simply (or at least largely) an artifact of the method used. In short, political and public considerations render some of the results non-comparable and even misleading.

Table 1 illustrates

As Table 1 indicates, as broader definitions were used more "time" was expected, though in fact little more, or even less, effect was actually involved.

A second relevant factor seems to be court size. Smaller courts produce fewer trial hours per judge. Some of this is clearly attributable to the minimum necessary administrative load (record keeping for example) required of any court, when divided among a smaller number of judges. This is the classic problem of fixed overhead costs as share of a growing enterprise. While courts do not provide sufficient information to establish fully what else is involved, it is clear that judicial councils and their statistical advisors believe other factors are operative. They conclude that civic and symbolic responsibilities also consume a relatively fixed, relatively greater share of small-court time. Then, too, smaller courts are rural courts. Judges may (as in Washington) have to travel between courts in different locales, or as in California, may be assigned to other courts to fill out a full load. Travel time obviously detracts from bench time and may significantly increase administrative load as well.

To date analogous claims by the Los Angeles Court that its administrative problems--it has a separate criminal court, and also maintains local courts at some distance from the downtown court--should result in a different work load total have not been accepted.

The argument that large courts create additional administrative load--in the loss of communication, required planning and the like--tends to be rejected by central planners. Generally there are more undersized and below average work-load courts; rationalization of their existing court power is both easier to accept and politically more expedient than inventing privilege for large jurisdictions. Larger court systems have advantages to compensate, particularly the efficiencies of para-judicial help not easily open to tiny courts. So those arguments fall not on deaf, but less attentive, ears. Los Angeles County (and San Diego now as well) have their own disposition weights for each type of case. But the total work time expected per judge is still similar to expectations for courts many times smaller. In 1977, Los Angeles had 239 judges. The Judicial Council has to date not officially adopted those weights. The range of court size has increased since the county has been permitted to add new judges after many years of frustration.

A third finding pertains to distributions of time among nominally similar courts. Perhaps because of the desire to effect uniformity, case weight reports do not provide much evidence on this point. An exception is the Florida case load

study (1977, p.88). This establishes the fact the percentage of time expended in dealing with cases in county courts varied incredibly--from a high 90 percent to a low 15 percent. While those counties with extremely high off-bench ratios are small counties in terms of litigation, those with very high case-related percentages were of all sizes. Dade County--with over 20 percent of all judge-time in the state--recorded 70 percent of judge-time as case-related but so did Pasco County, with one-thirtieth Dade County's case load. Circuit courts varied much less, spending from 14 to 41 percent on non-case related time. This disorderly pattern reinforces the suspicion that the differing times allocated to different size courts is a rationalization for existing judge distributions, designed to allow planners to produce politically acceptable results. The process is, after all, as the California shifts indicate, a form of "finagler's constant." ("Finagler's constant" is a standard device used by scientists and pseudo-scientists faced with unpleasant results to mitigate or eliminate bad news).

The details of the Florida report further indicate how diverse courts can be. For circuit courts, case weights (in real time-minutes), varied in different circuits by categories, but also in unpatterned ways: (See Table 3).

The high and low load circuits showed little patterning. Similar variation appeared as to non-bench time. While the breakdown of non-bench time into correspondence, travel, general

research, conferences, and administration is extensive, so is the variation. Gulf County recorded more ex officio hours than Dade but one-fortieth of the latter's administrative hours and about 10 percent of the other categories. Hillsborough Court reported twice the conference minutes Dade did. Other variations of even more complex nature exist. (Florida, 1977, pp. 95-96).

Some other attributes of bench-to-non-bench time emerge. (1) "Higher Courts" have higher non-bench times. (2) Civil Courts have higher non-bench ratios. (3) Ratios seem to be slightly affected by availability of administrative and research assistance, though whether proportionate absolute gains in judge time occur is not clear. (4) Some non-accountable minutes occur in every system.

In general, appellate courts have more complex cases and/or more need to justify their actions. As expected, more effort is expended on research time and other case-related work than in lower courts. Courts of more complex jurisdiction experience the same phenomenon; presumably this occurs not because the process is different (as with appellate courts) but because of the type of problem involved. While much of this difference is captured in different disposition weights presumably some is unrecordable. The judges and statisticians assume those patterns reflect the way work flows, while as we have noted, critics suggest judges become complacent as they rise. To date no one has tested the rival hypotheses. (An obvious one is that the

older judges, predominant on higher courts, or that more scholarly judges, have different types of work habits. These may be non-functional but still unavoidable habits, consuming time as a cost of system design, to please the type of person wanted.)

Criminal cases appear to be a special instance of this phenomenon, generally requiring less off-court work. Trials, however, and other bench time are extensive. They do not as a rule require involved legal study and the evidence is often weighed by the jury in any event. The difference in judges' allocations are well-depicted in the following Washington state study. Perhaps because of the intensity of trial activity, criminal judges amass more bench and total time and resemble minor-court judges in many ways.

The New York City data on criminal courts indicate research time of the judges is virtually nil. The availability of a full-time research clerk for each criminal judge reinforces the sharp tendency to simply conduct the trials and rely on briefs. In California where research assistance is more tightly rationed, significantly more judge-time is devoted to research. Similarly, the providing of administrative assistance to California courts seems to have some impact on time allocated to this function, though close examination of the figures suggests this is less impressive than the raw data suggests. (In point of fact, as we have noted, most of the change is an artifact of classification). Obviously, non-judge time can be employed in non-

judging functions of courts. Whether or not this is efficient depends upon the degree to which judges use the freed time effectively. As with other efforts to substitute paraprofessionals in professional functions the results have been poorly studied, if evaluated at all.

In assembling this data, compilers have had difficulties in assigning all judge-time to their categories. For example, the Washington report assigns time as "case-related" and "non-case related" under specific headings, chalks up calendar time (assigning it proportionally to civil and criminal cases) then totals "recess time" and still has residual categories at each step and at the end.

This is perhaps not surprising inasmuch as what is being pigeon-holed is complex professional activity. Earl Johnson and his associates report in a forthcoming compendium on judicial statistics (Oceana, 1981) that lawyers are able to allocate only 70% of their time as billable to clients. Obviously, the problems are analogous but not precisely on point.

III. JUDGES, SPECIALTIES AND COURT TIME

The differentiation in time distribution between different types of cases assumes particular importance as specialization takes place within the court system. Traditionally, judging is the last refuge of the professionalized amateur, the highly trained generalist. This remains reflected in the distribution of judges' functions as found by Ryan et al., in Table 5.

This suggests a fair degree of specialization, sometimes legislatively prescribed but increasingly achieved by internal administrative decision. Specialization is a response to the pressure for efficiency. It can lead to "assembly line justice" but it can also be a rational way of differentiating routine and non-routine functions as well as more complex from simple cases.

Suggestive, too, are Ryan et al.'s factor-analysis of judges' ratings of the importance of different functions:

> (1) High on reading files, keeping up on law, prepare decisions, general administration, moderate settlements.
>
> (2) Preside at jury trials, negative loading on others.
>
> (3) Socialize and discuss cases, both with attorneys.
>
> (4) Plea negotiate, waiting time management. (1980, p. 31).

While the authors do not speculate, these would appear to parallel known breakdowns of judges: (1) the general-civil-trial judge, (2) criminal judge, (3) presiding judge and (4) the civil-negotiator judge. Larger courts find it advantageous to have specialists on motions and discovery, and to "farm out" probate or minor family matters to lawyers working as part-time judges or full-time Masters. Such differentiation, already further along than might appear on the surface, is likely to increase.

There are arguments that justice suffers as cases are treated as modules and each judge sees only a portion of each case, perhaps losing true insight. On the other hand, the English express some concern with our system where the judge can coerce unfair settlements by hinting worse will come in a decision. The English discovery and settlement stage is the concern of a Master, a sub-judicial specialist who aids the regular judges. This concern with justice as being affected by the machinery of judging is a non-trivial one and has properly led to conservative patterns. But some arrangements have proven themselves and more can safely be expected to do so as strains upon resources make "new departures"--especially proven and not so novel small efforts--welcome and easy steps toward economy and efficiency.

Those court functions which are primarily administrative, recording or verifying processes are prime targets for simplified handling by court officers acting in the name of judges. Probate and uncontested divorce, like uncontested traffic tickets, constitutes a large fraction of court cases but consumes a small (and increasingly smaller) portion of judge time. It makes little difference whether one establishes a separate administrative office to receive uncontested traffic payments or does so as an office within the court building. It is doubtful that those perfunctory operations crowd out the rest of the Court's agenda or appreciably diminish court ability to handle more contested

or more significant issues. Society, after all, also has the right to decide how and to what degree it will subsidize public decisions on private matters. Other issues can still go to private mediation when the parties wish to absorb the costs involved.

All of this cumulative evidence is less than compelling, but it suggests conclusions requiring and permitting more exact testing. Judging is a complex activity, not easily homogenized or measured. A good deal, probably a growing proportion, of judge-time spent is not decision time but is overseeing or rapport building or training time. This involves external activity that might not take place or might be altered by that supervision. Those functions, such as supervision of the growing corps of surrogate judges, are on the increase.

Within the core of judging itself, rational-thinking, and the link between precedent and policy requires care and concern. Clearly, more decisions can be exacted from judges. At all levels some judges verge on neglect of duty. But inherent in the process is also a need for considered judgment. The allocation of non-bench time for different types of cases strongly parallels the need for deliberation. Some recognition that they also serve who prepare seems called for. Excessive emphasis on bench time has dangers, even as the public properly puts forth its demands not only for accountability but also for its money's worth. In so doing, it must also be leary of losing the dollars by focusing too narrowly on the pennies.

IV. QUALITATIVE EVALUATION OF JUDICIAL PERFORMANCE

Closely related to quantitative evaluation of judges is the new interest in qualitative evaluation. The roots of the efforts are similar to those discussed earlier. But where the motivation on weighted case load is primarily to monitor available resources, direct evaluation is zeroed in on judicial performance and on the very person of the judge.

Presumably, the effort to evaluate [illegible] judges is an attempt to improve performance from incumbents as well as to secure better new judges. Evaluation can encourage judges to develop proper skills, or suppress unwanted qualities. It also may suggest to incumbents they would more appropriately fit other positions, or encourage the best to stay in office.

These purposes can best be served by evaluations that assess judges in their most significant functions and in reasonable approximation to the actual skills needed by them. But the discrepancy between what is evaluated and the skills needed by judges is striking. While this occurs in many aspects of life, the myth of judging has remained static while the shift to mediational and administrative needs has been growing. Evaluation, which is still more an idea than a reality, has begun with a limited, archaic view of the position. To be sure, wisdom in judging would remain the paramount skill sought even if the hours spent were a tiny fraction of judge-time, much as the test of the surgeon is ultimately in the use

of the scalpel. But the conspicuous omission of many other qualities is a clear weakness in the efforts that have been made.

Officially sponsored programs still remain sparse, but additional programs are under consideration in a few places. Semi-official and media efforts and legislature concern are all likely to provide impetus to new programs.

(1) The state of Alaska has undertaken an impressive pioneering experiment to enhance and bring genuine public opinion to the Missouri plan. All judges must be initially appointed by the governor for a period of three years from a list recommended by the Judicial Council. (That group is composed of three non-lawyers appointed by the governor, three lawyers elected by the bar, and presided over ex officio by the Chief Justice of the Supreme Court). In 1975, the legislature mandated the Council to evaluate each judge up for retention or rejection by public vote and to transmit that recommendation (through the lieutenant governor) in the official Election Pamphlet. The law does not prescribe any method of evaluation. The Council has chosen to poll attorneys, peace officers and jurors. While the Council's first negative report on a judge was publicly opposed by the police officers, and repudiated by the voters, its careful and faithful program is respected nationally, in spite of its own inevitable feeling of disappointment at its first test. (Rubenstein, 1977).

(2) Under the Home Rule Act of 1973, the District of Columbia Commission on Judicial Disabilities and Tenure has disciplinary power, including removal power over D.C. judges. In addition, it has authority to evaluate sitting judges for reappointment. If it evaluates a judge as "exceptionally well-qualified" or "well-qualified" the judge is automatically reappointed for 15 years. If the judge is rated as "unqualified" no reappointment is possible. If the sitting judge is evaluated as "qualified," presidential discretion comes into play. The Commission has seven members: one chosen by the President, two by the Mayor, two by the D.C. Bar, one by the City Council, and one by the Chief Judge of the U.S. District Court for the D.C.

As might be expected, virtually all judges have been found to be "exceptionally well-qualified" or "well-qualified." Nevertheless the Commission's process and product are impressive. Polls of their members are undertaken by several bar associations, on behalf of the Commissions. The judge submits a resume of accomplishments including samples of opinions. The Commission meets with the judge, and may well indicate areas of concern prior, during or after the meeting. Finally, it transmits to the President clear, lucid, comprehensive and well-balanced evaluations including candid criticism, and a rounded picture of performance in many aspects. (I volunteer this evaluation as a former member of the comparable Minnesota Commission on Judicial Standards and one familiar with the work of many such bodies). (D.C. Commission, various dates).

(3) New Jersey has a well-publicized program under development and an impressive but little known program in existence.

The states' Office of Administrative Law (acting partly on the recommendation of the Supreme Court's Committee on Judicial Evaluation and Performance) instituted two programs on January 6, 1979, involving both qualitative and quantitative evaluation. They emphasize external (non-agency) evaluation of work-product as well as objective work-measures. These evaluations are mainly for re-appointment but also are continuously useful since the officers involved can be removed from office for bad performance. While this is a program for what are called "hearing officers," it has attracted attention as a possible model for programs for judges having more independence of office and greater authority. (Rosenblum, 1980).

In addition the New Jersey Supreme Court has after careful study voted to develop such a system of evaluation for the regular courts. Under Chief Justice Vanderbilt New Jersey emerged as the progenitor of the unified-court system and the Chief Justice has long had authority to "call on the carpet" judges found to be dilatory or inadequate in some way. (Individual work sheets have often been the basis of such reviews.) As of this writing no implementation of the broader system is in the offing though staff has been recruited to develop the approach. The prestige of the Court, and the heavy publicity attendant upon its decision mark this as a significant step. The fact that Associate Justice Handler is a major force

nationally for such efforts, and the author of a leading scholarly article on the subject gives it additional impetus and national attention. (Handler, 1979,).

(4) The Colorado legislature has (1979) instructed the Judicial Planning Council to develop a system of evaluation. Such action raises questions about separation of powers, which the Council has side-stepped by proceeding on the question on its own th ough alluding to the legislative interest. The Committee on Judicial Performance established and issued an interim report in January, 1980, and later published a more extensive version of its recommendations in August. It suggests a Commission be established to develop standards and methods, including consideration of means to disseminate results. Its summary of other practices is quite valuable.

In short, officially managed evaluation is in its infancy, and is characterized more by concern and intent than execution. Semi-official or unofficial polls of the bar are however much more extensive. Those are generally well-publicized and are often issued just before reelections. They have been studied and the techniques evaluated and refined. Their influence upon voters varies in different locales.

A most careful effort has been that of the Chicago Council of Lawyers. The organization has used non-polling evaluation of Illinois judges up for reelection, but more elaborate procedures to evaluate the Northern Illinois Federal District Court. The latter was widely viewed as a weak bench--perhaps the

worst Federal court--at the time the polls started. Since removal or political defeat could not be the goals, sound methodology was seen as needed to provide complex moral pressure on the judges to improve and on appointing authorities to aim for higher caliber appointees. Senator Percy's high standards had the most to do with dramatic transformation of the bench but the Council (and others) believe they also played an important part in transforming it into a prestigious unit. (Chicago Council, 1976).

The media occasionally run lawyer polls but are much more likely to publish unquantified assessments based upon informal and unidentified sources. They are not conducted on a regular basis and tend to be vaguely remembered by the public and bitterly resented by the judges. They tend to be more result-oriented, and less process-oriented than bar or official efforts. Thus a Texas paper published ratings based upon conviction rates and median sentences without considering sampling issues and simply assumed homogeneity of criminal cases. While television is less likely to conduct such polls, the burgeoning city magazines have run a number of judge-evaluations. The highly-middle class that is the audience for such publications is after all highly professional, law related or law-conscious. Finally, the expanded law dailies or weeklies seem likely publications of such studies. The American Lawyer (1980) recently evaluated the best and worst Federal District

Judge in each circuit, a relatively easy and uncontroversial task in most instances. It seems likely other such assessments are in the offing.

The American Bar Association Committee on Evaluation of Judicial Performance of the ABA Criminal Justice Section presented the ABA in summer, 1980 with a recommendation to endorse the development of questionnaires and other measures largely based upon traits identified by studies by Dorothy Maddi of the American Bar Foundation, Guttermen and Meidinger for the American Judicative Society and Cynthia Philip for the Institute of Judicial Administration. These would be supplemented by objective measurements all designed to provide for rounded assessment. Additionally, work-product would be professionally assessed along lines pioneered by the New Jersey Office of Administration and the Legal Services Corporation and recommended in Justice Handler's thoughtful and impressive article. Together with a National Center for State Courts proposal currently under consideration at funding agencies, this betokens a strong interest in such efforts, particularly among the infrastructure organizations servicing the court system.

This history and outline could be elaborated upon; good early studies need updating. Similarly many studies of the emergent disciplinary Commissions could be summarized or expanded upon here. The older tradition of judges-dealing-with-judges was simply timid; one commission voted "public censure" of a

judge but to avoid embarrassment decided not to name the individual. With public non-lawyer members now common, newly structured commissions have played a more decisive role in enforcing standards--and have not feared going public in extreme cases.

The purpose of this discussion, however, is to focus upon the relationship between what we now know judges do and the expectations we have of them. By and large public disciplinary cases tend to involve extreme departures from justice or flagrant violations of established, usually codified norms dealing largely with proprieties. Our concern is rather with the qualities and characteristics of judges adumbrated by the evaluations. These have been, as noted earlier, rather conveniently summarized in the ABA proposal and the Maddi and Philips and Gutterman and Meidinger studies. Meidinger's table is representative:

Interestingly, only one question on settlement skills is noted by the ABA compilers. It is clear, too, that "courtroom management" means control of decorum and not many other skills a judge needs. Only three polls ask about efficient use of court time, a narrow test of courtroom management. Calendar management might, for example, be a better focus in many jurisdictions where the judge is in complete control of scheduling.

It is interesting to note that . polling questions parallel only somewhat the sense of judge's competency in their own eyes as reported by Ryan . et al. (1980, p. 162). Their

evaluation was that they were highest on adjudication skills and weakest in negotiation. The judges' self-rating in order of competence was:

1. Adjudication
2. Administration
3. Legal research
4. Community relations
5. Negotiation

Lawyer's polls tend to rate judges almost exclusively on adjudication and legal research. The ABA report advanced, also, a number of quantifiable objective indicators as possible aids to evaluation. There is both overlap and outright redundancy in some of these, as clearly what is intended is heuristic:

> Some of those quantifiable "indicators" at the trial court level -- each of which would have to be used with great understanding, not by rote -- might include data concerning (1) frequency of reversal on appeal; (2) number of cases handled over a period of time (3 years for example); (3) types of cases handled (nature of cases, jury trial, etc.); (4) time between submission of a case and decision; (5) number of cases settled before trial; (6) number of cases settled during trial; (7) hours of attendance at continuing education

> courses; (8) number of postponements of hearings, conference, etc.; (9) sentence data; (10) number of complaints filed with judicial disciplinary agencies; (11) frequency of complaints with pertinent rules (such as filing Findings of Facts, etc); (12) data concerning movement of the docket (as well as study of content of the daily docket); (13) disparity of sentencing as compared to other judges of the same court and in the same court system; (14) industry - amount of time devoted to judicial duties and in furtherance of the administration of justice; (15) character - honesty, integrity and maturity; (16) participation and/or invitations to participate at law schools, legal seminars, etc.
>
> Some of those quantifiable indicators for appellate courts might include data concerning (1) conciseness of written opinions; (2) length of time from hearing and/or assessment of case to circulation of written opinion; (3) frequency of dissenting opinions; (4) number of cases handled over a period of time; (5) types of cases handled; (6) time between submission of a case and a decision; (7) hours of attendance at continuing education courses; (8) number of complaints filed with the judicial disciplinary agencies; (9) frequency of compliance with pertinent rules...

...

Some subjective "criteria" might include (1) legal ability and knowledge; (2) diligence and industry; (3) interpersonal traits; (4) judicial temperament and integrity; (5) conduct outside the courtroom. It was suggested that the above 5 criteria would be most helpful in judicial "self-improvement." Other subjective criteria, useful for "external" purposes, might include (1) comprehension of the applicable law in a given case; (2) willingness to consider novel theories and ability to understand such ideas; (3) consideration of briefs and arguments in an area of law which may be previously undecided or unfamiliar to the judge; (4) attitudes toward counsel and litigants; (5) industry; (6) judicial temperament - patience, courtesy, sense of humor, courage and dignity; (7) appearance of fairness and impartiality; (8) in fact, fairness and impartiality. (ABA, 1980, p. 13).

The National Center for State Courts proposal concentrates on the need to develop and validate scales designed to measure the type of subjective evaluators described above. As such, it has serious potential to supplement the ABA effort.

Another approach to measurement of judicial quality can be found in the report developed by Dr. Ariel Sharon of the Office of Personnel Management in an attempt to justify the

selection process for federal Administrative Law Judges. The method employs "content validity" where one enumerates and defines the principal qualities sought and measures this against the test items to suggest logical connections. This lowest level of validation does not test whether the qualities are in fact useful or necessary (performance validity) or whether the test distinguishes good and bad performers (validity). It none the less remains a meticulous effort to define important qualities.

V. CONFRONTING THE DISCREPANCY

What is striking when one compares actual evaluation--polling and the like--with description of performance is the strong discrepancy between performance needs and what is being rated. If not merely the tip of iceberg, courtroom performance hardly exhausts the repertoire of a good judge. Even negotiation skills are a distinct, significant and separate component of on-bench functions. This does not embrace the 25-50 percent of time the judges devote to non-bench duties. One significant component omitted from bar ratings for example is administrative skill, perhaps because this is not very visible to courtroom attorneys. But even trial court management skills are not fungible with general management performance. This dimension may affect general case disposition, as well as the tone of courts. (Studies indicate the courtesy by court clerks rivals outcome of the cause in determining litigant attitudes toward courts.)

Most polls have elaborate questions on courtroom demeanor of judges. The focus on courtesy to lawyers looms very large indeed in bar studies. (Lawyers argue this is not just an emotional reaction. Observors may be influenced in judgment of the lawyer's abilities and jurors even as to the worth of the case, taking cues from the judge's demeanor). Yet negotiation skills are asked about in only a few questions, often globally. Additional questions tend to be due process ones--(does the judge avoid coercive behavior?). Judges skillful in other aspects of court life have no great complaint about the jeopardy of questions currently asked--a judge who can't handle a courtroom is unlikely to be a good judge--but are being incompletely assessed.

Part of this is a consequence of the prevailing process of evaluation, and the effects on criteria noted in the early part of this essay. Lawyers have public interests also at heart in these matters, but they still see the world through their own eyes. A "civilian perspective," to use the Cahns' famous phrase, is called for. And still other perspectives, such as that of fellow judges, would be needed for a rounded view. Obviously the more cumbersome the process the less likely it is that review will take place at all. Attorneys do have the most information and perspective and willingness to evaluate of any single group. They constitute an efficient evaluation unit, but it may be possible to improve upon their efforts without

excessive complication. The proposal of the National Center for State Courts would experiment with polling of other groups including jurors and other judges, and still others being asked to appraise specific performance. Some new dimensions could also be taped. If new questions are not asked, or differences in viewpoints are not significant, such effort might still be worthwhile if it leads to greater authority and acceptance of results. (It would, however, be unfortunate and ironic if the search for the perfect evaluation led to forestalling of it all).

The major hope of such an effort would be, however, to achieve congruence between the job and what is valued and evaluated. The trend and the need is for more managment and for formal legal treatment. The increasing ratio of lawyers to population makes this needed shifting less, rather than more, likely to be carried out by the lawyers. Talent in this direction among the judges is valued (and increasingly) by over-worked colleagues, but public respect will be helpful. Certainly it should not handicap evaluation elsewhere.

SUMMARY

Clear recognition that judging is a complex set of activities would have considerable advantages. It would lead to more candor in presentation of data on courts and judging. In the medium-run (to say nothing of the long-run) this will be a better defense against demagoguery than disingenuous statistics. Critics of the courts should examine the limits

of "on-the-bench-mania" even as they should be applauded for their efforts to make judges responsible to their office and sensitive to the social costs of their lapses into laziness or imperiousness. Judges need more feedback as to their performance than is currently available through the delphic process of reversals on appeal or isolated and insulated gossip. Such evaluations, however, should be thorough if they are to serve in a meaningful way. We (the public) must be as judicious as we are demanding them (the judges) to be.

BIBLIOGRAPHY

American Bar Association, Criminal Justice Section, Committee on Evaluation of Judicial Performance, Draft concept paper, Project on Evaluation of Judicial Performance (Mimeo), January, 1980.

The American Lawyer, Federal Judges Survey: The Best and the Worst, July, 1980.

Carlson, Kenneth, et al, Citizen Court Watching: The Consumers Perspective, National Institute of Criminal Justice and Law Enforcement, 1977.

California, Administrative Office of the Courts (Mimeo), Weighted Case Load Study, California Superior Courts, February 1, 1977.

_____, California Municipal Courts, February 1, 1979.

_____, The Development and Use of a Weighted Case Load System in California Courts as a Means to Determine Judgeship Needs, n.d.

Chicago Council of Lawyers, Federal Judicial Survey, 1976.

Colorado, Interim Report of the Committee on Judicial Performance, January, 1980, (Mimeo). Final Report (Mimeo). (1980)

District of Columbia, Commission on Judicial Disabilities and Tenure, Annual Reports, 1975-76, 76-77, 78, 79.

Doyle, Joe, "Comparing Court Productivity," 61 Judicature, 416 (April, 1978).

Economic Development Council of New York, Supreme Court Task Force Court Utilization Study, 1976.

Florida Judicial System Weighted Case Load Study, 1974-1977, Office of the State Courts Administration, Tallahassee, Florida, Mimeo, n.d. (1978?).

Federal Judicial Center, The 1969-1970 Federal District Court Time Study, 1971.

_____, The 1979 Federal District Court Time Study (Steven Flanders, author), Draft, May 2, 1980.

Flanders, Steven, "Evaluating Judges: How Should The Bar Do It?", 61 Judicature, 304 (1978).

Friedman, Lawrence and Percival, Robert, "A Tale of Two Courts," 10 Law and Society Review, 267 (1976).

Fund for Modern Courts, Court Monitoring Project, 1976.

Gutterman, J.H., and Meidinger, E.E., In the Opinion of the Bar, American Judicature Society, 1977.

Goldstein, Joel H., "Bar Poll Ratings as the Leading Influence on a Non-Partisan Judicial Election," 63 Judicature, 377 (March, 1980).

Handler, Alan B., "A New Approach to Judicial Evaluation." 3 State Court Journal 3-10 (Summer, 1977).

Johnson, Earl Jr., "Thinking about Access" in Mauro Cappelletti and Bryant Garth, Access to Justice, Sitthoff and Noordhoff, 1979, Vol. 3. pp. 5-69.

_____, Statistical Dimensions of Justice System, (Forthcoming, Oceana).

Kentucky Weighted Case Load Study, Prepared for the Administrative Office of the Courts by Arthur Young and Co., October, 1976.

Lubbers, Jeffrey, "The Administrative Law Judge," Report, revised April, 1980.

Legal Services Corporation, Delivery Systems Study Report, Draft, June 2, 1980.

Ma di, Dor Linder, Judicial Performance Polls, American Bar Foundation, Chicago, 1977.

Meidinger, E.E., "Bar Polls: What They Measure, What They Miss," 60 Judicature, 468 (May, 1977).

New Jersey, Statistical Supplement to the Annual Report of the Administrative Director of the Courts, 1978-1979, Annual Report of the Director, 1978-79.

New York Trial Judge's Association Press Release, June 3, 1980.

National Center for State Courts, State Court Case Loads: The State of the Art, Washington, GPO, 1978.

Philip, Cynthia Owen, How Bar Associations Evaluate Sitting Judges, Institute of Judicial Administration, 1976.

Rosenblum, Victor G., "Evaluation of Administrative Law Judges: Aspects of Purpose, Policy and Feasibility," Draft, Mimeo, Administrative Conference, May, 1980.

Rubenstein, Michael, "Alaska's Judicial Evaluation Program," 60 Judicature, pp. 478-485 (May, 1977).

Ryan, J.P. and others, American Trial Judges, Free Press, 1980.

Schwarzer, William W., "Managing Civil Litigation: The Trial Judge's Role," Judicature 400 (April, 1978).

Sarat, Austin, "Understanding Trial Courts: A Critique of Social Science Approaches," 61 Judicature 318 (February, 1978).

Volcansek-Clark, Mary, "Why Lawyers Become Judges," 62 Judicature 166 (1978).

Vogt, Leona, et al., Field Test Results of Peer Review Quality Assessment of Legal Services, Urban Institute, February, 1976.

Wheeler, Russell, and Leo Levin, Judicial Discipline and Removal In the United States, Federal Judicial Center, 1979.

Washington, Weighted Case Load Project, District Courts, 1977.

_____, Superior Courts, 1977.

Table 1

Evolution of California Judge Year Values

1961	(Recommended by Legislative Analyst Office)	1200 trial hours
1966	Approved by Judicial Council (for Superior Court)	750 hours trial time
1968	Approved by Judicial Council for Municipal Court	60,000 minutes (1000 hours) of dispositional* time
1971	Approved by Judicial Council	Case-related time (See Table 2)
1977	Approved by Judicial Council	(See Table 2)

*dispositional time included bench work other than trial time

Source: Derived from The Development and Use of the Weighted Caseload System in California Courts as a Means to Determine Judgeship Needs (Unpublished mimeo, not dated).

Table 2

California Judge Year Values for Different Sized Courts

Court Size by Authorized Positions	Value in Minutes	
	1977 (proposed)	Prior
1-2 (7)	62,100	55,800
3-10 (11)	71,400	61,100
11-20 (14)	74,000	66,300
21 and over	74,000	71,600

Source: California Judicial Council Proposed Weights Memo, 1977.

Table 3

Range of Florida Circuit Courts as to Weights for Different Case Categories (in Real Minutes)

Case Category	High	Low
Probate	45	25
Delinquency	45	25
Dependency	40	25
Homicide	470	130
Rape	70	35
Robbery	270	55
Contracts	115	60
Dissolution	55	35
Auto	240	75
All other civil	210	75

Source: Derived from Florida Judicial system; weighted caseload study, 1977, table 1.

Table 4

Washington Superior Court Load

	No. of Cases	% of all Cases	Case Weight	Total Minutes	% of Judge time
CIVIL					
Tort	1003	7.8	156.1	156,517	16.5
Commercial	2048	16.0	16.8	126,617	13.4
Prop. Rights	395	3.1	166.6	65,819	7.1
Dom. Rel.	4214	32.9	33.3	140,124	14.8
Appeals	50	0.4	237.3	11,866	1.3
Writs & other	1357	10.6	74.8	101.497	10.7
Probate	1289	10.0	19.9	25,665	2.8
Mental illness	112	.9	45.8	5,130	.6
(Juvenile)	694	5.4	52.4	36,365	4.0
CRIMINAL					
Felonies	1300	10.1	193.2	25,141	26.5
Criminal Appeals	364	2.8	45.6	16,611	1.8
		100.0		944.344	100.0

Source: Derived from National Center For State Courts, Washington Case Load Study, Superior Courts, 1977, Table 7 and page 27.

Table 5

Judge's Subject Matter Specialization, National Survey (in %)

Purely criminal	12.2
Juvenile (mostly criminal)	1.5
General (civil and criminal)	59.1
Civil Exclusively	26.8
Misc., Administrative and motions	.3
Appellate	.1
	100.0%

Source: Derived from Ryan et al. Table 2.2.

Table 6

Criteria that Bar Associations Use in a Multiple-Attribute Poll

Criteria	% of polls that ask this question	N
Technical qualifications		
Legal ability	72%	18
Legal knowledge	28	7
Legal experience	20	5
Quality of opinions	20	5
Procedural correctness	16	4
Substantive correctness	8	2
Evidentiary correctness	4	1
Intellect	4	1
Work capacity		
Diligence/industry	68%	17
Punctuality/promptness	52	13
Trial management	24	6
Studiousness	24	6
Settlement skills	16	4
Age	16	4
Administrative skill	8	2
Efficiency	4	1
Physical/mental fitness	4	1

Table 6 Continued

Criteria	% of polls that ask this question	N
Interactive results		
Courtesy	64%	16
Attentiveness	28	7
Proper demeanor	16	4
Lacking controversial conduct	16	4
Patience	16	4
Considerateness	12	3
Respect for lawyers	8	2
Sense of humor	8	2
Character traits		
Judicial temperament	72%	18
Integrity	52	13
Impartiality	48	12
Lack of bias/prejudice	40	10
Political/economic independence	20	5
Decisiveness/firmness	16	4
Courage	12	3
Intellectual honesty	12	3
General character fitness	8	2
Judgment/perspective	8	2
Neutrality	8	2
Willingness to learn	8	2
General qualification for office	68%	17

Source: Meidinger, 1977, p. 473.

END