

# **Assessments, Monitoring, Evaluation, and Research: Improving the Knowledge Base for Judicial Reform Programs**

**Linn Hammergren**

## **Introduction**

Judicial reform is a new discipline, but not so new as to escape questions about its intellectual and practical value. In the past few years, external observers and long-term participants have begun a serious reexamination of its accomplishments on both dimensions. This process is not likely to end soon, but an early conclusion is that poor knowledge management constitutes a serious impediment both to disciplinary development and to its application in the resolution of real problems. Here critics have pointed to a series of issues, ranging from a tendency to design, and execute projects in a near informational vacuum, through an absence of systematic efforts to build and test theories and strategies, to a simple failure to draw on past experience or to consolidate and disseminate it adequately. The problems may be inevitable given the field's dominance by practitioners and advocates more inclined to action than to contemplation. However, they are sufficiently serious to threaten its immediate utility and longer-term survival. For those convinced that there is something worth saving, improved knowledge management has thus become a priority.

The current study looks at a fundamental part of this undertaking, the means by which reform participants can improve the information base for their individual activities and convert it into knowledge about the reform process. Four key elements in that effort are assessments, monitoring, evaluation, and research. While discussed separately, they are clearly interrelated, both in the case of individual projects and in the overall effort to produce a collective knowledge base for the discipline as a whole. The emphasis on field interventions might seem illogical given practitioners' bias toward action as opposed to reflection. However, this is where most knowledge continues to be generated and where its inadequate utilization poses the greatest costs. In the absence of a strong academic constituency with a vested interest in the discipline's survival, the practitioners, for better or worse, still have the major responsibility.

This is in the nature of a lessons learned analysis. Many of those lessons are negative, but I don't propose to spend much time on another issue, why things are as they are. Most of those points are obvious. Those interested in pursuing them are referred to essays, by the present and other authors, who have focused on perverse incentive systems, disciplinary biases, and a host of other contributing factors. At the risk of placing excessive faith in the powers of rationality, my assumption here is that an analysis of what we know about how we generate, collect, and use information might be sufficient to encourage improvements even in the face of organizational and political incentives for doing otherwise.

This incorporates a corollary assumption – the belief that knowledge-based action produces better results than stabs in the dark or uninformed good intentions. That still begs the question of how much you should know before you act and how much time and other resources should be invested in sheer knowledge building as opposed to working actual improvements. An emphasis on immediate action is embedded in most incentive systems, frequently voiced as the argument that we have enough studies and that what we have has not be useful (or used). While believing that more resources should be devoted to our knowledge base, I also would agree that we have made poor use of what we have. More with more would be better, but for the meantime more with the existing investment is a desirable and feasible goal.

The further development of this paper is very simple, looking at how the four activities have been conducted, identifying the strengths and weaknesses of the usual approaches, and deriving some lessons as to how they could be improved. Not discussed, but of equal importance are a series of subsequent steps involving the dissemination of the improved knowledge base, and the means by which the entire reform community can be encouraged to use it. It can be hoped, of course, that the availability of an improved product might provide additional incentives, and that those involved in its creation might also promote its use. However, more direct action will also be required, and more thought directed to the form it should take.

### **Assessments**

It may be only a matter of faith, but it is generally agreed that any reform program requires a good prior understanding of the problem to be resolved. This is usually achieved through a diagnostic, assessment, or appraisal. Whatever the term used, the intended product is a presumably objective study identifying the problems of the target system, attempting to prioritize and relate them, analyzing immediate and underlying causes, and suggesting likely solutions. It is also generally agreed that the assessment should be as broadly focused and open-ended as possible even though its recommendations will likely target only a part of the system. This is a way of avoiding the risk of premature diagnosis -- the study focuses on what the authors have already decided is the problem and thus overlooks what may be more important issues. Diagnosis is an iterative undertaking and is unlikely to occur absent some preexisting notion of what is wrong. However, a good diagnostic takes this notion as a hypothesis and is prepared to test and replace it as necessary. Finally, an assessment often requires original fieldwork and data gathering, but it can also be done as a desk study, especially in systems where considerable information and research is already available.

Despite the simplicity of these basic principles, a majority of reform programs do not follow them. Among the situations found are the following:

- No diagnosis is done: in recent presentations, several World Bank officials have noted that the Bank had 18 projects under execution and 12 sector assessments completed. The situation may be improved by the inclusion of diagnostics as a first phase of a project, but it is a matter of record that the Bank and other donors do initiate reform work without a sector assessment to guide it.

- Although a diagnosis is done, it prematurely focuses on a single issue and does not place this in the context of the broader system. If the topic is judicial reform, one would at least expect a review of the institutional setting and of the way non judicial elements impinge on judicial operations. We cannot assess judicial impact on credit operations or citizen security without reviewing the other contributing factors.
- The diagnosis is more broadly focused, but still skips over important problem areas. This is most likely when assessors are too concerned with pleasing the client (whether the judiciary, a government, or a major funder) and thus choose to avoid sensitive areas (judicial corruption, incompetence) or themes which for one reason or another they believe are of lesser interest – i.e. criminal law for a funder which can't or won't work in that area.
- The diagnosis is done from one point of view. There are many stakeholders affected by judicial performance. However, assessors too frequently consult only with those most easily reached or most vocal in their criticisms. This is also a problem of self-assessments where judges or governmental actors may focus only on their view of the problem.
- The diagnosis is guided entirely by what people think about the system with little effort to test this empirically. Like many of the above problems, this often originates in the scarce time and financial resources provided to the assessors who thus are forced to rely on collecting what is already “known” both about problems and their causes. It is also encouraged by the recent emphasis on participatory assessments – participants’ viewpoints are important, but they do not necessarily constitute the truth of the matter.<sup>1</sup>
- Assessors unfamiliar with what a reform program can do may focus on insoluble problems or impossible solutions. This is a particular problem for academics and activists less familiar with or less accepting of the usual political and resource constraints.
- The assessment becomes an inventory of structural and procedural characteristics presumed to constitute problems in and of themselves, or to be causes of problems (which are otherwise not identified). This approach would only be valid if we had an empirically based model of what a well functioning system looks like. In its absence, there is little basis for holding that salary levels, the lack of a training program, or the contents of specific laws are problematic in their own right.
- An assessment, of reasonable or unreasonable quality, is done, but it is further evaluated in a vacuum – there is no yardstick for determining whether what is found represents serious problems or the universal “norm” (e.g. citizens everywhere complain about court delays, even in systems which seem to move fairly rapidly.)
- Finally, there is the frequent phenomenon of multiple assessments, of varying degrees of completeness, conducted by several different agencies, or sometimes by the same one. This problem will not be addressed further here, but it warrants mention because it wastes resources, and, absent any effort to coordinate designs or share findings, can also produce contradictions in the resulting programs.

---

<sup>1</sup> This point has been well understood by U.S. researchers for years (see Kritzer, 1983). Recent World Bank sponsored research, conducted in several Latin American countries, has demonstrated that “conventional wisdom” about court performance is also inaccurate in that region. See Hammergren et als (2001, 2002).

These observations are all taken from real cases, and for each of them examples of the counterproductive results could be provided. However, rather than wallow in the negative, we ask the reader to assume that 1) having an assessment before beginning a reform or conducting one as part of the first stage is preferable to not having one at all and 2) that the positive principles to be followed are the converse of the situations described above. The guidance thus derived is still fairly abstract and much more remains to be said as to what a good assessment should look like and how it should be conducted.

Longer and more complex is not always better. When USAID began its administration of justice programs in Central America in the early 1980s, it sponsored a series of very complete assessments, which took up to a year to conduct, and ran to hundreds of pages each. This is a luxury that may no longer be affordable, in terms of time or money. Moreover, in the specific case of the USAID assessments, it has been frequently remarked that most of the material collected was never used. Proportionality is thus another key concept. Many activities that cannot be counted as reforms, but just as very targeted efforts to address an obvious problem, may require relatively little in the way of an assessment. Whether this is the case at hand is a judgment call. For example, the donation of a couple of computers, software and training to create a prison registry (and so keep track of prisoners' whereabouts and status) might not warrant an evaluation of the entire prison system. On the other hand, this approach could also give rise to a flurry of disconnected activities, each aimed at a discrete problem, with no consideration of their integration, prioritization, or coordination. Obviously, before this practice is too far extended, it is in everyone's interest that a general diagnostic be available. For example, given the large quantity of proposals to conduct training programs, it would be well to link these to an assessment at least of training needs, and preferably of a broader type – to help determine whether training is really where funds should be invested.

The remainder of this section focuses not on partial assessments, but rather on what else we know about the characteristics of a good initial overview. In suggesting how better assessments might be constructed, it elaborates on the four following arguments:

1) Even when the intended focus is narrow, it is preferable to begin with a broad survey. A proposal to work on court reform requires first situating the courts in a broader institutional context, and if with less detail, identifying and collecting information on other agencies and contextual details which will impinge on court operations.

2) Whatever the target institutions, the assessment should collect three kinds of information as the basis for its analysis and overall conclusions:

- Purely descriptive, qualitative and quantitative details on the country, sector, and institutions to be covered.
- A more qualitative and if possible quantitative analysis of institutional operations
- A discussion of the major problems claimed and identified as caused by current operations.

3) Although each assessment is done for and possibly by a specific country, it is becoming increasingly evident that a standardized format would be helpful. This is obvious in the case of international agencies conducting a variety of programs. It helps them to understand what they are doing. Less obviously, it is also useful for a country or country specific program. Generally speaking a problem is best identified, understood, and resolved in a comparative perspective. It is not uncommon to find reformers proposing solutions which have repeatedly failed elsewhere, describing as highly problematic a situation which is not that bad by international standards, or ignoring other conditions which by those same standards might be regarded as requiring attention.

4) Finally, and especially in the case of a program initiated and implemented by a country itself, some of the following may seem superfluous. Why would they want to pay someone to document what they already know? If the sponsors truly believe that the document will never be read, consulted, or otherwise used by those less familiar with the system, then they may downplay the general descriptions. However, even for internal purposes, some of the descriptive material serves a purpose. Population, GDP and income figures may help in making subsequent calculations; or in tracking progress; and even insiders may not have thought about some of the obvious details. Many US lawyers for example do not understand that country's administrative court system or what it implies for the burdens on the ordinary courts.<sup>2</sup> In Bank sponsored research in Mexico, the Mexican lawyers doing the field work found they required their own briefing on the specific procedures under investigation. As further discussed below, once we enter the area of judicial statistics and things like caseloads and time to disposition, it is now widely acknowledged that the estimates of insiders are frequently way off the mark.

#### ***Elements of an Overview Assessment:***

**1) Descriptive section:** Description is often seen as a mere prelude to the real assessment, and consequently either given short shrift or done with no particularly sense of what is relevant and what is not. It is not uncommon to find entire assessments composed of little more than a repetition of legal facts, a restatement of the laws with little attention to how they really operate. Thus, to facilitate cross-country comparison and to ensure the information gathered will be the most useful for country planning, the following guidelines are recommended:

***General Discussion of national context:*** There are some basic elements of the national setting which are critical to evaluating judicial performance. These include the population size, distribution (rural and urban, as well as regional dispersion), per capita income and income distribution, economic base, literacy rates, ethnic and other divisions, and so on. Certain historical information will also be needed – colonial tradition, form of government and any recent changes (e.g. redemocratization), recent ethnic conflicts and civil wars, and salient political issues. The treatment should be brief, and is largely for outsiders, but it may even be helpful to local planners in interpreting what follows.

---

<sup>2</sup> A recent international lecture by a former Chief Judge of the Federal Court of Claims found US lawyers in the audience asking each other what the court did.

***General Description of Legal Tradition and Basic Organization:*** While this general description is primarily for “outsiders” who will either review or make inputs to the reform plan, it may be a useful reminder to insider judges and others of what the entire system comprises. Emphasis here is less on how the system operates than on establishing its basic organizational and procedural parameters. Although a few basic statistics will enter here (how many members on the Supreme Court, number of court districts, etc) most of these are reserved for the next section.

First off, what is the legal tradition (common, civil or some subcategory or other mix)? Is there an indigenous or traditional source of law and what is its legal and organizational relationship to the state system? What are the basic laws shaping the court or justice system, when were they written, and how often have they been revised? How much of the organization, powers, etc of the sector are governed by the Constitution, by statutory law, or by regulations issued by the organizations themselves? What are the major institutions in the ordinary justice system, peripheral ones (administrative and other special courts, administrative police, bar associations, independent regulatory agencies with judicial powers, etc)? How is legal representation handled? Are there traditional or communal adjudicatory bodies outside the ordinary court system? If so, how are they organized and where are they located?

As this is the most free form part of the exercise, it would be helpful to establish a general outline to be used by those without a strong independent preference for their own descriptive scheme. This would make the results more useful for comparative purposes and also avoid the omission of important categories. Even experienced reviewers frequently overlook administrative or parallel court systems, or ignore things like the presence, organization and compulsory nature of bar associations and memberships. The section analyzing operational workings may remind them of their omissions, but a basic checklist of what should be included in the descriptive system would also be helpful. It might be a joint product of agencies doing many of these assessments. It then could be made available to all interested in undertaking one.

***Basic Statistics (so far as available) on Sector operations:*** It is extremely helpful, although often very difficult, to include statistics ranging from the number of judges, court staff, prosecutors, private lawyers, etc to the number of cases filed (by category, district, individual court), the caseload per judge, annual filings and dispositions, backlog, and if possible disposition rates and times. Figures on budgets, the funding sources (how much from the national budget, how much from judicial fees or taxes?), salaries, and expenditure categories are also useful as are crime rates (at the various points of reporting, including, if available, victim surveys). The reason for collecting these statistics is to flesh out the narrative descriptions of judicial and sector operations and potential problems, and to allow comparison with figures drawn from data bases on other systems.

As further elaborated below, single statistics tell us little, and even a complete statistical profile of a country must be interpreted with care. However, for a reader with some knowledge of worldwide trends, a few key numbers (judges and cases per 100,000 population, ratio of judges to court staff, prosecutors, defenders and private lawyers, budgets and salaries) may provide a much better handle on performance and performance

problems than the longest narrative description. This also suggests the utility of a standardized format for data capture and presentation. Experience suggests that some statistics are more meaningful than others, and that the manner of their presentation also facilitates their use. The lump sum spent on salaries is far less useful than the salary range for different kinds of employees (and this in turn is meaningful only judged against per capita income and salary scales for similar public and private sector employees). An understanding of the real workload usually requires several statistics: cases pending, filed and disposed on an annual basis, broken down by level of court, and if possible by major types of proceedings. A list is provided in the annex outlining the usual categories and conventional presentations of data.

Obviously, the availability and quality of statistics are highly variable. In some countries, it may be impossible to get a good count even of the judges and lawyers. Individual court districts may have no idea of the size of their caseload or what is being resolved and at what speed. Furthermore, available statistics are often of dubious quality. Used largely for end of year speeches by agency heads, they may never be checked and as a result supplied quite casually. Finally, whatever their accuracy they may not be provided in very useful forms. Socio-economic information on court users is frequently unavailable even in the most sophisticated systems, as are useful breakdowns by type of case, or comparable categories across jurisdictions. Despite long years of statistical studies of judicial operations in the United States, researchers still find it difficult to make comparisons across state systems. A recent effort<sup>3</sup> to accumulate comparable statistical data bases on a number of European courts has footnotes as long as the statistical tables – allowing the authors to stress the difficulty if not impossibility of making precise comparisons. In such more developed systems, even something so simple as an appeals rate is often impossible to calculate.

Nonetheless, numbers are important, and efforts should be made to collect them. They are helpful in determining the relative importance of phenomena identified more impressionistically and may also help put a different slant on problems reported by local observers. It is rare to find a judiciary that is not described as swamped with work, but figures for annual filings or dispositions often present another picture. In the worst of cases, assessors may have to supply their own statistics, conducting inventories or studies based on smaller samples. Most assessments will not have funding to do this, and thus the absence of a good, reliable statistical system may be among their most important findings. In that situation, they will just have to collect what they can, note what is absent, and interpret their findings with extreme caution.

It is also important to recognize that the numbers don't speak for themselves, and where they do speak it is usually cumulatively, contextually, and comparatively. There has been an on-going effort in recent years to collect and use court statistics as basic indicators of judicial performance. This, it should be stressed, is not why their collection is recommended and brings its own problems. Efforts to elevate certain statistics (average caseload, number of judges per 100,000 population, ratio of judges to court staff or to private lawyers, conviction rates, etc) to the level of indicators (representative of some aspect of the quality of judicial performance) or even worse, to develop one or ten

---

<sup>3</sup> See Blankenburg

basic indicators of the same have yet to be abandoned, although most who have attempted this have soon discover the practical and logical shortcomings.<sup>4</sup> Are 22 judges per 100,000 inhabitants better or worse than 8 or 30? What is a reasonable caseload for individual judges? A good disposition rate? A good appeals rate? To begin to develop answers to these questions we need far more information on the range of variations (one additional reason for urging the collection of statistics, and in some fairly standard categories), and even then, the answer is far more likely to be a range rather than a single figure, and still subject to contextual constraints.

In short, statistics, when available should be considered primarily as ways of enriching the description of the target system, not as a direct means of evaluating its performance. That there is, and probably never will be one statistic that can be used to measure the quality of justice, is not only a consequence of the various values pursued, but also of the different contextual situations of specific systems. These descriptive statistics may, once a problem has been identified, be a means, collectively and comparatively, of deriving some hypothetical explanations. The number of judges absolutely or per 100,000 inhabitants means nothing on its own. Combined with litigation rates, number of lawyers, and some understanding of the procedural rules and requirements for representation, it may begin to give us a picture of whether a complaint about delay or lack of access is valid or not.

The same need for caution should be applied to more complex statistics and statistical trends. Although they look more like good indicators, they are usually subject to a variety of interpretations and explanations. Ironically, the same statistical profile might characterize a very good and a very bad system, and as the situation develops over time, rates and trends may take unexpected directions. More or less is not always better. A high conviction rate on criminal cases brought to trial might mean a well trained prosecution with a good nose for what should go to court, or just an easily pressured judiciary. Where the latter is the case, an effort to impose due process rules and curb other abuses may well lower the conviction rate, at least temporarily. A drop in the percentage of untried prisoners (a usual goal for human rights groups) might mean a very selective use of pretrial detention, or that other, less desirable means have been found for reducing the prison population. As has been frequently observed, crime rates depend on many things besides an effective criminal justice system. A system may be so ineffective that citizens simply stop reporting crimes. As improvements are made, and citizens begin to have more faith in it, reported crime rates are likely to rise. The highly desirable 50-50 rate for overturns on appeals (indicating that only very difficult cases go to appeal) could also be produced by a system that operates like a lottery, and even worse a corrupt one.

For those used to dealing with general purpose indicators like infant mortality rates or GDP, the ambiguous significance of candidate indicators for judicial performance will be perplexing. It is a logical consequence of the judicial role, the variations in how it is defined and enacted in different countries, and the complex nature of its output. The judiciary is a reactive rather than proactive institution, and the raw material with which it

---

<sup>4</sup> USAID developed such a list of 75 indicators in the mid 1990s. The list has since been relegated to the status of “suggested” indicators when field testing revealed that most were relevant only for a few systems, and that even the desirable direction of change was contextually determined. See USAID ...

works (conflicts and rule violation) is shaped by a variety of other forces. Its own statistical systems track the tip of the iceberg (the cases getting to the courts), not the broader social phenomena. Limiting performance measures to how well it deals with the business it receives would be tantamount to judging a public health system only by the number of sick people treated. Both measures are important, but both, if used exclusively, would also create perverse incentives. Courts like public health systems have a preventive as well as a curative role, and if performance evaluation hinges only on the latter, they may be encouraged to waste their resources by attracting more business than is necessary and focusing on processing the easy, not the important cases.<sup>5</sup>

**2) Analysis of Internal Operations:** Because this is the area where most reforms will be directed, it is an especially critical part of the assessment. The objective here is not to focus on output problems (e.g. delays, bias, barriers to access), but rather to understand how the organization or organizations function. The immediate challenge is to select those characteristics which are most influential in shaping sector output and defining its quality. Early assessments often did this by measuring organizational and operational details against an implicit model – usually drawn from the assessors’ own country. The recognition that function can and should be separated from structure and that there are consequently various paths (or path dependencies) to improving performance has made this less acceptable. However, it has also removed the implicit yardstick. Fortunately, there are any number of suggested substitutes. In fact, over the past two decades, those called upon to do assessments have often seen part of their task as leaving a format that might be used by others.<sup>6</sup> There is still no agreement on which is best, and a decided tendency to reinvent rather than adopt some existing proposal.

Presumably this sort of checklist should be short, and its elements should have a direct relationship to predicting output. As many normal inclusions are already covered in the above sections (statistics, general description) most of the suggested templates could be substantially condensed. The suggestion here is to treat the judiciary (and other sector institutions covered) as an organization with all the typical requirements for effective functioning. This means a way of selecting appropriate staff, provisions for supervising and directing their performance (without, in the case of judges, interfering with the necessary degree of independent decision making), adequate resources and a system for administering them, and procedural rules congruent with quantitative and qualitative standards for output. The attached table is one suggestion of how this might be covered. It is divided into three dimensions, one covering simple organizational functionality (here called institutional governance) and the others relating to independence and accountability. It is particularly tailored for the judiciary, and would have to be slightly modified for other sector institutions, especially as regards the independence and accountability dimensions.

---

<sup>5</sup> In a study of the Mexican federal courts, two researchers found that an evaluation system based on cases resolved tended to encourage the admission of cases which would ultimately be dismissed for lack of merit. Since they counted as dispositions, judges had little reason to reject them out of hand. (Magaloni, and Negrete, n.d.)

<sup>6</sup> A sample of the efforts is found in Hammergren (1998).

Aside from a reluctance to adopt existing schemes, the biggest obstacle to constructing this sort of assessment guide has been the tendency for legal professionals to define the problem in terms of the adequacy of the legal framework. Given the often imperfect coincidence between legal requirements and what really transpires, this has not been a very productive approach. It has not been improved by an accompanying faith in certain doctrinal principles believed intrinsic to better performance.<sup>7</sup> At least in explaining internal operations, the principles of neo- and classical organizational (or institutional) analysis are arguably better guides. Judiciaries may have special characteristics, but they still have to provide adequate incentives to their professional and administrative staff, monitor performance, and administer their resources effectively. Their failure to do so appears to account for many common problems and thus a good part of reform programming will inevitably focus on these functions.

**3) Discussion of the Major Problems Attributed to Sector Performance:** While it may be a hard message for their strongest proponents to swallow, judicial reforms are usually not supported for their own sake but as instrumental means of achieving some larger goal. At the very least such goals involve improving the provision of services already delivered (greater efficiency and efficacy in deciding conflicts and otherwise dealing with existing demand). Often they refer to downstream events, the judiciary's impact on such things as crime, economic investment and growth, incorporation of marginalized groups, citizen security and so on. The links between the two levels of goals are not very well defined. Efforts to articulate them in any given country often rest on scant, and frequently inaccurate understandings of what courts actually do. The ultimate solution, a good, empirically based theory of judicial roles and impacts, is a long term project, however. For specific assessments, the immediate task is to inventory and test current complaints, and to identify other problems which may be less obvious to those standing too close to the system.

Any reform project must logically start with a problem or with a desired change in the status quo ante. Hence, early on and throughout the assessment, information will have to be gathered on what people believe to be wrong, as this will be the motivating force for the reform. Initial responses are likely to be vague and not very specific – the judiciary is not modern, the laws are out of date, there is too much delay, judges are overworked, biased, corrupt, or incompetent. Assessors will have to attempt to elicit more precision or at the very least some concrete examples of the alleged ills. Ideally, this will provide them with a list of more targeted areas for improvement with easily identified causes (measurable delays or paralysis of cases caused by procedural bottlenecks, inefficient administrative practices; or too few judges or support staff; limited access caused by the physical distribution of courts, high court fees, or requirements for representation; widespread use of speed money or petty corruption caused by low salaries, insufficient supervision, or court practices which facilitate bribes). However, they are just as likely to find that problems are exaggerated, misdiagnosed, or entirely ignored, or that widely touted remedies are unlikely to produce

---

<sup>7</sup> Many of these are now entrenched in international conventions which often seem to incorporate arbitrary, and occasionally ethnocentric assumptions about best practices. While much of this is so vague as to constitute only a nod toward good intentions, efforts like those in Latin America to stipulate a guaranteed percentage of the national budget for the judiciary are potentially counterproductive.

much of any effect. Corruption or delay may be less common, or more localized than usually believed. What looks like an excessive workload may be vastly inflated by a large number of inactive cases remaining in the courts. Judges may make good decisions, but execution may be avoided. Or reformers may believe that all it will take is an up-to-date law to increase the efficacy of the criminal justice system or bring foreign investment flooding into the country.

To continue the medical analogy, assessors are very much like a doctor confronted with a patient with a variety of complaints and symptoms who has already decided what is wrong and what needs to be done. The patient's complaints and symptoms cannot be ignored, but the self-diagnosis is frequently completely in error. What the patient would like to see fixed with a pill, a shot, or perhaps an operation may require a much more complex kind of treatment, including lifestyle changes and other less pleasant kinds of therapy. However, because judicial health is a much more subjective state, reported ills and even suggested remedies will necessarily be given more weight. On the basis of international experience, advisors should note the costs and other difficulties associated with making certain changes, or the likelihood that proposed solutions will not bring the desired results, but in the end what constitute a better system is in the eye of its users, not in any sort of universal standards.

**4) Analytic Summary:** Taken as a whole, the three parts of the assessment provide an overview of the situation of the sector or organizations reviewed, list and prioritize the principal problems attributed by the local stakeholders or found by the assessors, and identify traits and practices accounting for current patterns of operations and output. The analysts' further task is to use this information to derive an overall diagnostic of sector or organizational performance and suggest how it might be improved. Their reworking of the basic information will also separate spurious or second order problems from more basic ones, and note where solutions might be more practically pursued outside the sector or organization.<sup>8</sup> Areas where further investigation is required should also be noted; in this sense, the initial assessment should be regarded as a first cut at diagnosis. While the quality of the analysis hinges on the skills and knowledge of the assessment team, the emphasis on more standardized formats is intended to assist their efforts. It should help call their attention to details they might otherwise overlook, simplify the collection of information, and provide a basis for evaluating the significance of the immediate findings.

Because assessments (and evaluations) collectively constitute our best source of information on sector operations and reform programs, it is particularly important that they be made available to the entire reform community. There are evident obstacles to adopting this in practice, but unless they can be overcome, the discipline as a whole as well as individual reform efforts will find their progress limited.

## Monitoring

---

<sup>8</sup> For example, research in Mexico (Hammergren et al., 2001) on debt collection proceedings found many problems originating outside the courts – for example, poor practices in issuing credit, absence of credit bureaus, corrupt property registries, lack of debtor education.

Until the recent emphasis on results management, monitoring of judicial reforms was conspicuous by its near absence. When it was done, the emphasis was on tracking inputs or outputs delivered, not the impact on overall goals and objectives. While the shift to tracking results is important, it poses problems for this kind of project because of how institutional change is accomplished. The underlying logic of any institutional change project is as follows:

- Certain external impacts are pursued by changing the ways institutions operate.
- Institutional change is affected by shifting the mix, composition, and quality of internal variables.
- Until internal behaviors have responded to these shifts, outcomes and impacts are not likely to vary, and if they do, there may be a temporary decline in quality or quantity while the new patterns are learned and perfected.

The process described thus has two main stages: first alterations in internal operations are effected, and once they are in place, outcome should improve. This suggests a sort of progress by plateaux in which the second order, outcome changes will not occur incrementally, or not begin to do so, until the first order changes are relatively complete. In reality, there is still a third stage, in which improved outcomes (court performance variables) produce changes in extrajudicial behaviors and thus in overall societal well-being. We would not expect to detect outcome changes (and the subsequent impacts on societal goals) until the internal change process is well advanced. This resulting dilemma can be phrased as two related questions:

- How do we know anything is happening in the meantime?
- How can we be sure the long-term goals will ever be achieved? (If, for example, the hypothesized links between internal change, outcomes, and impacts do not hold).

Only the first is really a monitoring problem, but its solution is closely related to the second. If one cannot directly monitor progress through incremental changes in the final objective (whether reduction of delay, improved decision making, greater trust in courts, or further downstream, higher growth rates, poverty or crime reduction), then one will have to monitor it by tracking progress in making intermediate process modifications. In essence, this means monitoring by benchmarks – the accomplishment of the various stages in the execution of a change strategy which in the end will produce fundamental alterations in how things are done and thus in the value of outcomes. For example, a delay reduction effort might start with the establishment of baseline data, proceed through an analysis of process bottlenecks, develop means for changing those not regulated by law and encouraging any needed legal regulations, introduce improved manual or automated systems, and so on. The benchmarks generally coincide with the project strategy and plan, following the chronological order they establish.

It is also possible, going to a much lower level of detail, to track the impact of these partial internal changes on lower order behaviors. For example in a project aimed at improving the coordination of police and prosecutorial operations in Panama and thus

their success in detecting, investigating and prosecuting crimes,<sup>9</sup> intermediate progress was measured by tracking changes in specific procedures – types of evidence collected, rate of consultations between police and prosecutors, or even something so simple as improvements in the format of the police charge sheet. These are extraordinarily context specific indicators, not only as regards the specific country, but also the problems identified for resolution. In Panama, it was thought important to encourage police to number the pages of the charge sheet. This thus became an indicator of progress. In the delay reduction project mentioned above, behavioral changes might include participants' adoption of the new systems, evidence that once the means to do so are installed, those responsible are keeping track of the progress of individual cases and trying to meet deadlines, that records of individual judges or courtrooms are being monitored, etc. In both examples, internal changes should be visible long before the outcome goal (more investigation completed within a shorter time, higher clearance rate of criminal cases, or reduced average time to disposition and backlog reduction for all cases) shows any signs of impact.

In the end, the significance of achieving the benchmarks or effecting the intermediate behavior changes hinges on the validity of the overall strategy. If training judges is not going to produce better decisions, than monitoring stages in implementing a training program or judges' absorption of new skills and knowledge will not have much point. This type of monitoring only makes sense in the context of a viable change strategy. The solution to this quandary is not to return to an effort to monitor outcomes and impacts. However much those responsible might want to do so, they are not going to have visible system-wide results for years. Instead, the emphasis will have to go to improving the quality of strategies and finding some way to test the hypotheses on which they rest. There is, most would agree, considerable room for improvement here. Much of it rests on the other elements of better knowledge management, including improved analyses of country specific problems (assessments), a better use of the lessons of past experience (starting with more systematic evaluations to bring them forth) and a concerted effort to explore hypothesized linkages for which empirical evidence is scant (research). This reinforces the relationship already noted among the topics explored in this article as well as the importance of increased attention to knowledge consolidation, dissemination and debate.

Although the topic of improving strategies extends beyond the themes discussed here, two further measures warrant noting: a effort on ensuring that reform programs do have explicit strategies, thus lending themselves to monitoring, ex ante quality control, and ex post evaluation, and the use of pilot programs to do a quick check on the anticipated outcomes. As regards the first measure, as critics have frequently noted, large and small programs alike are too often characterized as collections of activities with only the most tenuous links to their presumed downstream goals. They embody strategies only in the loosest meaning of the term. Forcing them to articulate the means-ends changes they presumably incorporate might be a way of better aligning outcomes with inputs (how much impact will a code or a course on ethical behavior have on curbing rampant corruption?), weeding out some of the less likely proposals (will \$19 million of

---

<sup>9</sup> This was a USAID project, undertaken in the early 1990s. I am indebted to Tim Cornish, a consultant with the project for providing documents and explanations of the benchmarking process.

new buildings and \$1 million of computer equipment really increase access to justice in a country where judges are believed to be corrupt, incompetent, vulnerable to political pressures, and guided by outdated rules and procedures?), and establishing a better base for both monitoring and evaluation. This does not eliminate space for innovation. It does mean that an innovative approach which produces no results will less likely be reattempted in the future.

Novel approaches which cannot be tested by logic or past experience can be tried out in pilot projects. Rather than build all the new, computer-equipped courtrooms, or train the entire judiciary, the proposed approach can be applied in one district to see not only whether it has an impact, but also how that impact might be enhanced. Pilots can also be used to assess the costs of successful implementation, evaluate competing approaches, and make adjustments so that replication, if desired, will be more within the means of the country. Pilots, of course are special cases, and should be evaluated with this understanding. What works for a group of judges who see themselves as, perhaps are selected for, being in the vanguard, will probably have less impact for the entire universe. Pilot projects are themselves no novelty in judicial reform programs. Unfortunately, they have too often become ends in themselves – neither created with an eye to their replicability, nor utilized as inputs to a broader knowledge base. This accounts for a growing skepticism about their value, for which their backers must take a large part of the responsibility. The other part lies in an inadequate evaluation of their results and potential impact, and thus leads to a third element of the knowledge management agenda.

## **Evaluation**

While it is commonly acknowledged that evaluation is essential to program development, this lesson has had little apparent impact on judicial reforms. For the quantity of work that has been done, evaluations are remarkably few, and all too often neither widely consulted nor even available. Everyone reads the evaluation of their own project; almost no one reads those of anyone else's work. This suggests an amazing lack of interest in acquiring information and an incentive system which allows and possibly encourages it, but it is also evident that by intent or mere oversight, evaluations are not easily accessible, even to members of agencies which conducted them. A recent suggestion that major donors share their evaluations is a good sign, but it will be hard to implement if only because they may not know where they have stored them.<sup>10</sup>

Participants in judicial reform programs seem remarkably resistant to evaluation. Courts and governments rarely evaluate their own efforts; NGOs seem to regard evaluations as an infringement of their independence, and even major donor agencies have been very lax in this area. It has been suggested that this is a consequence of a disciplinary bias. Lawyers dominate the programs, and law is not an experimental science. Lawyers are more inclined to argue than to test their positions. Moreover,

---

<sup>10</sup> A series of informal interviews with individuals charged with evaluating programs for the UNDP, IDB, USAID, and the World Bank made it clear that none of them had access to all the documentation that should have been available. I suspect, as all the work was commissioned by the respective agencies, that this reflects an information storage and retrieval problem, not a conscious effort to keep their evaluators in the dark. However, it also demonstrates an inadequate internal usage of the documents; were they being read and used, they would have been easier to locate.

because of the values involved (justice, rights, fundamental principles) and the politically charged nature of the topic, the notion of measuring or quantifying achievements is often regarded as somehow denigrating the importance of the theme. While there is a recent effort among several donor agencies to conduct evaluations of their entire programs, observers are still wondering whether the results will be released and if so, in what form. Preliminary reports from the evaluation teams suggest that their findings may not please agency leaders who have highlighted their dedication to promoting judicial reform.

Aside from evaluation-phobia another major disincentive is that evaluations do add to costs. Like assessments, they are frequently seen as absorbing funds that might be put directly into reform. And since they come at the end, when funds are most likely to be scarce, they are still more obvious targets for elimination. However, in-house policy, including the funds initially allocated for evaluation and monitoring, and decisions, like that of USAID, to make evaluation optional<sup>11</sup> indicate that more than last-minute cost-cutting is at work. Nonetheless, the increasing impression that programs are beginning to repeat old errors, that strategic programming has if anything weakened over the last years, and that the number of objective pursued has increased with no particular relationship to the content of programs, demonstrates a serious failure to learn from and build on experience. Whatever the reason for the inattention to evaluation, it may be the largest single contributor to this situation and an obvious place to work immediate changes.

A first lesson and recommendation is thus that systematic evaluations must be done, because without them there is little way of determining whether programs are producing results and how they might be improved. Results indicators, like those USAID has attempted to adopt and other agencies are also pursuing, are no substitute. They will only become one once we have a better internal understanding of how judiciaries and reforms operate. The further advantage of evaluation is that it looks not only at progress but also analyzes the path for arriving there. That second kind of knowledge is essential for anyone attempting to replicate a presumed success. It does little good to know that country X has reduced the average time to disposition of cases by 25 percent if one does not know how it did this. Evaluation can also focus on additional consequences and contributing factors. Delay reduction could be achieved by simplifying proceedings, improving in-court management, automation, discouraging filings, doubling the number of judges, or dismissing more cases for lack of merit. Each of these mechanisms would shorten times to disposition, but they have different implications for financial or other costs (e.g. reduced access). Perhaps the reduction coincided with a substantial upturn in the economy, which is usually associated with a reduced recourse to court services. This might mean that the program itself had no impact, as the reduced demand might have been sufficient to allow speedier processing of the remaining caseload.

Evaluating individual projects is important, but the value of evaluation substantially increased when evaluations are compared. Comparison makes it easier to identify the

---

<sup>11</sup> This is not only for judicial reform but for all programs and was motivated by the agency's shift to a management by results mode. Most internal and external observers regard the latter as a poor substitute, as it lends itself to manipulation (participants select results they are sure will be achieved, regardless of their wider significance) and because of the problems, discussed above, as regards selecting performance indicators.

causal relationships and the intervening variables that may affect them. Especially in an area as complex as judicial reform, where many programmatic interventions are introduced simultaneously and where exogenous variables may have a still stronger influence, it is all too easy to pick out spurious relationships. Was it the training, the higher salaries, or the internal monitoring that caused judges to speed up their handling of cases, change the patterns of their decisions, or become more resistant to bribes? Could one have achieved the same results with only one of the interventions, or achieved more by doubling one of the inputs? Or was the result a consequence of some external change with no relationship to the actual reform? These questions are not easily answered, but there is more opportunity to weed out the irrelevant or to define the conditions for relevancy when several examples are under investigation.

Thus a second lesson and recommendation is an emphasis on comparisons of evaluations or even cross-project evaluations of single activities – training programs, automation, delay reduction, or changes in basic laws. As this latter effort approaches research we can leave it for a moment, and look further at the process of comparing evaluations themselves. For this to happen the first obvious steps are that evaluations be done, that they be widely available, and that a program or incentives be introduced to encourage comparative study. A second need is for evaluations to be structured to allow comparison. This has generally not been the case. Evaluations whether done by in-house experts (either in evaluation or on judicial reform, generally not on both), contracted outsiders, or external agencies (a governmental body or an NGO), tend to be shaped by the project at hand and by the specific interests of the evaluators. Consequently, they are usually directly comparable only when the same evaluator does them, and that person's interests may still not jibe with those of the parties wishing to use the evaluation to improve future programming.

Even when general protocols are established they usually are so open ended as to allow the evaluators to pick and choose as to what they will review. The World Bank's standard evaluation format (prior situation, strategy adopted, what was done, results and lessons learned) is as good as any, but obviously leaves enormous leeway to the evaluator. More specific terms of reference tend to be determined by the project under review and thus often avoid some of the key questions (was this project worth undertaking? Was training or law revision the best way of achieving the desired ends?) as opposed to things like whether the inputs were provided as stipulated, the immediate outputs achieved, and could the quality of the individual activities have been improved. In training programs for example, attention often goes to the training methodologies, the content of individual courses, and the means for selecting students. A training expert, the usual candidate for this part of the evaluation, is hardly likely to question the value of any training at all, but rather look to how what was delivered might have been improved. Where the evaluation is entrusted to one expert, he or she may ask some of those harder questions about less favored elements, but there again, the standards remain dependent on the evaluator's inherent preferences.

What this suggests is the need for a dual focus in evaluations: one part examining the project at hand (to determine how well it has carried out the proposed actions) and the other focusing on more general questions, both the adequacy of the strategy itself and the way it addressed certain common problems. For an individual agency doing many

evaluations it should be easy to adopt this format. It may be harder to implement given the inherent limitations of evaluators themselves – their own biases and preferences, a tendency to not want to be overly critical (an evaluator who finds himself declaring even a few strategies not worth following may not be an evaluator for long), and the difficulty of finding people with both specific knowledge and the ability to take an overview. Using generic evaluation experts is usually no solution because of the specialized, substantive knowledge required. USAID's now famous overview of its rule of law programs<sup>12</sup> came under heavy criticism from project managers for just this reason. It was claimed that the authors, coming from other disciplines, never understood the purposes of the projects, confounded their objectives with cause lawyering (the use of existing systems to reach immediate benefits for the poor), and misread the multiple and interactive aims of many of the common elements.<sup>13</sup> Participants have been equally critical of macro-economic analyses of judicial performance and reform impacts and especially of their efforts to derived composite or unidimensional scores for both.<sup>14</sup> Thus, while good evaluation techniques need to be introduced, this may be more appropriate in setting the standards for the evaluation, but not in actually conducting it.

The third need and recommendation thus is thus to make evaluation a little less free form, by establishing general standards (the work of the evaluation experts) and common themes and questions (the responsibility of those designing and overseeing programs). The project-specific criteria will be set by the sponsoring agency (assuming there is one) and the project itself. This is the within-system evaluation, the one which focuses on whether the project followed its intended strategy, complied with any rules the sponsoring agency requires, and achieved its proposed results. These are important questions for evaluating any specific proposal but they are less important in terms of furthering programmatic knowledge. For the latter, the evaluation will be expected to address the adequacy of solutions in addressing certain common problems – delay reduction, combating corruption, ensuring the selection of the most capable judges, encouraging judges to make speedy, fair decisions, increasing access for marginalized groups, and so on. Once again, selecting these categories will be easier for assistance agencies doing multiple programs, but they are also the ones best situated to compare results.

## Research

Because judicial reforms have tended to become very routinized (a mixture of training, administrative restructuring, legal change, automation, ADR, and subsidized services for the poor), there is a danger that they will only lead to improving a second-

---

<sup>12</sup> Gary Hansen and Harry Blair, USAID, 1994.

<sup>13</sup> For example, training which is often used to build support for reform, collect more information on common problems, and provoke judges to take a different view of their role or develop sensitivities to the needs of a widely variety of clients, was relegated to simple capacity building.

<sup>14</sup> Many of these analyses rely on surveys or expert opinions to assess quality. A common criticism has been that those surveyed often have a unidimensional view and that in any case, even knowledgeable informants tend to misjudge real operations. See Kritzer In one recent study, lawyers from prestigious law firms were asked about the duration of debt collection cases. As several discussants noted, these lawyers probably rarely if ever did such work.

best approach. They are very much aimed at making an existing process work better, but rarely address the question of whether that process represents the best use of resources in addressing high priority needs. They also have tended to take their impact on extra-system objectives (economic growth, equity, political stability, and citizen security) on faith because of the obvious difficulties of tracking extra-system changes. Better assessments, monitoring and evaluation are unlikely to tackle these broader issues. Reform, however radical it may propose to be, like every other human activity tends to follow patterns that have already been set. Whether or not training has ever accomplished any improvement, it is far easier to mount a reform program highlighting training than one excluding it. Additionally, certain elements are added largely because they are popular –judges want new laws, equipment, training, and buildings. External assistance agencies tend to feature activities conventionally included in all their programs. And finally, there are the various preconceived notions about what is wrong with judicial systems which are again far easier to incorporate in a program than to contest. For all these reasons, knowledge building based only on actual reform programs is inherently limited. The full range of possibilities, whether of solutions or underlying causes, is unlikely to be explored because they have never become part of the standard reform repertoire.

Research is thus a vital fourth element in the knowledge management program. It can examine conventional understandings, test dominant hypotheses, and even attempt novel solutions which would not be possible in an ordinary reform. Research on judiciaries and judicial reforms is far from nonexistent, but as currently conducted suffers from some potentially serious problems:

- Research itself is not very coordinated in the sense of pursuing a set of common themes and questions, building on prior work, or confronting differing findings.
- Many researchers and most practitioners appear to lack a good overview of the relevant body of existing studies.
- Research is rarely policy relevant in the sense of exploring themes which might improve reform programming.

The explanations for the first two phenomena largely lie in the absence of a well-established academic discipline focusing on the judiciary and its reform. Researchers come from a variety of disciplines. This in itself is desirable, but some of the negative side effects are that content, methodologies, and eventual dissemination are usually driven by the parent discipline's priorities and that published articles are scattered through a vast number of journals, only a few of which focus specifically on legal and judicial reform. The fact that academic careers, access to funding, and publication depend on a disciplinary community less interested in judicial reform for its own sake creates two interlocking vicious circles which also explain a good part of the third observation. Researchers select topics, questions, and methodologies defined as important by their disciplines, and so tend to keep cross-disciplinary, thematically determined issues off the agenda, thereby guaranteeing they will continue to be seen as less significant. Practitioners' inherently lesser interest in reviewing research is thus reinforced because much research is not relevant to their activities.

The cycles tend to affect even research sponsored by assistance agencies, which often seems more influenced by academic than practitioner concerns. For example, much of the recent work sponsored by the World Bank has involved the statistical analysis of relationships between aggregate indicators of judicial and economic performance. Its apparent purpose is to demonstrate the need for judicial reform, but beyond that its impact on reform content is dubious. Even discounting questions about concept validity,<sup>15</sup> the further problem is in unpacking the indicators to understand the nature (and direction) of the relationships. Efforts to breach that gap by linking major variations in judicial structure or legal tradition to performance face similar criticisms. The construction of both the dependent and independent variables often involve some fairly arbitrary distinctions,<sup>16</sup> but even if we accept their findings, the implicit recommendations (change to a common law system, or a parliamentary as opposed to presidential one) will not be easy to enact. While the grand distinctions (civil versus common law, inquisitorial versus accusatory criminal proceedings) may matter, the emphasis on tracking their impact fails to recognize that they are the least amenable to change. It also overlooks the potentially greater variations within the grand categories (e.g. between the inquisitorial justice system as practiced in Haiti and in France). Such within-category variations are intrinsically less interesting to academic researchers because they hinge on a multitude of differences of detail, most of which will not be statistically significant. The obvious conclusion is that if practitioners are going to benefit from research, they must help define the agenda and finance those studies most likely to have an impact on their own work. Inevitably these will involve more intensive comparisons of a limited number of cases. As they will be more costly to conduct,<sup>17</sup> it is also important that the themes be selected with care.

As a start toward a new research agenda, the following general topics are suggested, ordered from least to most ambitious:

---

<sup>15</sup> These are legion. Most involve the judicial performance indicators which are usually based either on surveys or expert opinions. Even in a single country, there are questions as to what is really being measured (public perceptions, the perceptions of a small group of citizens, experiences respondents are willing to report, and so on). However, cross-national comparisons are still more problematic because of the different perceptions of what is normal. One researcher using this approach notes the perplexing finding that residents of London reported more exposure to street crime than those of Caracas. Further interviewing revealed that the threshold for reporting incidents was far lower in England than in Venezuela. (Juan Jose Toharia in a presentation at the Law and Society Association meetings, Miami, 2000.) A 2000 Gallup poll on citizen satisfaction with government handling of crime is also illustrative. The second highest ranking went to Nigeria, the fourth to the Philippines, while France, the Netherlands, Finland, and Sweden received only half the positive responses of the two former countries.

<sup>16</sup> A series of studies (la Porta et al, 1998) focusing on differences between civil and common law systems have been criticized (Pistor, 1995) for overlooking the enormous variation within and cross fertilization between the two traditions. Another Bank study (Buscaglia and Dakolias) relating delay to such traits as computerization of courts, investments in infrastructure, and proactive judges raises many questions as to how the last three variables were constructed, as well as to the posited causal links. Courts which invest in computers may simply be courts interested in reducing delay, and it may be that interest, not the computers, which explains their success in shortening time to disposition.

<sup>17</sup> Aside from the disciplinary biases of the researchers doing macro studies (most of whom are economists) another explanation for their methodological focus is that they can use data bases that already exist. This means that they can do a study covering 80 countries for less than it costs to do original research in one.

- A review (possibly as cross-project single theme evaluations) of the activities (training, automation, law reform) common to most projects to identify variations in approaches, difficulties encountered, costs, and results. This is the most basic form of knowledge generation needed, and while it might not substantially alter the cafeteria approach to programming, it might reduce redundancy and increase efficiency of actions. Despite a recent aversion to the notion of best practices,<sup>18</sup> at the level of individual activities, the concept still warrants consideration.
- A review (via cross-project single theme evaluations) of progress in advancing the common subgoals (delay reduction, increased access, anti-corruption measures) found in most programs. While more complicated than the above (because it usually includes multiple activities), the value of this approach lies in its emphasis on affecting outcomes.
- Research on the factors commonly blamed for poor judicial performance and giving rise to many standard reform policies. Among the high priority issues are salaries (for their impact on corruption, quality of judges, efficiency and efficacy), size and control of budgets (as related to quality of service), variations in appointment systems and criteria, tenure and lateral entry, variations in forms of judicial governance; and equipment and infrastructure.<sup>19</sup>
- Research on some of the overlooked candidate causes of poor performance – for example the role of the private bar and variations in its organization, power, or the sheer number of private professionals; the influence of educational institutions and patterns in legal training; the effects of judicial fees, prohibitions on pro se (self) representation, legal (il)literacy, and the types of services provided in limiting access to courts; the adequacy of and disincentives for using alternative and community dispute resolution mechanisms.
- A simple examination of how courts and other sector institutions really operate and what incentive patterns and other constraints shape their output. Most reforms are based on uncontested conventional wisdom (there are delays, large amounts are commonly at stake, large users avoid the judiciary because they don't trust it and so on). Many of these do not hold up in fact and to the extent they can be replaced with a more accurate picture of what is occurring, overall reform planning will benefit.
- Finally, because of the use of extrajudicial impacts as the principal justification for reform programs, more attention needs to go to exploring them. This is the most ambitious part of the research agenda, and the one least directly related to the immediate interests of reformers. However, it is critical to ensuring the value of their work. Here the current emphasis on aggregate, multi-country analysis can be abandoned in favor of more intensive exploration of the themes within one or several countries. The macro-studies are suggestive of where one might look; what is now required is an exploration of the specific linkages. If delay is critical, what kinds of

---

<sup>18</sup> For some reason certain donors (most notably the World Bank) have become extremely critical of the term. Taken as blueprint solutions, the criticism may be well placed. However, as more flexible rules of thumb or basic principles for conducting activities, “best practices” are demonstrably effective. For example, we know that training is best done with a needs assessment; that the focus should be not on simple knowledge transfer, but rather on improving concrete practices and behaviors; and that repetition and complementary organizational changes substantially enhance training's effect. .

<sup>19</sup> Equipment and infrastructure often constitute the major part of investments in reform programs. Ad hoc arguments and some empirical work (Buscaglia and Dakolias) suggest an impact on performance. However despite (or because) of their popularity, critics have expressed doubts which are worth pursuing.

delay and critical for whom and for what? What kinds of compensatory mechanisms are adopted and with what results? Do the answers vary across and within different types of legal systems and societies?

Donors will have to sponsor this research, but to the extent they can help build a body of relevant work they may also help legitimate it within the academic community. An obvious first priority is that they reorient their in-house research programs to coincide more closely with programming needs. A second is a larger research budget both for in-house and sponsored work. And a third is that research be reviewed and utilized. That is has not been so far may be a disguised blessing, but it is time to replace the two vicious circles with a virtuous one. If reform planners are forced to incorporate research findings in their own designs, they will pay more attention to what is proposed and what is delivered. When the results are, as is currently the case, contradictory,<sup>20</sup> someone will have to resolve the differences.

### Conclusions

There is a special irony in judicial reform's problems with knowledge management. The sector itself is, one assumes, an area which prioritizes intellectual effort and which is dedicated to processing information. Clearly there are also disciplinary and agency biases at work. Lawyers are not scientists and reformers tend to be activists, not dispassionate analysts. Multidisciplinary contributions are a plus, but the absence of an applied disciplinary base has made their input all too academic. It has also discouraged cross-disciplinary dialogue – the various disciplines are present but they don't necessarily build on each others' work. However, both reformers and academics with an interest in the theme, and not simply advancing in their respective primary homes, will need to overcome these problems if the field itself is to survive.

The topic does have more status, as an academic and applied discipline, in the industrialized nations. It would thus also be helpful if those involved in Third World reforms paid more attention to the developments in the "North."<sup>21</sup> Some clearly have, but their readings tend to be highly selective, and one suspects they are not following the emerging arguments in full. Otherwise they might put a little less faith in the remedies they adopt, many of which are now being questioned in the countries where they were first introduced.<sup>22</sup> One characteristic of a mature discipline is its development of major unresolved issues and internal conflicts. Members argue with each other because they agree on where they disagree, and it is this process of confronting contradictions that eventually leads to a better shared understanding. As practiced in the Third World, judicial reform and judicial studies still lack that trait. It is as though the participants had agreed not to enter into conflict, and thus to endorse a common dogma. There are exceptions, but they tend to be unwelcome because they are violating the fireman's syndrome by treading on the hose.

---

<sup>20</sup> For example two recent Bank publications found that salaries were 1) unimportant to and 2) highly predictive of levels of judicial corruption. The different sources of data easily explain the discrepancies but to my knowledge no one has even noticed them.

<sup>21</sup> This is the symbolic North – some very interesting work is also being done in Australia. See Australian Commission (2000).

<sup>22</sup> The benefits of ADR and pro-active court administration are two examples.

Improved knowledge management clearly threatens that cozy status quo. It requires more work, and it also endangers conventional routines. As always, the basic challenge is to get participants to focus on the long term losses as opposed to the short run gains. So long as those who control the funding continue to believe that judicial reform is important and feasible, there are immediate advantages to working in the dark. Over the short run, the funders won't notice the difference, and they might actually be disheartened by reports that things are not as easy or as certain as they seem. Still if the interest is in working real changes, as opposed to the merely cosmetic and in enhancing our ability to continue making them, our knowledge base requires continual attention. Better assessments, monitoring, evaluation, and research are a start. If we have collectively failed to ascend the learning curve, that is primarily because we have failed to construct one.

## REFERENCES

- Australian Law Reform Commission (2000), *Managing Justice: continuity and change in the federal justice system*. JS McMillan Printing Group.
- Blankenburg, Erhard (2000), "Indicators of Growth of the Systems of Justice in Europe of the 1990s: The Legal Profession, Court, Litigation, and Budgets," unpublished paper prepared for Stanford Law School.
- Buscaglia, Edgardo and Maria Dakolias (1995), *Judicial Reform in Latin America: A Framework for National Development*. Stanford University, Essays in Public Policy.
- Hammergren, 1998 "Diagnosing Judicial Performance: Toward a Tool to Help Guide Judicial Reform Programs," paper prepared for Ninth International Anti-Corruption Conference, Durban, South Africa.
- Hammergren, Linn, Ana Laura Magaloni, Layda Negrete, Alfredo Ramirez, and Rosario Trellez (2001), "The Juicio Ejecutivo Mercantil in the Federal District Courts of Mexico: A Study of the Uses and Users of Justice and their Implications for Judicial Reform," World Bank, Poverty Reduction and Economic Management Unit, Latin American and the Caribbean, Report No. 22635, June 22.
- Hammergren, Linn, Germán Garavano, Héctor Chayer, Milena Ricci, Carlos Alendro Cambellotti, Ana Laura Magaloni, Layda Negrete, Alfredo Ramírez, Rosario Tellez (2002), "An Analysis of Court Users and Uses in Two Latin American Countries," unpublished draft prepared for the World Bank, Latin America and Caribbean Region.
- Hansen, Gary and Harry Blair (1994), *Weighing in on the Scales of Justice*. USAID/CDIE.
- Kritzer, Herbert (1983), "The Civil Litigation Research Project: Lessons for Studying the Civil Justice System," *Proceedings of the Second Workshop on Law and Justice Statistics*, U.S. Department of Justice, Bureau of Justice Statistics, pp. 30-6.
- Magaloni, Ana Laura and Layda Negrete (n.d.), "El Poder Judicial Federal y su política de decidir sin resolver," unpublished draft, Mexico, Centro de Investigación y Docencia Económicas
- USAID, Center for Democracy and Governance (1998), *Handbook of Democracy and Governance Program Indicators*. Washington, August.
- Zuckerman, Adrian (1999), ed., *Civil Justice in Crisis: Comparative Perspectives of Civil Procedure*. Oxford University Press.

## **ANNEX I: LIST OF BASIC STATISTICS FO AN ASSESSMENT**

[Under preparation]

## ANNEX II: A PROPOSED INSTRUMENT FOR ASSESSING JUDICIAL OPERATIONS

*(The list includes three dimensions and presumes that an assessment will focus on all of them; a reform program which focuses on only one or two, without taking the others into consideration, may well create new impediments to satisfactory performance.)*

|                            | Selection of Judges   | Management of the Judicial "Career" <sup>i</sup>  | Internal Administration   | Resources   | Judicial Processes   | Legal Profession  |
|----------------------------|---|---|---|---|--|---|
| <b>INTERNAL GOVERNANCE</b> | <ul style="list-style-type: none"> <li>• Judges are selected through pre-established criteria based on job-relevant<sup>ii</sup> merit. Candidates are selected from those recommended by an independent professional body.</li> <li>• Selection process and criteria are periodically reviewed to ensure they are obtaining the best candidates.</li> <li>• In addition to screening for substantive knowledge and skills, criteria also include background checks and other means of identifying (and eliminating) candidates with questionable personal and professional histories.</li> </ul> | <ul style="list-style-type: none"> <li>• Training in compliance with Code of conduct and process for monitoring compliance, disciplining violators, and appealing disciplinary decisions.</li> </ul> <p>Code of conduct includes conflict of interest provisions requiring recusal.</p> <p>Code requires public disclosure of assets.</p> <ul style="list-style-type: none"> <li>• Standards for performance (number of cases decided, average time limits, reversals on appeal, service to users, etc.) exist and are monitored to help judges improve their work and where relevant, to affect decisions on tenure, promotions, transfers, and discipline.</li> </ul> | <ul style="list-style-type: none"> <li>• Administrative processes (at the systemic and courtroom level) follow set rules and procedures.</li> <li>• Budgets, procurement and management of resources are planned, monitored, and audited.</li> <li>• There is a management information system (manual or automated) to facilitate planning and budgetary oversight.</li> <li>• Administrative staff are chosen, promoted and retained through objective, merit-based procedures.</li> </ul> | <ul style="list-style-type: none"> <li>• Changes in the overall judicial budget are commensurate with the growth of the national budget and also reflect increases (or decreases) in demands for judicial services.</li> <li>• Staffing, equipment, and offices provided to judges and administrators are adequate to allow performance of their duties.</li> <li>• Staffing, equipment and offices provided to judges and administrators are no worse (no better?) than that for the rest of the public sector.</li> </ul> | <ul style="list-style-type: none"> <li>• Procedures for handling cases are standardized and mechanisms exist for ensuring they are followed.</li> <li>• Rules of evidence and standards for evaluating arguments exist and are applied in a predictable fashion.</li> <li>• Assignment of cases follows standardized procedures and results in a reasonably equitable distribution of work.</li> <li>• Procedures are reasonably efficient and designed and reformulated in the interests of eliminating unnecessary steps and bottlenecks.</li> </ul> | <ul style="list-style-type: none"> <li>• There is a transparent process for entrance into the profession, based on educational background and other relevant criteria.</li> </ul> <p>Law school curriculum includes professional ethics.</p> <p>Pre-appointment training and continuing education.</p> <ul style="list-style-type: none"> <li>• Professional codes of ethics governing the profession are widely known and enforced.</li> <li>• Denial of entry or disbarment is subject to objective, published rules, and is carried out by an</li> </ul> |

|  | Selection of Judges | Management of the Judicial "Career" <sup>1</sup>   | Internal Administration   | Resources   | Judicial Processes  | Legal Profession  |
|--|---------------------|--|---|---|---|---|
|  |                     | <ul style="list-style-type: none"> <li>• Promotions, transfers, dismissals, and/or renewal of appointments are based on preestablished, publicized criteria.</li> <li>• There is a transparent appeals process to an independent judicial body for judges in the case of denial of promotion, transfer, or renewal.</li> <li>• Training programs are available and participation is encouraged and facilitated; some sort of entry level training or orientation is compulsory.</li> </ul> | <ul style="list-style-type: none"> <li>• Administrative staff have rules of conduct, performance standards and their own career and disciplinary systems.</li> <li>• Adequate training is provided for administrative staff.</li> </ul> | <ul style="list-style-type: none"> <li>• Internal resource distribution is based on need and workload.</li> </ul> | <ul style="list-style-type: none"> <li>• Judges have the power to move cases along and to punish or deny efforts to create excessive delays.</li> <li>• Where judicial decisions are not complied with, courts have additional means to enforce them.</li> <li>• Judicial decisions are reversed only through a regularized appeals process.</li> <li>• The pre-trial settlement of disputes is encouraged but not forced.</li> <li>• There exist duly recognized alternative dispute resolution mechanisms, both court annexed and free standing, which provide a viable alternative to judicial processes.</li> </ul> | <p>independent professional body.</p> <p>Judges can be prosecuted for non-judicial misconduct.</p> <ul style="list-style-type: none"> <li>• Where there is a shortage of qualified professionals, there is a provision for lay representation or performance of some legal duties, but these individuals are also subject to rules of conduct.</li> </ul> |

|   | Selection of Judges   | Management of the Judicial "Career" <sup>1</sup>   | Internal Administration   | Resources   | Judicial Processes  | Legal Profession   |
|---|---|--|---|---|---|--|
| <b>TRANSPARENCY/<br/>ACCOUNTABILITY</b> | <ul style="list-style-type: none"> <li>• Public input is solicited as a part of the judicial selection process.</li> <li>• Appointments are adequately publicized.</li> <li>• Selection process is open and transparent.</li> </ul> <p>No future employment by the government after term expires.</p> | <ul style="list-style-type: none"> <li>• Standards for judicial performance and ethical behavior are publicized.</li> <li>• There is a process for registering complaints about judicial misconduct.</li> <li>• Public input is solicited as part of the judicial evaluation process.</li> </ul> | <ul style="list-style-type: none"> <li>• There is a process for registering complaints about administrative misconduct.</li> <li>• Adequate information is publicly provided on the roles and responsibilities of administrative officials attending the public.</li> </ul> | <ul style="list-style-type: none"> <li>• Judicial budgets, salaries, and results of audits are publicly available.</li> <li>• Judicial requests for additional resources are presented publicly.</li> <li>• Proposals for major investments in infrastructure or equipment are presented publicly with opportunity for discussion.</li> </ul> | <ul style="list-style-type: none"> <li>• The rules for how cases will be processed are well publicized.</li> <li>• Court users have access to information on the status of their case.</li> <li>• Hearings are publicly announced and open to the public.</li> <li>• Judicial decisions and statistics on case flow are publicized.</li> <li>• Ex parte meetings or conversations with litigants are expressly prohibited.</li> <li>• Press and other nonjudicial groups may comment on decisions without fear of reprisals.</li> <li>• Court services are readily accessible to the entire population, and there are no unreasonable geographic, monetary, or legal barriers.</li> </ul> | <ul style="list-style-type: none"> <li>• Information as to accredited bar members and any paralegal profession is easily available to public.</li> <li>• Disciplinary actions and disbarments are publicized.</li> <li>• There is an easily accessible process for providing complaints about attorneys' actions.</li> </ul> |

|                                   | Selection of Judges  | Management of the Judicial "Career" <sup>i</sup>  | Internal Administration  | Resources   | Judicial Processes  | Legal Profession   |
|-----------------------------------|--|---|--|---|---|--|
| <b>INDEPENDENCE<sup>iii</sup></b> | <ul style="list-style-type: none"> <li>Any external input (by other branches of government or private individuals and organizations) to the appointment process is subject to transparent rules and occurs only in accordance with established procedures.</li> <li>Evaluation of candidates is done by a body or office separate from that making the final selections.</li> <li>Judicial appointments are made as vacancies occur, not to coincide with changes in national administration.</li> </ul> | <ul style="list-style-type: none"> <li>Judges have permanent tenure or fixed, renewable appointments.</li> <li>Judicial salaries meet living wage and some reasonable proportion of good wage in private sector.</li> <li>Additional privileges (housing, vehicles, trips, training) are allocated through a transparent process with no nonjudicial input.</li> <li>Where external actors have complaints about judicial performance these can only be entered through the normal disciplinary process.</li> </ul> | <ul style="list-style-type: none"> <li>The selection and further management of administrative staff is handled through objective rules and regulations and is not subject to intervention by officials not legally authorized to provide specific inputs.</li> </ul> <p>Whether handled by an external body (e.g. Ministry of Justice) or by the judiciary itself, oversight of internal administration responds to judicial needs, not to the administrators' agenda.</p> | <ul style="list-style-type: none"> <li>Salaries and budgets cannot be reduced nor their distribution altered by other branches of government.</li> </ul> <p>When judicial workload reaches unmanageable limits, the judiciary is able to obtain more resources.</p> | <ul style="list-style-type: none"> <li>Other branches of government do not override or ignore judicial decisions, and when they do, they are subject to legal action.</li> </ul> <p>Decisions and powers accorded to the judiciary are not usurped by other governmental actors.</p> <ul style="list-style-type: none"> <li>Judiciary is able to set its own rules for internal operations; where those rules are limited by enacted law, they have substantial input into shaping the latter.</li> </ul> | <ul style="list-style-type: none"> <li>Access to the professional status is managed only according to official rules.</li> <li>Whether the judiciary or the bar association is responsible for admittance and discipline, it does this without irregular outside intervention.</li> <li>Ability of lawyers to form professional associations is reasonably open.</li> <li>Internal operations of bar associations are determined by the members themselves.</li> </ul> |

<sup>i</sup> For countries without true career systems (e.g. where judges are elected or appointed "for life" to a single position) some of these categories will not apply and should not be scored.

<sup>ii</sup> The kind of explanation offered here is doubtless needed for some other criteria. The term "job-relevant" is used because many merit-based lists, while properly objective, focus on traits of questionable importance to what a judge does – lists of publications, ability to recite laws or entire codes by memory, performance on psychometric tests supposed to reveal judicial vocation. Here, credit would be given for having objective criteria, but graded down for relevance.

<sup>iii</sup> Independence refers both to that of individual judges and of the judiciary as a whole. The dual definition does pose problems, but it is important in that it distinguishes the judicial model from that of ordinary bureaucracies.