

Public Disclosure Authorized

USING SURVEYS TO IMPROVE JUSTICE SECTOR PERFORMANCE IN THE MIDDLE EAST AND NORTH AFRICA

Public Disclosure Authorized

EXISTING PRACTICES AND LESSONS FROM OTHER REGIONS

ACKNOWLEDGMENTS

The World Bank team wishes to thank the Human Rights Trust Fund for funding this research.

This Report was prepared under the leadership of Mr. Klaus Decker (Senior Public Sector Specialist and Task Team Leader). The Lead Author of the report is Dr. Linn Hambergren. Experts who contributed to research include Ms. Fatma Raach, Mr. Ghadanfar Kamanji, Ms. Hadeel Abdel Aziz, Ms. Kenza Chakir, and Mr. Rami A. Y Rabah. Mr. Domagoj Ilic, international survey expert, provided guidance on survey design and development. Mr. Srdjan Svircev reviewed the draft. The team would also like to thank Ms. Alma Hernandez and Ms. Farishta Ali for their support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	2
TABLE OF CONTENTS	3
ACRONYMS AND ABBREVIATIONS	5
EXECUTIVE SUMMARY	6
CHAPTER I: PURPOSE, AUDIENCE, ORGANIZATION, METHODOLOGY AND EVALUATION CRITERIA.....	10
CHAPTER II: GENERAL CONSIDERATIONS ON SURVEY DESIGN, ANALYSIS, AND USE.....	12
2.1. Definition of terms	12
2.2. Participation in survey design by members of targeted organization	14
2.3. Definition of target survey population and of topics to be explored.....	14
2.4. Selection of respondents	15
2.5. Questionnaire design -- contents.....	17
2.6. Questionnaire design – length.....	19
2.7. Pretests and translation.....	19
2.8. Conducting the questionnaire	19
2.9. When targeted interviewees do not respond	20
2.10. Avoiding bias	21
2.11. Checking validity of responses.....	22
2.12. Sample size	22
2.13. Analysis and use	23
CHAPTER III: REVIEW OF SURVEY TYPES USED IN THE JUSTICE SECTOR	26
3.1. User Surveys	26
3.2. Surveys of system actors.....	29
3.3. Multi-Stakeholder Surveys.....	31
3.4. Legal Needs Surveys	32
3.5. Use of non-justice specific surveys	34
3.6. Selection among survey types, timing of repetitions, and other strategic issues	35
3.7. Lessons learned	36
CHAPTER IV: MENA’S EXPERIENCE WITH SURVEYS AND RECOMMENDATIONS ON THEIR USE	38
4.1. Introduction.....	38
4.2. Examples from MENA	38

4.3. Recommendations	47
4.4. Recommendations for global and regional action to promote justice surveys	47
4.5. Recommendations for Immediate Adoption by Countries Interested in Surveys	48
4.6. Recommendations for Later Action	49
REFERENCES	51

ACRONYMS AND ABBREVIATIONS

CEPEJ	European Commission for the Efficiency of Justice
COE	Council of Europe
ENCJ	European Network of Judicial Councils
EU	European Union
HIIL	Hague Institute for Innovation in Law
IFCE	International Framework for Court Excellence
IPSOS	No translation. An international marketing research firm
MENA	Middle East and North Africa
OECD	Organization for Economic Co-operation and Development
OHCHR	(UN) Office of the High Commissioner on Human Rights
OSF	Open Society Foundation
SAWASYA	Strengthening the Rule of Law in the State of Palestine
TI	Transparency International
UAE	United Arab Emirates
UK	United Kingdom
UNDP	United Nations Development Program
UNICEF	United Nations Children's Fund
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women
US	United States
USAID	United States Agency for International Development
WJP	World Justice Project

EXECUTIVE SUMMARY

1. The present report consists of a review of international practices and lessons on using surveys to explore justice sector performance issues and an overview of experience with justice surveys within the Middle East and North African (MENA) region. Its purpose is to assist MENA countries in using surveys to assess and improve the quality and quantity of their justice services. Secondary audiences include governments and justice sector organizations in other regions, the World Bank and other donors, and anyone else interested in how surveys can identify and promote ways to create better services.
2. The use of surveys is now a recommended practice for justice institutions as a means of assessing performance, identifying the results of recent legal and operational changes, and determining where improvements are still needed. Surveys complement the equally important and more common reliance on statistical analysis for these purposes, and should replace, but often have not, a dependence on sheer intuition or “what everyone knows.” Even where surveys are accepted, there is considerable misunderstanding as to what a good survey requires, both in technical terms and as regards the involvement of the principal client – the agency or institution about which responses are collected.
3. A survey is a means of collecting feedback on agency performance from stakeholders (users, non-users, internal and intermediary actors). It taps their experience and perceptions of performance quality and its conformity to their expectations. Among the many types of feedback (interviews, focus groups, suggestion boxes, man-on-the-street interviews) a well-designed survey is the most reliable and scientific. However, reliability depends on following the rules as regards sample selection and size, avoidance of bias, questionnaire design and testing (including careful translation to local languages where applicable), mode of application, checking validity of responses, and dealing with non-responses and segments of the population that for one reason or another are not captured in the initial sample.
4. These rules are the same for surveys in justice as for any other kind. Random (probability) samples/surveys are the gold standard, but logistical issues may require some “bending” of the rules; these should always be acknowledged in the accompanying methodological annex. Quasi-probabilistic surveys and others that are still further removed from the gold standard still can provide important information but should never be taken as representative of the entire target population. In some cases, these less technically rigorous alternatives can be intentionally or unintentionally misleading; while some information is usually better than none, bad information can be harmful and push policy in undesirable directions.
5. Justice (and other) surveys assessing service quality come in several types: user surveys, surveys of internal actors (judges, staff, lawyers), multi-stakeholder surveys (capturing all the above and some non-users) and legal needs surveys that ask respondents about their justiciable (potentially legal) problems and whether and how they resolve them. Each has a different purpose and the selection among them depends on what one wants to know, funding and logistical constraints, and to some extent, what the target institution can reasonably attempt to resolve. Even when a country has not done or had done a dedicated justice survey it will often have results from some international surveys and survey-based indices (e.g., the World Justice Project, Transparency International and in MENA, the Arab Barometer). These can be a useful source of ideas as to what a national survey should explore.

6. The sections in this report on MENA specific experience are short, as so few justice surveys have been done in the region. In some countries this is part of a wider disuse of surveys to measure performance in any sector, but it is also influenced by European practices (which also came late to surveys); a preoccupation with unresolved conflicts over fundamental issues like judicial governance, independence, and distribution of functions within the sector; a focus on “modernization” often emphasizing only automation; limited budgets; and a distrust of public opinion as a measure of the quality of performance.

7. Divided into three categories – self-initiated surveys, those done by donors for their own (usually project-linked) purposes, and inclusion of some MENA countries in international indices with survey components – few examples could be identified. Conceivably given a regional tendency of less than total transparency (seen also in a failure to publicize basic performance statistics across much of the region), a few examples may not have been captured by this report. However, it is unlikely that many were missed.

8. Among country self-initiated surveys, nearly all with donor technical and some financial assistance, examples include a 2011 survey in Jordan, one in Tunisia in 2013 as part of a national consultation on justice reform, one or more in Morocco as part of surveys on enterprises, and a recent survey initiated by Egypt’s State Council (administrative jurisdiction). Four UAE members also contracted a justice needs survey, which was never published, and have begun online “surveys” to evaluate online programs – the online exercises are not technically surveys but are positive examples of seeking feedback from system users.

9. Roughly the same number of donor-initiated surveys were identified – one in West Bank and Gaza, two and possibly more in Egypt, one in Morocco, and the five or more user needs surveys conducted by the Hague Institute for Innovation in Law (HIIL). None of these surveys could have been done without country acquiescence, but they are distinguished from the first group, not only by total donor funding and implementation but also by the relative absence of country participation in designing the questionnaires. However, it bears mention that the Morocco example fortuitously coincided with the country’s emphasis on the principal theme, trust in public sector institutions, and thus stands a good chance of being used by the government authorities.

10. Inclusion of MENA countries in international surveys and survey-based indices is spotty. There is no example, except possibly the Global Competitiveness Index (with only a few questions on justice in its Executive Opinion Survey) that includes even half of the MENA countries. Exclusion of others has various explanations, but a refusal to allow entry arguably accounts for a part. Although except for the World Justice Project (covering only eight countries) none focuses on justice, the results even for the others do indicate areas where the covered nations could improve sector performance.

11. Surveys done by donors for their own purposes and international indices have other impacts – in the first case justifying or promoting changes in donor activities while facilitating dialogue with local actors, and in the second, expanding global knowledge about regional justice systems. Over time the latter could motivate greater attention within the region as to sector shortcomings, but it still would be important to determine how that impact could be hastened.

12. Recommendations are provided in three sections: one for donors and others interested in promoting survey use within MENA (and elsewhere), a second focusing on immediate actions to be taken by countries that are interested in expanding or introducing their own surveys, and a third on longer term actions for these same countries. The first section, for donors and others interested in expanding the use of surveys is intended to help overcome a general impression, hardly limited to MENA, that surveys primarily convey bad news and have no real role in modernization efforts. This impression needs overcoming and focusing on MENA would be an important step in that direction.

13. Even when initiated by country authorities, it is important that key local stakeholders participate in all stages of the undertaking, especially during the preliminary analysis of survey results; this is no less important for surveys conducted by donors and might be considered by international comparative efforts. However, given different arrangements in each country, the key stakeholders will vary. In some instances, a Ministry of Justice will take the lead; in others it may be a Judicial Council, or one or more apex courts (e.g., a Supreme Court and/or a Council of State, a separate set of religious courts) and/or the head of Public Prosecution. Other relevant stakeholders (e.g., bar associations, judges' association, civil society organizations working on justice should be consulted in the process as appropriate.¹

Recommended changes to global and regional discussions of surveys, primarily for donors, internal stakeholders and others interested in overcoming surveys' image problem	
Title	Details
Globally, stress importance of feedback mechanisms	Link their use to design reforms advancing values important to governments like foreign investment, national reputation, and political stability.
Globally, emphasize modernity of seeking feedback	Stress that the modernization inputs (ICT, training, legal change, and infrastructure) can be improved, and investment better directed, with knowledge of stakeholder reactions, before and after any changes.
Regionally, sponsor workshops in MENA	Focus on organization and uses of feedback mechanisms, noting the advantages of surveys, and invite representatives from judiciaries in other regions to discuss their experience.
Provide funding/technical assistance	Where interest develops, offer funding and/or technical assistance, depending on country situation to encourage early efforts. Involve government/judicial agencies from the start.

Immediate Tasks for national authorities/judiciaries with interest in survey use	
Title	Details
Review existing surveys with data on specific country	Countries with any interest in surveys likely have information from some national or international surveys. This can serve to focus discussion on the questions deserving more exploration, especially from these and other feedback mechanisms.

¹ Which agencies are essential depends not only on local organization (which for example in Egypt might include five separate apex judicial bodies and the Ministry of Justice) but also the survey topic. If it is only performance relating to commercial or administrative cases, then agencies not involved there would not be included.

Organize a workshop or national consultation to review ideas about necessary information vis-à-vis perceived reform needs	As the recommendation is directed at countries with some information from user (and non-user) feedback/surveys, it should provide some important ideas and suggestions. Recommended reforms are less likely to come from an eventual survey, but the latter can help evaluate them and point to where they are most needed.
Design new (first?) within country survey	Use information from first two exercises to inform design. The resulting survey should be publicly announced; both the announcement and the questionnaires should use the major local language/s. Depending on the type of survey, mobilizing donor funding and technical assistance may be an option.
Publicize results	This is important and should not be ignored. Donors and others with experience with justice surveys can provide advice and models as to how best to address different audiences.

Tasks to be taken later within countries	
Title	Details
Develop a sector-wide assessment combining survey data with other information sources	There are models for such assessments, and they do suggest the variety of sources – dedicated surveys, statistical analysis, targeted interviews and focus groups, process mapping, existing evaluations and surveys, and others – that can be used for an overview of issues.
Build surveys into justice reform/modernization plans	Just like statistical analysis, but measuring different things, surveys should be part of any reform program, to identify issues, track results of changes and catch emerging trends. Absent these inputs (and both are needed) a plan is moving blindly. Statistical analysis, with a good database can be continuous; surveys should be scheduled periodically, less often than annually (unless extremely focused) but at least every 5 years.
Use surveys to review treatment of vulnerable groups	Relevant vulnerable groups would need to be identified and surveys should be designed to capture their experience and opinions as well.
Plan for series of surveys	One survey captures a baseline; multiple surveys at staggered intervals tap progress, emerging or disappearing issues, ² and other trends relevant to longer term planning. Thus, a modern judiciary incorporates these as well to assure it is on the right track and not working off initial impressions.

² These are often forgotten, but in justice a sudden flood of some type of case or client may disappear once conditions change. This is important to capture to avoid planning to resolve issues that no longer exist because of external changes in policy, laws, or socio-economic conditions.

CHAPTER I: PURPOSE, AUDIENCE, ORGANIZATION, METHODOLOGY AND EVALUATION CRITERIA

14. The use of surveys is now a recommended practice for justice institutions (International Framework for Court Excellence, IFCE, 2020). It complements the more common reliance on organizational statistics to evaluate performance and identify areas needing improvements. Analysis of a good statistical database, assuming one exists (unfortunately not always available), can provide information on total caseloads and their distribution among work units (e.g., courts, prosecutors' offices) and their members; time to and types of resolution; outcomes (including penalties assigned or amounts awarded); and which cases and users get left behind or experience unusual delays. However, even the best database offers no insights into ease of access for different types of users; satisfaction with treatment, process, and results; distrust (some of it merited) of organizational personnel; and perceptions of bias toward specific groups. It also gives no hint as to how system operators (ranging from judges and public defenders, through their staff to private attorneys and others) feel about their jobs, the efficiency and integrity of their colleagues; and their own notions as to how things could be improved.

15. The present report consists of a review of international practices and lessons on using surveys to explore these issues, World Bank fieldwork on experience within the Middle East and North African (MENA) region, and a set of recommendations addressed to MENA countries. Its purpose is to assist MENA countries in using surveys to assess and improve the quality and quantity of their justice services. Secondary audiences include governments and justice sector organizations in other regions, the World Bank and other donors, and anyone else interested in how surveys can identify and promote ways to create better services. Following this brief introduction, the report is organized as follows: Chapter II on general survey practices; Chapter III on the different types of surveys used in the justice sector; and Chapter IV on fieldwork findings and recommendations applicable to MENA countries.

16. In reviewing international experience, the emphasis is on identifying good practices, which here refer to several aspects: the use of different types of surveys, questionnaire and sample design and other technical details; use of surveys focusing on other issues but including questions on justice; recommended frequency of the same survey type or alternations among different ones; and analysis and dissemination of results. Although some examples of good practices are drawn from fieldwork findings, the search for good practices covers examples from all regions, legal traditions, and country contexts. It draws on a search of available literature on the topic; in person, phone and email interviews; and World Bank and other donors' experience.

17. Conducting a survey of user and system actor views, analyzing, and publishing its results can be a daunting experience for organizations used to "operating independently" even of the public that depends on their services. In an era where the quality of public services has come to the fore it is both symbolically important and a way to realize that mandate, while also challenging the less scientifically based criticisms and praise that often serve as the only external source of assessments.

18. Although technical discussions of survey design abound, information specific to their use in the justice sector is relatively recent and often limited to individual countries or courts. Nonetheless, the generic technical recommendations on designing and conducting surveys are applicable everywhere, with justice being no exception. They are briefly referenced in Chapter II and accompanied by a short list of key sources. However, in looking for good practices, more attention is given to examples that address justice-specific issues – how to select questions that will be useful both in evaluating performance and designing change programs; how to select respondents, encourage their participation, and deal with low response rates; how to conduct technically sound surveys in countries lacking adequate registries of all citizens or even of justice employees; how to deal with the tendency of sector actors to dismiss surveys as “only opinions” from those who do not understand the justice system; and how to analyze, interpret, and disseminate the results.

19. The present report reviews experience with justice surveys internationally and in MENA; methodologically it examines what has been done and to the extent any evaluation is conducted, this principally is against international criteria developed by survey experts. While the latter may disagree on details, there is a well-established reference framework allowing for identification of a sound survey and the deviations from that standard.

20. Added to this is a second set of evaluation criteria – whether justice surveys are in fact conducted, and if so, whether they are used as part of an improvement plan. As noted, surveys provide an important source of information on justice performance, areas where attention is needed, and how stakeholders of various types assign priorities to possible actions. In this they are as important as the more usual analysis of organizational databases. Determining the existence of justice surveys is straight forward. In MENA, those identified seem to hold up well against international standards. But the fact remains that justice surveys are not common in the region, and they are often not country-driven. At the same time, determining whether and how they are used is more challenging given the low level of transparency in the sector across much of MENA. Wherever possible, this report mentions when survey data are actively used by countries to inform reform initiatives and management decisions.

CHAPTER II: GENERAL CONSIDERATIONS ON SURVEY DESIGN, ANALYSIS, AND USE

21. This chapter is not intended as a do-it-yourself guide to conducting surveys; it simply offers an overview of common practices and issues that any agency considering doing a survey should know. Surveys, in justice as in other areas are highly technical exercises, for which reason those sponsoring them nearly always hire an experienced firm or academic institution to carry them out. This is true even of organizations like the Hague Institute of Innovation in Law (HIIL) that have their own internal experts and methodological approaches developed, in HIIL's case, while conducting over 15 legal needs assessments. As elaborated below, questionnaire design also requires expertise in both the general principles and the substantive topic, including country specifics.

22. It bears mentioning that there is still an unfortunate tendency to assume that anyone who can draw up a list of questions and ask them of whomever is conveniently present is doing a survey. This "method" can produce some interesting responses, but they only represent the ideas of an arbitrarily defined small group.³ The purpose of a survey is to systematically tap the perceptions, opinions, and experiences of representatives of a specific population – perhaps all those who use courts or a legal assistance agency. Well applied, survey techniques allow this to be done without asking questions of every court or legal assistance client. This is one part of the technical expertise required, along with questionnaire formulation, and analysis of the results.

23. As noted, there is a wealth of information on survey design, analysis, and use, and to repeat it here would take excessive space. Instead, a few general cautions and recommendations are provided, drawing on literature listed in the bibliographic annex. As noted, when surveys are funded by donors, this is information they already have, but however, financed, it is important that organizations either conducting their own surveys or receiving them from donors, understand the basics, to ensure both their informed participation in survey formulation and their optimal use of the results.

2.1. Definition of terms

24. To facilitate the following discussion, key terms used are defined below:

25. The **justice sector** is the collection of agencies and actors involved in providing justice services -- i.e., dispute resolution, including criminal cases, as well as in some systems, authentication of documents, company and other registration, and provision of information on how to access these services. Here we are primarily interested in the judiciary (all branches, all jurisdictions), a Judicial Council (if separate from the judiciary), the Public Ministry (prosecution and sometimes involvement in non-criminal proceedings),

³ Also, the various non-technical uses of the term "survey" (for example as an overview of some phenomenon) in English may cause further confusion. A questionnaire sent to a few experts may "survey" their views and knowledge, but it is not, technically speaking, a survey.

any government-funded legal assistance offices or programs, and the Ministry of Justice. Other entities within the sector (bar associations and lawyers, arbitration centers, police, notaries, private bailiffs, prisons, NGOs, and so on) are less likely to sponsor surveys to evaluate their own services but may be included as respondents in surveys financed by others.

26. A **survey** is "the collection of information from a sample of individuals through their responses to questions" (Check and Schutt, 2012; 160). This is a minimal definition, and as elaborated below, the technical soundness (and thus reliability) of a survey depends on meeting various other conditions.⁴ A survey, technically sound or not, can record respondents' opinions/perceptions or real experience. It may be conducted face-to-face, by telephone, electronically, or with the questionnaire provided by physical mail or made available within the court (or other agency) building.

27. **Questionnaire:** this is the standard list of questions provided to all respondents to a survey. Standardization is critical in allowing tabulation, analysis, and comparison of results. Surveys may include some "open-ended" questions, where the respondents provide answers in their own words, but for most questions, answers comprise a closed list of options among which the respondent must choose.⁵

28. **Samples:** these are critical to ensuring survey quality as well as keeping costs reasonable, especially where the target population is large and widely dispersed. Applying the questionnaires to all citizens or only all service users or employees (in effect a census) is not only impractical but also unnecessary (and possibly less accurate) as opposed to focusing on a smaller group (sample) that replicates the profile of the entire targeted universe. Ideally the sample is drawn at random from a list (**sampling frame**) of all members of the targeted population – for example all citizens over 18 years of age, all individuals and/or enterprises involved in a court case, or all first-instance judges. In complex surveys, a multistage stratified approach uses separate sampling frames for each sub-sample (e.g., businesses, court employees, court users). Where such a list does not exist or is known to be incomplete, other methods of random or quasi-random selection may be used. Depending on the intended use of the survey, some of these criteria can be relaxed as further discussed in Chapter II, but they should never be set aside completely.

29. **Complementary additions to surveys:** a survey provides important information on respondents' experience, perceptions, expectations, and (unmet) needs, but it can benefit from elaboration through complementary approaches. Some of these are discussed below as used to expand and qualify information derived from a survey. Aside from statistical analysis (either complemented by or a complement to surveys), they include focus groups, detailed interviews with some respondents, and

⁴ As universally there are many examples of "surveys" that do not meet all or many of the conditions for technical soundness, the minimal definition was used as the best way to elaborate on these issues. Restricting "survey" only to the most scientifically rigorous examples would leave us with no term for or way to discuss those that do not qualify.

⁵ Respondents may be asked to rank services numerically (1-10), give a yes-no answer, or choose among a short list of ranked categories (e.g., "very often," "sometimes" "never). In this a questionnaire differs from an open-ended interview.

invitations to interested individual and groups to provide suggestions and comments. They can serve as a check on respondents' answers,⁶ a way of explaining certain reported behaviors, or a source of more specific ideas as to what works or needs improvement.

2.2. Participation in survey design by members of targeted organization

30. Many justice sector institutions have little experience with surveys, and even those that do rarely have the internal expertise and staff required to carry one out. Thus, they nearly universally rely on specialized organizations to conduct the survey as well as various degrees of design and analysis. However, no matter who finances or implements the survey, if members of the target institution do not understand the process and participate in the elaboration of questionnaires and procedures, their use of the results will be limited. Thus, just as sector institutions are cautioned not to do it on their own, they also are advised to participate in survey planning and analysis. Good implementers understand this and will insist on full ownership and participation from the start.

2.3. Definition of target survey population and of topics to be explored

31. These are the two most important questions; which comes first is a judgment call. No matter who funds or implements the survey, the answers should be endorsed by the “recipient organization” and its members. Outside funders and contracted implementers can provide sample questionnaires and ideas about the target population, but ownership of the results belongs to the organization that will use them. Target populations can include service providers (e.g., judges, prosecutors, court staff), users (e.g., citizens, businesses), and intermediaries (e.g., attorneys).

32. If this is a “court user” survey, is the target population all those who visit the courts, only those who have done so in connection with an on-going case, or only those with specific case types (e.g., family violence, commercial, criminal)? Whatever the answer, what does the court (or other organization) want to know about their experience and perceptions? Typically, questions focus on issues like ease of access and use; quality of treatment and services; impediments like cost, geographic distance, and uncertainty, distrust, or lack of information about the formal rules; and satisfaction with outcomes. Respondents are usually not accurate reporters of details more easily captured in a good **Case Management Information System (CMIS)** like time to resolution, number of hearings and adjournments, or costs. However, where one does not exist, some approximation can be obtained through surveys, and as noted, even their inaccurate responses are important signs of where better communication may be needed.

⁶ For example, many attorneys report lengthier delays in case processing than the statistics indicate. This is called **selective recollection** – the most extreme example is the most remembered. Users may not remember accurately the number of hearings in their cases, so statistical analysis can serve as a check on their responses. However, users' impressions, even if inaccurate, are important because they are likely to affect their behavior and even their willingness to access formal institutions.

4.4. Selection of respondents

33. Ideally, a survey should draw respondents at random from a sampling frame that incorporates all members of the targeted population (be they court users, staff, or ordinary citizens). This ensures each member has an equal chance of being included in the sample and allows generalization to the entire population. Problems often begin with a sampling frame that under or over represents certain categories or lacks accurate contact information. Construction of a sampling frame is easiest in surveys of justice sector actors and most difficult for users for whom information is likely to be missing or incomplete. However, in many countries, there may not be a list of all employees from which to choose “at random;” and/or adequate contact information (mailing or email addresses or phone numbers).

34. Random (probability) selection has been called the gold standard for surveys. Even in complex, multi-stage surveys, it can still be maintained within each stratum so long as they have adequate sampling frames. However, where incomplete sampling frames, logistical issues, and similar impediments defeat pure random selection, one of a series of alternative approaches may be used. These are considered quasi or non-probability samples, but they vary considerably in the quality of their results. Where lists of users are absent, incomplete, or potentially skewed, one frequent solution is the court (or legal aid/prosecution) door capture in some random fashion (e.g., every 10th person) of potential respondents as they exit. As further discussed below (see box on purposive sampling), courthouse door interviews, like many such solutions, do pose risks, which if recognized can be partially mitigated. Another solution, especially for very large countries is multi-stage sampling as applied by HIL in its legal needs surveys. Survey experts can also suggest additional alternatives.

HIL’s use of multi-stage sampling in underdeveloped and/or very large countries with dispersed populations

Although HIL uses this approach for legal needs surveys, it is applicable to other survey types in the absence of good registries of targeted populations. HIL has conducted over 15 legal needs surveys, typically in less developed countries with large populations and geographic areas (e.g., Nigeria, Kenya) and even in less populous states with highly dispersed populations (e.g., Fiji and other archipelago nations). Its solution has been “multi-stage” sampling whereby there is first a selection of certain geographic districts or regions (typically not at random, but based on population size or demographic composition), then a random selection of internal districts, followed by a random selection of respondents within each one. HIL’s explanation of selection procedures at the lowest level is sketchy, and this is an area where bias can enter. Other organizations have reported that their field interviewers left to their own devices opt for those easiest to locate or living in the most desirable areas. This is not to suggest HIL’s surveys suffer this flaw, but only to emphasize the need to include these details in a methodological annex or section.⁷

⁷ HIL’s reports include a brief section on methodology, but it contains less than the usual methodological annex, which, if it exists, is not referenced. See a selection of HIL’s reports in References.

35. The preferred quasi-probability alternatives, like the HILL example, incorporate some random elements and are typically larger than needed for a simple probability sample. HILL’s samples average about 6,000 respondents whereas 1,000 or 2,000 would be sufficient for a simple probability sample. Larger size is advisable where target population parameters are unknown or when the aim goes beyond “average” views to those of smaller subgroups, as further discussed below. To be avoided are what are called haphazard or convenience samples –the “man-on-the-street” interviews conducted by the media or **opt-in** and sometimes paid respondents pre-selected because of their willingness to participate. Most of the justice surveys reviewed here and listed in the references (including those in high income countries⁸) made compromises with fully random selection, either for lack of adequate sampling frames, costs, or because of operational issues like low response rates. They nonetheless provided important, if not broadly generalizable information about the opinions and experience of system users, potential users, and employees. Even flawed applications can provide valuable insights to interested governments and institutions. For example, what eventually had to be called a “nonrandom” sample of 6,000 visitors to roughly half of Colombia’s justice houses (USAID, 2013) still generated theretofore unavailable user evaluations of the menu of services offered.⁹

36. The required degree of selection rigor depends on the use to which the survey is put. Predicting an election outcome requires hewing to the rules, but in soliciting feedback on or even evaluating court or prosecutorial services, so long as the potential for selection bias (e.g., limited geographic coverage, court-door capture conducted unsystematically or only in the most easily reached areas) is recognized and to the extent possible avoided, and the number of respondents, if not 6,000 is adequately large, the results should still be useful to the court or other justice institution authorities. Local standards, often set by a planning institute, may pose more rigid rules, however, and these will have to be followed.

“Purposive sampling” and other non- or quasi-probabilistic means of obtaining user feedback

The term “purposive sampling” is used by IFCE (2000; 29) to describe a simple technique to give “immediate voice to all court users without jargon or methodological barriers.” The recommended approach (which includes a sample questionnaire piloted in Macedonia) involves interviewing every user of the courthouse(s) on a single day (or on several days if the courthouse/s have a low volume of users or where certain proceedings like arraignments are only conducted on specific days). The Colombian example falls into this category despite the much larger sample and the likelihood that not every user was interviewed. For the purpose of getting user feedback, there are a variety of still less complex non-survey approaches, for example suggestion boxes, questionnaires left at the courthouse door, community meetings, and invitations to participate in focus groups. Sector institutions interested

⁸ Even in high income countries courts usually cannot or will not provide a list of all users. Consequently, many of their surveys start with administrative citizen or business registries and require a further step to identify those who have used the courts. Where such registries exist, there are always some excluded groups (often as few as 1 or 2 percent) who, if of interest, may require additional samples.

⁹ The problem was the non-systematic selection of the justice houses included and of exit interviewees, the latter probably conditioned by the hours the interviewers could work and some sort of daily quota, with no further instructions as to selection procedures. There was no indication of any intent to bias the “samples,” but simply of a failure to consider how what was convenient for the organization conducting the survey might affect the results.

in obtaining user feedback often use several of these methods, each of which adds something to improve overall information as well as enhancing public contacts and trust.

For a judiciary or other sector institution without the funds to finance a formal, probabilistic survey, these methods can be appealing while indicating an interest in knowing what users think and experience. The advantage of a probabilistic sample is the ability to generalize to an entire universe of users, something that non-probabilistic surveys and non-survey methods do not allow. Nonetheless, following some of the rules applied to probabilistic samples can improve the quality and utility of the results. These include:

- Adequate design and testing of questionnaires before they are applied (see sections below on survey content, length, and pretesting).
- Where interviews are done face-to-face or by phone, adequate training and monitoring of the interviewers.
- Pre-testing of the approach (including interviewer performance) to guard against bias, off-putting presentations or introductions, and other factors that may discourage or influence user responses.
- Guaranteeing anonymity while incorporating information on important characteristics like gender, type of court involvement (party, lawyer, witness) and of case (including simple administrative matters).
- Recording of response rates, and if possible, characteristics of non-respondents
- Where conducted in specific courthouses or regions, their selection to be as representative as possible

Both in reporting and using results, organizations applying them should be careful about overgeneralization as these quasi or non-probabilistic approaches do not permit interferences about entire populations. In short, the most important step is eliciting user feedback. While probabilistic surveys are the gold standard, for those who face insurmountable logistical obstacles or cannot afford them (or do not have a donor offering financing) there are alternatives, which if done carefully and with sufficient understanding of their limitations, can provide information critical to making and even assessing service improvements.

2.5. Questionnaire design -- contents

37. As surveys are both expensive and time consuming, a first recommendation is to ensure what is asked has a purpose. This means considering what one needs, not might like, to know. This is important not only to reduce costs added by superfluous information, but also to encourage responses from the target population. Experience demonstrates that lengthy questionnaires can cause respondents to stop answering questions or diminish the quality of their answers. Donors can and often do finance longer questionnaires and hire survey companies to apply them. However administered, long questionnaires are only advisable with face-to-face or possibly telephone interviews in which interviewers can keep the respondent engaged. ¹⁰

¹⁰ However, care must be taken to avoid individual interviewers adding too much of their own interpretations.

38. Second, survey language should be appropriate for the targeted respondents. For the general public, as opposed to subject experts (e.g., lawyers and judges), language should be simple and easily understood. Use of the local language of the target population is key at both testing and implementation stage. Specific recommendations for the US (Legal Service Corporation¹¹) range from what could be understood by a third grade to sixth grade student. Face-to-face and phone interviews have an advantage here as the interviewer can provide explanations and in the former case, use visual aids as well. It is obvious, but merits mention, that in countries with high levels of functional illiteracy, face-to-face interviews may be the only reasonable choice. And even where literacy levels are high, written, self-administered questionnaires may exclude responses from a small, but potentially significant portion of the population.

39. Third, when mixed groups are surveyed (e.g., general public, court users, sector personnel of different types) questionnaires should be tailored for each group – an exception can be made when administered by an interviewer who can skip the irrelevant parts for the specific interviewee. Otherwise, using a single questionnaire for all groups can discourage answers or provide those that are not relevant for the specific respondent. This does not preclude the use of some of the same questions for all (to allow comparison among them), but these questions should be written in the simplest terms.

40. Fourth, question wording and order can influence answers. Obviously, leading questions (those pushing the response in one direction) should be avoided, but there are additional factors to consider. Immediately prior questions can influence subsequent answers while alternative wording has also been shown to radically change results. In one survey asking respondents whether amounts spent on prison populations were too much, too little, or exactly right, there were major differences in responses among three wordings: one with no amount given, one with the total amount spent, and one with the amount per prisoner.¹² There is no obvious answer as to which best reflects popular opinion, although the implications for those interested in prisons is quite clear. It is likely, moreover, that using two choices sequentially would also have an impact on the second answer.

41. Fifth, provide a “prefer not to answer,” “do not know,” or “not applicable” alternative for those responding to closed questionnaires. There is no point in forcing answers respondents cannot or do not wish to give. It will only frustrate or antagonize them as well as, if they feel they must choose something, provide inaccurate information. However, the number (and demographic characteristics) of those selecting these options constitutes important information that can also be analyzed to determine, where for example, potential users require better information on practices or areas where staff should be better attuned to users’ sensitivity.

¹¹ At <https://www.lsc.gov/grants-grantee-resources/resources-topic-type/comprehensive-needs-assessment-priority-setting>.

¹² Katz et al. (2008), citing Michigan State University research conducted in 1997. Interestingly the greatest difference was between the total amount and that per prisoner although the two expenditures were equivalent.

42. And finally, although respondents should be guaranteed anonymity, the questionnaire should include basic demographic information (age, education, gender, occupation, and so on). If questions about income are too culturally sensitive, face-to-face interviewers may be asked to provide additional information, based on observation, that give some indications (e.g., type of dwelling, visible possessions like cars, computers, smart phones).

2.6. Questionnaire design – length

43. Length directly affects time taken to answer, but the Ideal length hinges on how surveys are conducted. CEPEJ (2010) recommended that a questionnaire should take the respondent no more than 15 minutes to complete.¹³ CEPEJ, however, made its recommendation for self-administered questionnaires – those sent by internet, mail, or left for the respondents to pick up and fill out at later. Face-to-face applications with a recommended maximum response time of 30 to 40 minutes, because the interviewer can keep the interviewee engaged. Phone interviews lie somewhere in between. Both methods have other issues, but length seemingly is not a major one.

2.7. Pretests and translation

44. All questionnaires should be pretested with the same respondent types and in the form (e.g., by phone, online, in person) in which they will be applied. Questions not understood or posing other problems should be reformulated and tested again. As noted above, factors like specific wording and sequencing can also change responses even when the questions are easily understood.

45. Questions are usually first drafted in the language of their authors but may have to be translated to other languages. Where the initial draft is also in the language of those to whom they will be applied, this is less problematic although pretests are still necessary to ensure questions are interpreted as intended and that they elicit sufficiently varied answers to be of value. Where all respondents give the same answer, thought might be given to removing the question. Where the drafter does not work in the local language or where several languages are involved, questionnaires must be translated, and the pretest should also focus on uniformity of interpretation across the linguistic versions. Here reverse translation (translating back to the original language) can be useful.

2.8. Conducting the questionnaire

46. There are several choices here, the utility of which depends on country conditions and in multi-stakeholder surveys may vary with each type of respondent. Face-to-face contacts are preferable but may not be possible because of costs, difficulty in locating those selected or other logistical issues. Also, for some very sensitive issues (see IPSOS 2010; Yan and Cantor, 2019 on respondents involved in criminal proceedings) respondents may be intimidated with a face-to-face interview. Thought also should go to where an interview will be held – respondents may be less forthcoming in their office (assuming they work

¹³ Examples from CEPEJ (2010) took even less time – 8 – 10 minutes. CEPEJ’s model survey has been adopted by others (see USAID, 2020).

within the sector) but answer more freely in another venue. The other choices, only if contact information is available, are by regular mail, email, and phone. Exit interviews while often done face to face can also distribute questionnaires to be filled out and returned later or get contact information for a later phone or email interview.

2.9. When targeted interviewees do not respond

47. Response rates are increasingly an issue for all surveys because of questionnaire fatigue (in countries like the US where every online purchase seems to elicit a satisfaction survey), distrust based on experience, or, conversely, unfamiliarity with the practice. None of the examples reviewed for this report registered response rates of 100 percent, ranging from 15 percent or lower¹⁴ to 96 percent. However, response rate calculation is a tricky undertaking with much depending on the means of calculation (Law and Justice foundation, 2012: 12) – for example whether for those in the initial sample actually reached, or those reached who agree to answer the questionnaire.¹⁵ Some of the highest rates are reached through the questionable practice of using “opt-in panels,” whereby the sample is drawn from potential participants who already indicated willingness to be interviewed (Pleasant et al.; 11-12). This is not a recommended practice given the strong suspicion that those who opt in do so for reasons not shared by the entire target group.

48. Response rates tend to be lowest with email questionnaires (Hamlyn et al. 2015, citing a 30 percent response rate for emailed surveys in the UK, or the ENCJ, 2019 judicial survey with an average response rate of 20 percent) or those distributed “at the courthouse door” to be completed later. However, in the US, responses for general phone interviews are now said to be in the lower one-digit category.¹⁶ Face-to-face interviews tend to do better but nowhere near 100 percent.

49. Low response rates pose two issues for surveys: ensuring an adequate sample size and avoiding sample bias. Sample size (typically defined before the survey begins based on target group size, known variations of interest within the target population, and available budget) is more easily dealt with. A common solution, where budget does not preclude it, is to draw an initial randomly selected sample large enough to produce the desired target size once the anticipated response rate is considered. Thus, the

¹⁴ This was the ENCJ (2019) survey of CoE judges, in which response rates in four countries (Romania, Croatia, Germany, and Spain) were either at 15 percent or lower. For some reason (perhaps how the survey was conducted) the ENCJ considered 15 percent adequate; other sources (IFCE,2020) accept a minimum of 50 percent. However, holding surveys to that number would mean that many should not have gone ahead.

¹⁵ The UK and Australian surveys seem to use the first formulation for phone surveys; what they do for courthouse-door surveys is not explained. Since with phone surveys in particular, the biggest gap may be between those identified in the sample and those who answer the phone, a response rate based on those reached who agreed to be interviewed is likely to be much higher. E-mail surveys pose their own calculation issues, but assuming nearly all receive the emails, it would only make sense to calculate the response rate on those who fill out the questionnaire, presumably in many countries (ENCJ 2019 as an example) a relatively lower percentage.

¹⁶ This is probably a result of survey fatigue, a surfeit of spoofed calls, and constant warnings not to answer calls from unknown numbers. However, even the 2008 phone survey in Australia required 500,000 tries to get its targeted sample of users (Pleasant et al; 11). Problems encountered include not having the right person answer a landline call and for cell phones registered to the intended individual, their turning off their phones to avoid distractions.

civil court user survey conducted in the UK in 2015 (Hamlyn et. al., 2015) started with a sample of 8,464 individuals to achieve 2,213 completed questionnaires. The actual response rate of 26 percent was as anticipated “not untypical for government self-completed surveys.” For exit interviews or fieldwork that does not use a sampling frame to select respondents, a low response rate can be compensated with a substitution plan that replaces a non-respondent with another randomly selected interviewee (e.g., in an interval system using every 10th person, if the 10th interviewee opts out, the 11th is taken instead, with subsequent every-ten-person intervals starting there). This method typically assigns each interviewer a quota. If the random selection system is respected, this should not skew the sample unless the interviewers’ working hours do not coincide with those of a significant proportion of service users.¹⁷

50. These adjustments do not, however, deal with response bias, the tendency of response rates to vary among population sub-groups. This is a more serious issue, and harder to deal with, in part because it may not be identifiable from the start. Depending on the country context, different methods of survey application can reduce (but not eliminate) some types of non-response bias. Thus, in the pretest, those applying the survey are advised to check which methods elicit more responses and whether they over or under-represent specific groups. Where response bias is detected after the fact and identified as potentially important, **booster samples**¹⁸ or separate surveys may be added for the affected groups.

2.10. Avoiding bias

51. Low response rates are not the only source of bias. Sampling design has been a recognized problem ever since the US public opinion polls before the 1936 presidential elections. The largest of the two polls (10 million sample size and 2.4 million responses), done by the *Literary Digest*, failed in its predictions because the sample was drawn from a biased list – people with telephones, car club membership, and magazine subscriptions producing a “universe” that skewed toward the upper and middle class.¹⁹ The low response rate (25 percent) is believed to have exacerbated the issue because responses were also skewed. A smaller sample (50,000 respondents), used by George Gallup made a more accurate prediction because of how it was constructed:

52. Gallup had worked out what kinds of personal characteristics (including state, urban/rural residence, gender, age, and income) related to voting patterns, and used these in the design of his sample. He set quotas for the numbers of individuals needed for each type of respondent, so that the number surveyed would reflect the population distribution.²⁰

53. The *Literary Digest*’s effort at full coverage of a flawed frame and Gallup’s representative quota approach were later replaced by random sampling, but issues of bias remain, whether in the

¹⁷ The Australian surveys using this method e.g., Family Court 2015) do not mention this issue, but clearly if certain users tend to arrive earlier or later than the interviewers’ schedule or however long it takes them to fulfill their quota, this could pose a problem.

¹⁸ Boosters add more interviews from the underrepresented sub-group to ensure there are enough of them to ensure adequate representation.

¹⁹ Katz et al. 2008.

²⁰ https://amsi.org.au/ESA_Senior_Years/SeniorTopic4/4b/4b_2content_4.html.

construction of the sampling frame or in its application. A frame that excludes no one and avoids duplicate entries is nearly impossible to construct. More importantly exclusion is often systematic even in presumably complete lists of all citizens, businesses, or court users. Outdated or absent contact information is also rarely random.

54. Means of application can also introduce bias, most notably in internet or phone surveys. Even in countries with high levels of connectivity, willingness/ability to respond to internet surveys can vary significantly by population group. In countries (the US for example) where many citizens have either landlines or cell phones, but not both, choice of one or the other will introduce significant bias by both age and location. Given the seeming unavailability of some kind of bias, good surveys identify possible sources and furthermore warn about the generalizability of their findings to the entire target population. This does not discredit the value of their results, but only suggests the need for further exploration of the views of the likely underrepresented members of the target group.

2.11. Checking validity of responses

55. With professional survey firms, this is less a problem, but it is always well to ensure 1) that interviewers are contacting the designated respondents and that 2) answers are internally consistent. Many survey firms (at considerable additional cost) now give their interviewers laptops or tablets, not only to record responses but also to ensure the interviewers are where they are supposed to be.²¹ This addresses a concern originating in some of the earliest surveys on topics aside from justice, in which interviewers falsified answers, sometimes in bulk, just to meet their quotas.

56. Consistency of answers refers to preventing illogical responses – for example, where a respondent evaluates the quality of services s/he says were not received.²² This can be tracked when responses are entered into a database, but when laptops, tablets or other electronic systems are used to record responses can also be checked during the interview through a program incorporated in the electronic questionnaire.

2.12. Sample size

57. This is added separately only because of a common belief that more is better. As the case of the 1936 *Literary Digest* and Gallup surveys demonstrates, a smaller appropriately designed sample can be more accurate than a much larger one. The size of an adequate sample varies little among very large populations and absent other considerations (e.g., need to include adequate numbers of “marginal” or minority groups) is an extremely small proportion of the total. Rather perversely, when the target

²¹ A method used by Vanderbilt’s Latin American Public Opinion Project (LAPOP) available at <https://www.vanderbilt.edu/lapop/> and HILL in its legal needs surveys. However, as one interviewee noted, even GPS tracking can be gamed by errant interviewers.

²² In a rather rudimentary survey of prosecutors done for USAID in one African country, respondents evaluated courses they had not taken. As this was not captured when the survey was applied, the answers had to be eliminated.

population is small (say only a few hundred) an adequately representative sample may include most members.

58. The principal problem with small samples is their inability to capture sufficient representatives of internal sub-groups. A random sample of 2,000 court users can accurately represent their average views. However, if one is interested in the experience of say low-income, single women living in rural areas, the resultant sub-group may be too small to be considered representative. One solution, seemingly adopted by Australia in its various legal needs surveys, is to expand the sample size (over 20,000 in the Law and Justice Foundations 2012 survey) so as to allow this type of internal analysis. Another still requiring a larger sample, but smaller than the Australian choice is to use **booster samples** for the underrepresented groups.

59. A small sample may also collect few members of minority populations, for example those who are disabled, homeless, or belong to indigenous groups. In this situation, simply increasing the sample size or using a booster sample may not be enough, and a further solution is a stratified design with separate sub-samples of these groups. This can be difficult when there is no way of identifying group members within the sampling frame or where its construction already underrepresents or excludes them. Here a separate survey would be needed once a way to capture these groups is developed.

2.13. Analysis and use

60. The general purpose of justice surveys of the various types discussed in Chapter III is to broaden knowledge within the sector institutions about how their services are perceived and experienced by users, potential users, their own staff, and other stakeholders. Legal needs surveys take a different approach, focusing on all citizens' experience of potentially legal disputes, what they do with them, and why they access or not available services, including those offered by government agencies. This knowledge is also important for policy makers within and outside the sector.

61. The survey responses must be entered in a database that facilitates their analysis. This is a task for those conducting the survey, but depending on their expertise in justice, their ability to do further analysis may be limited. This is simply because they may not know what is important to sector institutions. Typically, the initial analysis is fairly simple – a tally of results on answers to specific questions, and where the survey allows differentiation among user and case types, then according to each of these. Results are typically presented in tables or graphs for use by sector institutions. More complex statistical analysis (e.g., regression) can be done to further link demographic characteristics to frequency of specific problems or to map problem linkages and prioritization within different groups. However, it is recommended that any significant results be presented in simplified form in general reports since few readers will be prepared to understand the more mathematically sophisticated versions.

62. The essential issue is how the analysis will be used: to test adequacy of existing practices, identify areas for necessary improvement, or evaluate the results of reforms already made. The choice among, or the inclusion of these uses should already have determined the questions asked and groups

included in the survey but will also shape the analysis and the interpretation of the results. Except for surveys designed with actionable indicators (e.g., the International Consortium for Court Excellence questionnaires on user and employee engagement) most surveys are very good at identifying and even prioritizing problems. However, they do not provide guidance on solutions especially for “big” issues like trust, independence, bias, or even “excessive” delay. This does not make them less valuable but does imply the need for further analysis of other types to investigate both causes and potential remedies.

63. Here focus groups, in-depth interviews, reviews of laws, process mapping, and statistical analysis (assuming a well-developed statistical system) can be useful. For example, delay, one of the more easily explained big issues, has a variety of potential causes and consequently of solutions; both also vary by country. Delay may be facilitated by something as simple as a failure to monitor timelines, but even where this is done can be complicated by unrealistic deadlines, overly complex procedural requirements, procedural abuse, inefficient programming of hearings, judicial cherry picking (focusing on simple cases and leaving the complex ones behind), poor notification systems, excessive use of panels of judges to decide simple cases, employee (and party) absenteeism, petty corruption (staff “losing” case files), and even equipment failures. Moreover, it may only affect certain cases or parties or have different explanations for each type.

64. This example may appear to argue in favor of surveys with “actionable indicators,” but there are caveats. Principal among them is their frequent reliance on a closed set of problems and solutions and their consequent exclusion of issues and remedies that may have a greater impact on performance. Perhaps employee engagement or the number of steps required to enforce a contract do explain performance quality, but conceivably far less than political interference, limited access, and corruption. Thus, a big picture survey might be a better first step, with those focusing on more actionable issues coming only once the larger problems have been identified.

65. Once a survey has been done, some dissemination of results is in order, if only because it will be difficult if not impossible to deny its existence. The extent of dissemination is a function of the survey objectives, but broader dissemination is encouraged. It may involve different formats for different audiences, taking into consideration what the targeted institutions want to disclose and to whom. Regardless of local preferences, an initial analytic report will have to be prepared for the leaders of the target institutions, laying out all the findings and conclusions. The more critical decisions involve the reports for other institutional members and stakeholders and for the broader public. Even when funded by external actors, the process may stop with the first report, as happened to HILL with one of its legal needs surveys in MENA.

66. When the results are published, even for limited dissemination, it is important that they include a section on methodology. Absence of a good description of the methodology (not only the questionnaires, but also sample size, how selected, response rate and so on) does not mean the survey was of poor quality, but unless included it can raise questions that could undermine its results. Admittedly, most readers will skip this section. However, in the surveys reviewed for this report, there were several that lacked this section raising question about their validity.

CHAPTER III: REVIEW OF SURVEY TYPES USED IN THE JUSTICE SECTOR

67. Although subject to the same requirements as those in any other area, the types of surveys used in justice are somewhat peculiar to the sector. User surveys are common everywhere, but in justice as in other sectors present their own specific problems as to selection of and contact with respondents. In part thanks to donors, and especially the World Bank, the other survey types – of system actors, multi-stakeholder (mixed), and of unmet needs – represent an innovative approach often not seen in other sectors, and not even among justice sectors in more developed countries, where the emphasis is most commonly on the user or single groups of internal actors.

A note on the implications of the Covid and Post Covid environment

The pandemic clearly accelerated the adoption of e-justice techniques, meaning that actions requiring presence in the courthouses are being reduced, if not to an equal extent in all countries. This makes certain survey techniques like courthouse door interviews less practical and may curtail other face-to-face approaches at least over the short run. It may imply a greater reliance on phone and e-mail surveys, with all their attendant problems especially in countries with low levels of connectivity and where capturing marginalized groups thus becomes more difficult.

Justice institutions, donors, and other international organizations that had ongoing or planned surveys when the pandemic hit may have suspended or terminated them. If not, whatever adjustments they made could provide lessons on how to proceed in the immediate future. Countries need feedback on how their e-justice innovations worked and the extent to which they excluded marginalized groups from any service. Statistical analysis, for those who can do it, offers one way of tracking results, but the quality of statistical data in many countries (and its near absence in others) may not allow the necessary finetuning. If not a boon to surveys the pandemic should offer incentives to improve justice sector record keeping and for donors to put more emphasis here. Statistics cannot replace surveys, but in many countries they could be improved considerably while a good, well analyzed statistical database can also provide ideas as to where surveys can shed most light on operations.

3.1. User Surveys

68. User surveys constitute an important and often first attempt to assess the adequacy of services offered by specific sector institutions beyond what their own statistics can tell them. For this reason, they are the survey types most often recommended to justice institutions (IFCE 2020, CEPEJ 2010). Client/user²³ most often refers to individuals or businesses that for some reason need to access justice services. However, these surveys may also include private attorneys, and encompass additional “users,” such as witnesses and members of the public seeking information or attending public hearings or (see below and World Bank, 2014) or be part of **multi-stakeholder surveys** that incorporate justice system

²³ Client and user are utilized without any intended distinction. Neither one adequately captures case participants like expert or lay witnesses although “user” more frequently is intended to incorporate them as well.

actors. Still, most self-financed versions, unless done at the courthouse door, focus on parties to court proceedings. When surveys draw their samples from registries of all citizens or all businesses (World Bank 2019) they will capture and may eventually exclude the non-users,²⁴ but this method does allow an estimate of frequency of use. Within the countries that use them systematically, user surveys are frequently conducted by legal aid agencies. These agencies have the advantage that they, unlike courts, usually have sufficient information on all users to construct a sampling frame.²⁵

69. Justice institutions everywhere have been late in coming to a recognition of the importance of client evaluations, frequently assuming the “client does not know best.” This assumption too often discourages citizens from accessing sector institutions because of their various shortcomings in treating all but the most sophisticated clients. Once justice system actors recognize that they are not “serving the law,” but rather using the law in their service of citizens, they may, as many have already, see surveys as a means to identify how to improve their mandate. As mentioned above, developed common law countries seem to have taken the lead here, but as noted by CEPEJ (2020; 93), “each year a larger number of [COE] member states and entities have in place mechanisms to assess the perception of court users of the service delivered by the judicial system.” This is the first CEPEJ report to track these surveys, and it notes that “user” is frequently expanded to include lawyers, court staff, and judges.” Among the 37 countries with some survey mechanism, slightly more (31) included lawyers as opposed to the 30 including parties. Unfortunately, neither the report nor CEPEJ’s interactive database documents how many surveys were conducted, when and how often, or whether they combined several categories of users or had separate surveys for each one. Judging by these numbers and ENCJ’s series of surveys conducted across the EU (see next section), sector professionals may still value their colleagues’ opinions more than those of lay users.

70. When limited to parties,²⁶ user surveys may focus on all clients of agency services, or specialized groups – either by case (e.g., commercial, family, debt collection, tenancy) or user type, for example only firms (World Bank, 2019). Sampling frames for businesses are usually easier to construct given the likely availability of business registries. For individual users, this can be an issue, thus explaining the variety of imperfect approaches to identifying them – at the courthouse door, from a national citizen registry, by neighborhood location, or even (the snowball technique) asking lawyers for names of clients.

71. There are numerous model user surveys available online.²⁷ Most are relatively short (a common recommendation) and couched in language easily understood by respondents without legal training.

²⁴ However, the World Bank multi-stakeholder surveys do use answers from the general population, employing a simplified set of questions to be asked of all respondents. This allows a comparison of those without court experience with those who have used the courts.

²⁵ This is a record keeping issue that should not be taken as a court problem. Legal aid agencies register users of their services because this is their primary data. Court records usually record the name of the plaintiff or defendant but when done manually or in a simple CMIS, not in a format easily converted to a sampling frame.

²⁶ This remains the more frequent approach in common law countries to this day.

²⁷ See CEPEJ (2010a); Center for Court Innovation (2020) with links to surveys in the US; Legal Services Corporation, all citations but only for legal aid; World Bank (2014) on multi-user surveys in Serbia with a separate module for parties; and World Bank (2019) for businesses.

Topics vary, but they typically focus on the users' experience and satisfaction with agency performance. Thus, the IFCE (2020) calls these user satisfaction surveys. These surveys may be administered by any of the means listed above. However, because of costs, those funded by the sector institutions themselves tend to rely on "courthouse-door" interviews, questionnaires left to be filled out by visitors to the sponsoring organization, emailed questionnaires, and phone interviews (Hamlyn et al., 2015 for the UK; New Zealand Ministry of Justice, 2019). When donors finance surveys in less developed countries, they often go top-of-the-line, hiring companies specialized in survey implementation, and using longer questionnaires and face-to-face interviews. Face-to-face surveys are inevitably more expensive and justice budgets may not allow their use, even once and especially not repeatedly. Moreover, as noted above such face-to-face applications may be less feasible until countries emerge from the restrictions of the Covid era.

72. Among the common law countries that adopted user surveys earlier, many have conducted them repeatedly, using quasi-probabilistic sampling. In the US this is nearly always at the state level and uses courthouse-door in-person interviewing, online questionnaires, or phone contacts, the latter two especially for legal aid agencies. Australian examples include both federal and state surveys, most commonly with phone and/or courthouse-door interviews. Repetition of surveys often includes modifications to questionnaires (addition of questions or slight changes to the wording of others). However, given the interest in tracking performance over time and the results of any recent changes in practices, they typically leave much as is.

73. Unlike other methods for assessing performance (and especially legal and statistical analysis) the surveys' principal purpose is not comparison across systems/countries, but rather an evaluation of progress over time in a single one. A single survey provides a good baseline as well as ideas about what could be improved and what seems to satisfy users. Repetition allows an assessment of whether recommended changes have occurred and with what results, as well as any emerging issues. Countries or courts that can afford this often attempt repetitions every few years. Where, as in the current era, budgetary constraints may not permit this frequency, some non-survey approaches can still provide important information.

74. Where user surveys are repeated, the example of the Seattle Municipal Court in Washington state (US) is especially interesting for having done this three times (2011, 2015 and 2020) and in the most recent iteration, with an accompanying program of focus groups.²⁸ The 2020 survey combined courthouse-door with online interviews, allowing a comparison of the results. Although the methods targeted different populations, with online interviews sent to those paying traffic fines online, the comparisons still revealed some important differences as to who responded. Unfortunately, response rates were not given, but it was noted that online respondents had markedly higher education levels and

²⁸ CEPEJ (2020) notes that Slovenia has also programmed biennial surveys from 2013 onward, using the broader definition of user (thus a mixed survey) and that these are complemented with workshops, in-depth interviews, and observation in the courts.

incomes.²⁹ Although the survey preceded the Covid crisis, one conclusion from comparing the two groups was that more online facilities (beyond paying fines) were needed.

75. The Seattle focus groups (The Vida Agency, 2020) did not alter the general areas of criticism and positive responses. They did provide, as focus groups usually do, more details on what participants liked, believed was needed to improve user experience, and why they objected to certain practices. Criticism was most common and ranged from time and expense of getting to court to (for criminal cases) legalese as a barrier to understanding their case. As jury members were also included as users, there were important recommendations here that would probably not appear in an ordinary survey. Focus groups, like suggestion boxes, community meetings, and similar means to elicit feedback, cannot be taken as representative of larger populations, but well conducted can provide important ideas as to potential changes in “business as usual” that rarely emerge in a survey.³⁰ Two nearly opposite issues with their use are that 1) when participants do not fear repercussions, they can be much more negative than in a survey, but 2) when they do not have this trust or are chosen for this purpose, they will be far less critical. The Seattle Court made the results available online (while protecting the anonymity of the participants), but it is less usual to publish focus group results than those of surveys, if only because of their less than representative nature.

76. Where “user” includes more than ordinary citizens and/or businesses (the parties to a case), it is generally recommended that separate questions be developed for each type, tailored to their likely experience and knowledge. This is especially important where questionnaires are sent by email or “distributed at the courthouse door.” It is suspected, but not proven that undifferentiated multi-user questionnaires can reduce response rates because of their length (even when the portions for each user are short). They can also lead to confusion in providing answers to sections not intended for the specific respondent. Face-to-face and telephone applications can address these issues, but without an interviewer’s intervention, the problems are likely to remain.

3.2. Surveys of system actors

77. Surveys of system actors (e.g., judges, prosecutors, legal aid attorneys, their staff, bailiffs) are generally easier to conduct than user surveys (smaller target group, more likely to have a complete list of members for the sampling frame). However, they are often used only internally, without wider dissemination. The contents and purpose of these surveys vary considerably as suggested by the three examples below.

78. That recommended by the International Consortium for Court Excellence (2020) focuses on “employee engagement.” It features a short (20 question) format that can be answered in 10 minutes.

²⁹ Since there is no reason to believe more educated people with higher incomes commit most traffic infractions, one conclusion is that online responses are easier for them. Or alternatively, online payment may be preferred by higher income groups who can afford not to contest the fines.

³⁰ One issue is that absent many open-ended questions, responses on improvements will follow a predeveloped list, thus eliminating ideas that did not occur to the questionnaire’s authors.

According to the Consortium, it has been applied in 100 US state courts. As the title suggests the emphasis is on how court employees feel about their jobs. It does not include questions about overall court performance of the type typically addressed to court users. Respondents include both judges and other court staff. The purpose, and suggested use, is to identify “trouble spots” or “bright spots” that can be “easily translated into improvement actions,” as well as to track trends and changes over time. Questions are largely actionable – for example, “In the last month, I was recognized and praised for doing a good job” – although possibly not equally relevant in all cultural settings.

79. The surveys conducted by the European Network of Councils of the Judiciary (ENCJ, 2019) take a different tack, focusing on judicial independence and accountability and so omitting questions about other aspects of court performance.³¹ As of 2019, ENCJ had conducted three judicial surveys among EU member states (plus Bosnia-Herzegovina in 2019). It uses a common format (translated into the local language/s) and asks the national judicial councils, ministries of justice and other governance bodies to provide this information to all professional judges (and in a few cases prosecutors) in the country. In all, 11,335 judges participated. Curiously, response rates varied enormously, from less than 5 percent in Romania to over 60 percent in Norway. The average was roughly 20 percent. Possible reasons for the differences were not given. While the questionnaire was not included in the report, the sensitive nature of some of the questions (e.g. whether the specific judge had felt pressured in making a decision) may have dissuaded some respondents or significantly reduced government interest in promoting the exercise. A recent ENCJ survey of lawyers, co-sponsored with the Council of Bars and Legal Societies of Europe (CCBEJ) focused on the same themes. It had lower responses rates and in the end, only 4,250 participants. The responses had not been analyzed by the time the 2019 report was published, but one early conclusion is that lawyers rated the independence of judges lower than did the judges themselves. Since the results are publicly available, a principal purpose and intended use is for participating judiciaries and governments to compare their scores and motivate them to make changes to raise them.

80. Finally, there are surveys of sector actors incorporated in the multi-stakeholder approach developed by the World Bank (2014); here questions posed to system actors replicate many of those concerning performance asked of lay users, including but not limited to independence, fairness, and corruption. These surveys and the logic behind this replication are discussed in greater detail in the next section on multi-stakeholder analysis.

81. As noted by IPSOS (2010), system actor surveys of any of these types are one area where face-to-face interviews may not be advisable. This is largely because respondents may fear the guarantee of anonymity will not be realized. Since this observation came only in Serbia, it merits testing elsewhere, but another World Bank survey (not cited here) also warned against having self-administered surveys conducted in the presence of a supervisor. Supervisor presence is good for a high response rate, but not ideal for frank answers as in this case, the supervisor dictated the responses. USAID and the World Bank

³¹ Although ENCJ has not yet done a user survey, it seems likely that it will share the focus of those done with judges and lawyers. For comparative purposes this makes sense, but it may also indicate something about the valuation of user opinions on other themes.

faced similar issues with focus groups and surveys in two African countries where it proved impossible to remove the supervisor from the room. The International Consortium (2020) also cautions about fears of non-anonymity in its employee engagement surveys because in smaller courts it is easy to identify respondents. The solution offered, when this was done across a court system, was to merge responses of smaller and larger courts to protect those in the former.

3.3. Multi-Stakeholder Surveys

82. Multi-stakeholder or mixed surveys (including individuals and businesses with and without court experience, system service providers like judges and court staff, and intermediary groups like attorneys) are larger and more complicated. The World Bank has promoted and published these surveys although so far largely in Eastern Europe (IPSOS 2010, World Bank, 2014). If funding is available, they offer several advantages over a series of separate exercises. One is cost – the larger the samples, the more expensive the survey, but there are some economies of scale when multiple surveys are conducted under one contract. The greatest advantage is, however, the ability to compare responses from different groups – private lawyers, public defenders, parties to cases and other service users, judges, prosecutors, and so on. The differences are sometimes large, but often in surprising directions, especially on issues like delay, independence, fairness, and ease of access. In one survey (Serbia in 2019), judges’ estimates of citizens’ awareness of legal aid availability were far greater than what citizens themselves acknowledged. As in the ENCJ surveys of lawyers and judges regarding judicial independence, the Bank-sponsored surveys also show significant differences between the two groups.

83. Using multi-stakeholder surveys to compare responses does require overlap in survey questions. If length of case processing or decision quality is a question, it should be asked of all respondents in the same fashion, if a comparison of the results is wanted. World Bank (2014) offers an example of two multi-user surveys conducted in Serbia. The report’s promised methodological annex was seemingly not published, but the implementer, IPSOS (2010,) does provide a good description in a separate report (unfortunately, only for the first survey, which excluded judges, and is not publicly available).

84. Constructing samples for multi-stakeholder surveys is also a complicated undertaking, and one more reason to use advisors with experience in the methodology as well as established survey firms. When it comes to users, the World Bank approach has been to use an initial sample drawn from a registry of all citizens, and then to use a booster sample to select those with court experience. Additional samples may be needed to capture minority groups (commonly in Eastern Europe, Roma populations). Separate samples for businesses are drawn from business registries and for system actors from registries of court employees, lawyers, and any other professionals targeted. In Eastern Europe, this works because most of these registries exist and are reasonably complete. In less well documented countries, this will likely be more difficult, as even registries of all lawyers or judges may simply not exist. Still, these impediments will affect surveys of more limited scope (only users, only judges) in these same countries and so are not unique to a multi-stakeholder approach.

85. Even if only done once (and with World Bank support some countries have repeated the approach), the multi-stakeholder survey provides an excellent baseline against which to measure system performance from the perspective of both the users and actors. It is particularly important for countries that have never done their own surveys before, consolidating into one dataset the perspectives of all the relevant actors involved in the justice system. This allows a comparison of views and so provides rich, and theretofore absent information on where each sees problems and strengths, as well as the differences among them. Actionability can be an issue, however; unlike the simple International Consortium surveys of court users and employees, multi-stakeholder surveys often identify large problems for which the remedies are not obvious. If a court employee believes s/he receives insufficient information on her/his work or a user feels the explanation of a judgment is overly technical, the potential remedy is more evident than if stakeholders agree that judges are biased or insufficiently independent in their decisions.

3.4. Legal Needs Surveys

86. Legal needs surveys have been conducted since the mid-1990s, but their numbers and the countries employing them have burgeoned in the last few years. The two principal overviews of this experience (Pleasant et al. 2013 and OECD/Open Justice Foundation, 2019) thus document an increase from 20 to 55 national surveys over the ensuing period. Had they considered subnational applications the number would be considerably more. As no one has counted the number of user or system actor surveys it is impossible to say whether they are losing ground to this newer approach. As legal needs surveys also include court users, asking some questions about their experience, it is possible they are seen as a substitute or useful variant. Also, many justice systems have additional ways of tapping user reactions and employee perceptions, but legal needs surveys are unique in documenting the needs and experience of those who do not or cannot use courts.³²

87. Although legal needs surveys existed before its publication, Hazel Genn’s 1999 book, *Pathways to Justice*, clearly gave a boost to this approach, as well as introducing an increasingly popular methodology. This entails surveying all citizens, rather than, as was often done in the US, only low-income individuals who are potential users of legal aid. It also asks about the problems/disputes/justiciable issues (and the term used does influence the answers³³) experienced by respondents, what and whether they did anything about them, and whether the results were satisfactory. Thus, it is not only about lawyers, courts, and legal aid, but also addresses alternative ways of acquiring information about the issues and the means to address them.

³² The World Bank multi-stakeholder surveys do incorporate citizens not using the courts, but focus on perceptions of independence and integrity, rather than what “legal” problems they experienced and whether and how they resolved them. The questionnaire for businesses does ask about alternative means of solution but this is not included in the questionnaires for citizen users and non-users.

³³ On this see Pleasant et al., who discuss this issue in great detail, possibly excessively for those who simply want to conduct a survey.

88. While typically aimed at actual and potential system users, these surveys are especially important to assess where sector institutions should place more emphasis to reach the “under or unserved.” Legal needs surveys have been done most frequently by legal aid agencies in the US and by courts and foundations in Australia. England and Wales have done the most repetitions, although Australia, New Zealand, and the Netherlands have also conducted a series of legal needs surveys. Still, HIL arguable has the longest experience using this approach in less developed countries, financed by donors or in some cases by the government itself.

89. By definition, unmet legal needs are “justiciable” issues (potentially legal cases) for which an individual has found no solution. This definition of legal needs, met or unmet, has various interpretations, in turn giving rise to vast differences among countries as to what percentage of the population has met and especially unmet needs. This has made cross-country comparisons difficult as have many additional technical issues ranging from sampling frames through mode of application (Pleasant et al. 2013; Law and Justice Foundation, 2012). Moreover, respondents’ willingness to tolerate what might be considered legal needs or even to classify them as such seems to vary among countries. This could be cultural but may also be attributable to survey design and application. (Pleasant, 2013; 5-6).

90. While comparison of results among several countries is of interest to researchers, this is not the principal issue for the country where a survey is done; here the key questions are which needs are most common, among which groups, whether they are resolved, if so how, and the reasons for the use or non-use of formal sector institutions. As access to justice has emerged as a major concern (in part because of the Sustainable Development Goals and especially SDG 16’s subgoal of promoting “the rule of law.....and ensur[ing] equal access to justice for all”), these surveys constitute an important means of measuring it. It bears mention that if the principal issue is finding a satisfactory resolution, this does not necessarily mean via the courts. This also relates to an occasional criticism of these surveys – that they over-legalize everything. That is to say that a work-related problem might have legal implications but is just as likely to originate in educational deficiencies, economic downturns, and technological innovations within the country. For example, in Tumaco, a Colombian port city, stevedores who used to unload cargo have lost their jobs to technological innovations. The port no longer needs them, but their demands for retraining for another job have no clear target – the companies, the port, the government? (USAID, 2015). They are unlikely to win their case in court should they take that route, and the few accommodations made by the port to keep some on cannot fix the larger problem, hardly restricted to Tumaco’s stevedores. Such issues arguably require a government policy to forecast future human resource needs and train workers for the next generation of jobs. In short, where these surveys convert everything to a legal issue, they can call attention to the problem, but getting people to court or just providing them with legal assistance may not be the best way to resolve it.

91. This said, user needs surveys are an important innovation that can help not only justice institutions but the government writ large to understand the types of problems experienced by ordinary citizens. In some cases, they could motivate legal change or that in justice sector operations to address unmet needs. In others, they should be a heads up to governments about how their laws and policies

exacerbate their citizens' most pressing problems. Unfortunately, except for a few surveys conducted in more advanced countries, there is scant information on follow-up, and this should be addressed after these surveys are conducted.³⁴ And even more than with multi-stakeholder surveys, many of the issues uncovered will simply exceed the potential for immediate, simple remedies.

3.5. Use of non-justice specific surveys

92. Whether or not justice institutions do their own surveys (or have them done by donors or other outside groups) they also typically have relevant information available from surveys done for other purposes. These include household surveys, often conducted by their national governments,³⁵ as well as international surveys like those done by Transparency International and its local chapters, the various regional barometers (Latinobarómetro, Arab barometer and so on), and surveys, like the Global Competitiveness Index, aimed at businesses. Most include few questions on justice, the majority of which focus on corruption, trust, and, if less frequently, quality of performance. They are thus limited in the information provided, but what is there should be considered as important and as a clue to where services might be rethought or improved. There are also indices that use surveys or quasi-survey approaches, like the now discontinued Doing Business, and while it is focused on justice, the World Justice Project.

93. Again, there is scant information on how governments and justice institutions react to these surveys or whether they use the results. Most commonly (and universally) they pay attention only when their scores improve in which case this becomes a media event. What happens when they decline is unknown, except for those in Doing Business, which had the advantage of linking scores to actionable indicators. This inspired some countries to work directly on the elements measured by Doing Business – for example, reducing the number of procedures required to enforce a contract. Whether this improves performance except in these specific areas and the specific cases³⁶ used by Doing Business is unclear. Still, a survey that incorporates actionable indicators is more likely to produce positive reactions than one that simply documents a problem.

94. It should concern a country and justice sector that it has a low and possibly declining score on corruption or trust; still the explanations behind these trends can be complicated when even within the reach of the justice sector. And unfortunately, this is typically not the case. Moreover, a low or declining score, especially on corruption, may only feed a government's inclination to further interfere with sector

³⁴ Both Pleasance et al. and OECD/UNDP ask about impacts, but for whatever reasons this seems restricted to English speaking and thus common law countries. In both, references to specific policy and procedural changes, as opposed to simple familiarity with the surveys' contents and implications were vague. This may be because of the selection of interviewees, few of whom would have been responsible for changing government practices (as opposed to those within a single NGO).

³⁵ OECD/OSF (2019; 28) contains a very short list of government household surveys with justice relevant questions. In fact, there are many more, both in the 6 countries/surveys listed and in others.

³⁶ To improve comparability of its results, Doing Business used a single case type to measure performance. This can hardly be called a "typical case" as for example in enforcement of contracts, the amounts involved were far higher than those usually claimed. As numerous Bank studies (and those of others) have found, most contract enforcement cases are really recovery of amounts owed, and these are typically toward the lower end of the spectrum in countries where courts are used to recover debt.

institutions – as demonstrated by countries as diverse as Ecuador, Peru, and several Eastern European governments. Although low-scoring judiciaries often ignore these results or dismiss them as the views of uninformed citizens, they might take these lessons of experience to heart. As a Brazilian Constitutional Court President once said to his judges, these views may be based on disinformation and citizens' ignorance, but they are still our problem to resolve.³⁷ This caution should also cause sector institutions to consider sponsoring their own surveys, appropriately designed and implemented of course. This might counter those to which they object, but also allow them to explore the reasons for what they consider to be citizen ignorance and misinformation. External surveys are unlikely to stop and thus merit both attention and a closer examination of the problems they reveal.

3.6. Selection among survey types, timing of repetitions, and other strategic issues

95. For countries already accustomed to using surveys, these questions are less important; for those without this experience they are critical. A first step, however, is to recognize that any justice system needs to evaluate its performance and that feedback from users and other stakeholders is an essential part of that process. Performance statistics are important, but they provide only part of the picture. And unfortunately, countries collecting minimal performance data, and analyzing and publishing even less, are also those least likely to use and value survey information. If a country chooses not to evaluate justice (or other) services, the rest of this chapter is irrelevant. For those taking that step, the following applies as it may also for long-time survey users ready to reconsider whether the information received is sufficient to identify areas meriting action. For newcomers, the issue is where to begin.

96. The first and most basic question is what you want to measure, which in turn determines the survey type (although not necessarily, as discussed next, the questions to be asked). Is it how those familiar with the existing system perceive and evaluate it or what they think of recent changes? In both cases, a multi-stakeholder survey or one of its component parts (users, system service providers, intermediaries like lawyers) is the answer. If, as the Paths to Justice approach argues, it is what justiciable problems citizens have and how and whether they resolve them, a justice needs survey is in order. Where both sets of questions are relevant, as they are in many less developed countries, both types of surveys might be done. Much like the distinction between statistical and survey data, the two approaches measure different things, and so are equally valuable. Conceivably with outside support, both survey types could be used. For countries not in this enviable position, a user or multi-stakeholder survey could be more appropriate if only because what it reveals will be more actionable. However, the choice depends on local priorities, where countries feel changes are most needed and/or most possible, and the types of criticisms already voiced by local populations and to some extent, by the results of the international indices and surveys discussed above.

97. The choice of survey questions is not dictated by the survey type, as demonstrated by the variety of survey contents discussed above. User or employee engagement figuring in the International

³⁷ Nelson Jobim, President of Brazil's Supremo Tribunal Federal and National Judicial Council 2004-2006). During a 2004 meeting attended by the author.

Consortium model questionnaires (focusing on how respondents feel they were treated) is very different from the questions on fairness and decisional independence asked of both groups in the multi-stakeholder or the ECJ surveys. Such differences hinge on two predetermined assumptions: the problems identified, possibly through external surveys and indices, and their hypothesized causes. Countries and judiciaries not concerned about judicial independence, whether or not they should be, will exclude questions about it, while those emphasizing employee morale or user friendliness as determinants of performance will emphasize both. What a country or those designing their survey decides to ask is again a local responsibility. The only guidance here is to be careful about assumptions if the survey is to provide useful information on what needs to and can be improved. At the risk of making surveys impossibly and impracticably long, those commissioning/designing surveys might review questions asked in others if only to see whether they have missed something important.

98. And then, assuming one has done a good baseline survey or two (assuming both a user/stakeholders and needs survey are done), the issue is not only to act on it but when and whether to repeat it. Repetition is good, but institutional change is a slow process, and repeating a survey annually or even biennially is likely not worthwhile or can be counterproductive.³⁸ Moreover, absent rapidly ongoing changes either in the system or in the country context, anything less than every five-years makes little sense. Performance statistics, assuming they are adequately reliable, can record more rapid progress or lack thereof in output, but people's perspective inevitably take longer to adjust.

99. Finally, to be of broadest value, any survey should be technically sound, meaning primarily that its design avoids introducing bias from the start. It should not be intended to prove anything but rather to tap feedback, perceptions, and experience from targeted respondents. Factors like survey content (questions asked) and credible guarantees of anonymity are important, but the critical issue is sample design. Samples can be chosen to reflect the desired answers, but why anyone would invest the necessary time and resources in doing this is a good question. Even lay observers can detect this intention fairly easily so the final word is if a survey is to be done, it should be done well. Otherwise, even as a symbolic investment it is money wasted.

3.7. Lessons learned

100. There are some core lessons learned that may be particularly useful as MENA as the practice of justice surveys is only taking root in the region.

101. Placing a value on user feedback requires a sizable cultural change in any public or private service or professional organization: Judiciaries may be among the last professionally-based service providers to pay attention to user feedback, but all public (and many private) agencies accept this change with enormous difficulty. Medical doctor and teachers are only adjusting now in many countries, universally. Eventually, and inevitably, most professional

³⁸ CEPEJ (2020) does highlight Slovenia's use of biennial surveys, but does not indicate why they are done so frequently.

agencies come around, but what is missing is any idea as to how to speed the transition. There are no lessons here from MENA. Still, even without surveys, a few countries seem more open to the concept, less because they value surveys and other feedback but because of an announced commitment to a public or “people” oriented approach to their policies.

- 102. Because surveys per se typically do not provide solutions to problems, countries may struggle with deciding on what to do to address areas of weak performance:** And when solutions are provided they do not necessarily improve overall performance. Here the challenge is that when the “problem” identified is the absence of something (e.g., simpler procedures, training, a specific law) the implicit solution –provide the missing element -- may not be adequate to fix the larger issue (e.g., delay, access, judicial independence). Procedural complexity can indeed slow resolution of cases but reducing steps may not override other causes not identified in the survey (for example, corruption, duplicitous counsel, staff’s inattention to procedural deadlines, political interference). Absent a more nuanced understanding of the causes of poor services, surveys and indices designed to promote a one-size-fits-all remedy can push those interested in improvements in the wrong or at least an incomplete direction.
- 103. A survey can be an important first step in identifying issues, but developing solutions nearly always requires further investigations of other types** – where possible, statistical analysis; where not, case file analysis and process tracking, as well as focus groups, targeted interviews, and on-site observation of staff performance. The answer to the what-works-and-why question is inevitably more analysis. This can discredit the value of a survey, but only if one expects it to do more than its principal objective – identifying the strengths and weaknesses in service delivery as perceived by respondents, a necessary first step in setting reform priorities.
- 104. Depending on how samples and expert witnesses are chosen, surveys and expert opinions can provide very biased and often overly positive views.** This was an issue at the start with Doing Business, which arguably selected the lawyers to answer its questionnaire from an elite group. As one critic noted, these lawyers had probably never handled a bounced check (Doing Business’ first sample case) in their lives. Most real surveys done in MENA (see next chapter) do pass the test of adequate sample design even if the objectivity of the complementary expert (or government) response to questionnaires remains uncertain.
- 105. Just as newcomers to justice services often have unrealistic expectations, countries doing their first survey are likely to be overly optimistic about what it can tell them:** A survey will reveal issues, but as noted above, is unlikely to provide solutions and moreover can be hard to interpret. Interviews with those beginning the process suggest their relative confusion as to the likely results and how they might be used. For example, one agency, used by a small proportion of citizens still expected significant findings on public perceptions of its importance – from a population most of whose members probably had no understanding of what it did.

CHAPTER IV: MENA'S EXPERIENCE WITH SURVEYS AND RECOMMENDATIONS ON THEIR USE

4.1. Introduction

106. Although surveys are a recent addition for justice programs everywhere, in MENA, their adoption has been still more delayed. Few countries in the region have conducted them on their own, and even then, still with donor financing and technical assistance. To these can be added several surveys implemented by donors for their own (often project-related) purposes and the inclusion of selected MENA countries in survey-based indices (e.g., World Justice Project, Global Economic Index, Arab Barometer) with sections on justice. Still in total, the examples are few, making for a very short chapter but many recommendations on the road ahead.

107. The explanations for this short and so far, limited history are multiple. For many MENA countries, costs are an impediment, although hardly the only or even principal one. Other explanatory factors include a judicial culture and practices in justice institutions in many parts of the world, including MENA, which have also been slow to adopt surveys. While surveys have been practiced in high-income common law countries for longer, OECD countries with a civil law tradition only followed over the last 15 years. Historically, justice institutions consider themselves as exercising authority rather than delivering services. The global tendency has been that a shift towards the latter is generally required for an institutional culture to embrace surveys. At the same time, there are now important examples of justice surveys from the MENA region.

4.2. Examples from MENA

Self-initiated surveys

108. Few MENA countries have conducted their own surveys of justice programs. There are a handful of exceptions here described as country-initiated surveys. Conceivably the field research has overlooked a few more, possibly because they were never publicized as per accompanying limitations on transparency.³⁹

109. “Self-Initiated” may be too generous a term as all have benefited from external assistance and in most cases funding. However, those included are distinguished from the next two categories in that government authorities were actively involved in planning, overseeing or doing implementation, and disseminating the results. Unlike most donor-initiated surveys (the next category) the self-initiated are

³⁹ For example, a case file analysis financed by the UNDP in 2017 in cooperation with the Jordanian Ministry of Justice was not released and as noted below, the legal needs study done by HIIIL for four UAE emirates was only published in its summary form.

not linked to a specific donor project/program and often were intended as input to a sector reform program. Those identified include one survey in Jordan; at least one and possibly more in Morocco on commercial justice; a survey in connection with a 2013 Tunisian national consultation on justice reform; a needs assessment done by HILL but requested and funded by four UAE members; online surveys conducted by Abu Dhabi and other emirates soliciting user reactions to online services; and a multi-stakeholder survey recently initiated by Egypt's State Council (administrative justice). Except for the online surveys in the UAE all follow recommended survey protocols, with so far as could be determined, adequate sampling frames, sample selection, testing of questions and so on.

Jordan

110. Jordan's 2011 survey was initially funded by the World Bank, benefited from World Bank input to its design and analysis, and from its inception involved participation from the government, a principal local NGO (Justice Center for Legal Aid, JCLA), and the World Bank. The parties agreed to its implementation by the government's Department of Statistics and so guarantee governmental involvement in dissemination as well. Plans to repeat the survey later were never realized.

111. The 2011 survey is the largest covered in the "self-initiated" category tapping views from 10,000 households in both rural and urban areas (Prettitore, 2013, 2014 a and b). Its primary objectives were to identify the most common types of legal disputes, linking them to the characteristics of the households and individuals in the sample and to tap respondents' perceptions of the quality of sector institutions and services. The survey was designed in cooperation with the World Bank, which also providing financing, later supplemented with funds from the Ministry of Justice. Although informants registered some doubts as to the government's initial commitment, the survey did build on a series of government initiatives including the 2006 "We are All Jordan" strategy promoting social justice and the Justice Upgrading Strategy (JUST, 2010-2012), which included the enhancement of access to justice as part of its plan. The survey's impact on subsequent reforms has not been documented. Reportedly it influenced the 2017 alteration to the rules on automatic eligibility for legal assistance adding it for criminal defendants facing up to 10 years of imprisonment if found guilty.

Morocco

112. So far as could be determined Morocco has conducted no general user or stakeholder surveys on justice. However, its Planning Commission does periodic surveys of private enterprises and at least the most recent (Morocco, 2019) included questions on court use and satisfaction with the service provided. In total 2101 enterprises were included in the stratified, random sample. The sub-samples were designed to separate enterprises by size and areas of activity (industry, commerce, construction, and non-financial services). Questionnaires were done face-to-face; 80 interviewers and 10 supervisors were contracted for this purpose.

113. The survey found that among private enterprises, court use was the most frequent means of resolving disputes, representing 70 percent as opposed to mediation, arbitration, and other methods.

Micro enterprises were more likely than large companies to report difficulties with the courts (22.7 percent as opposed to 6.9 percent) but were less bothered by delays (42 percent as opposed to 68.7 percent). Whether this was a one-off inclusion of questions on justice is not known; nor are the use to which the findings may have been put.

Tunisia

114. Although Euromed (2012) reported a user survey in Tunisia, the most important experience was the survey conducted in 2013 as input to a proposed reform of its justice system. The few details provided on the earlier experience suggest it was less a survey than the “citizen supervisors” review of issues they found in various courts. After its 2011 revolution, Tunisia organized a National Consultation on the reform of Tunisia’s Justice [system].⁴⁰ The Consultation was done in partnership with (and probably with financing from) the UNDP and the (UN) Office of the High Commissioner on Human Rights (OHCHR). It was implemented by a Tunisian research entity (Elka Consulting) with survey experience. To what extent this was a purely Tunisian initiative is unclear, but the Ministry did require it in preparation for a strategic reform plan.

115. The Consultation was conducted in 2013 (from April to September) and was done in three stages. In order, these comprised regional meetings to produce a preliminary assessment of the existing system; a series of in-depth interviews and focus groups with system actors to develop themes introduced in the regional meetings, and a national survey covering justice actors, court users, and the general public. The justice actors were contacted by telephone, but the 1,248 respondents from the public were contacted directly with face-to-face interviews. Although the final report does include a section on methodology, the more detailed annexes including questionnaires and information on the selection of respondents were not available online.

116. The Consultation did produce improvements in sector organization and operations, but these came from the first two stages and hardly incorporated the pages of suggestions listed in the final report. The most important changes, if implemented gradually, were the creation of regional administrative tribunals, and increases in the number of the lowest level ordinary courts. Both measures were adopted to expand access to justice services. The opinion polls did not ask for recommendations but focused on issues like trust in justice, staff competence, and corruption. The responses provided a dire picture, although one largely shaped by operations under the prior Ben Ali administration. The survey indicated that 59.2 percent of the respondents believed that the judiciary was politicized, and that 54.1 percent thought it is biased and partisan. These are high numbers but still lower than the 68 percent saying it favored certain groups. Questions about the performance of various justice actors featured corruption as among the three most important characteristics, with it coming first in several cases.

⁴⁰ Republic of Tunisia (2013) with information on the methodology and findings.

United Arab Emirates (UAE)

117. Of the two examples here, the user needs survey, conducted by HIL and covering four emirates (Abu Dhabi, Dubai, Sharjah, and Ajman) is the most methodologically sound. The online surveys of online service users demonstrate an interest in feedback, but hardly a systematic effort at random sampling. These efforts to capture feedback are a good sign so whether this is an example of a survey is less important.

118. HIL conducted the legal needs survey in 2016, following its normal protocols (discussed in Chapter 3 above). The survey was completed, and an initial summary of results published (HIL 2016), but the Emirates (or at least one of them) did not authorize publication of the complete report. HIL does not disclose reasons for this. The summary appeared fairly innocuous, but significantly, this is the first time this has happened with any of HIL's exercises (although it was apparently also the first time the country studied contracted the work).

Egypt

119. Egypt had at least two donor-initiated and funded surveys previously. However, so far as could be determined, that recently begun by the State Council is the first to be initiated by the judiciary. The Council is the separate judicial body handling administrative disputes as well as providing advice to government on proposed legislation and serving as legal counsel to various ministries, governors, and other public sector actors. The survey design is similar to the World Bank multi-stakeholder surveys. To design and conduct the survey, the State Council contracted a local firm (Baseera) with extensive survey experience, although none in justice.

120. As of March 2022, this survey had begun its first segment, interviewing court staff. Later steps will address the views of judges, lawyers, court users, and the general public. Sub-sample sizes vary as do the means for their selection and for applying the questionnaire – ranging from face-to-face, to email to phone interviews. The 3,000 citizens to be surveyed will be interviewed by phone and selected through randomized digital dialing. Although the selection will clearly exclude anyone without a phone, or phone access, for a survey focused on stakeholder perceptions of and experience with an entity specializing in administrative disputes (including a larger number involving public sector employees) this is not the problem it would be for a survey with another purpose – for example identifying the common justiciable issues of the entire adult population. Moving forward, this will be an important experience to follow.

Donor-initiated and Funded Justice Surveys

121. Several countries which, whatever their intrinsic interest in surveys, may lack funds or an immediate reason to conduct one have benefitted from donor programs to fill that gap. The examples found (and there may be more) were Egypt, West Bank and Gaza, Morocco, and the various countries with HIL's legal needs studies financed by a donor within or outside the latter's programs there. Although

it can be assumed that all these surveys required an agreement from the targeted country it is apparent that the impetus came from the donor and was often linked to the donor's in-country projects or related interests. This means that here unlike the surveys described in the prior section, even more than funding and assistance, donors also defined survey content to answer questions they considered important, either linked to their in-country activities or to broader concerns. There is no inherent problem with these developments except that they may reduce the likelihood that countries will act on the results and repeat the surveys out of their own initiative.

USAID in Egypt

122. USAID has financed at least two surveys in Egypt, most recently on its project with the economic courts. USAID (2019) provides the questionnaire and results for the survey begun in 2018 along with a brief description of methodology. Respondents were 90 lawyers and businesspeople who used the courts, 30 each in Alexandria, Cairo, and Asyut. No methodological information could be found so it is impossible to determine whether these were probability or other types of samples. As USAID has conducted other projects in Egypt, including with family courts and personal status (PSL) issues, it may have done other project-related surveys, which the authors of this report were unable to locate. Such surveys are typically part of USAID projects and evaluations worldwide, are always shared with the host country, but inasmuch as they are project focused, may be of less concern to the latter.

UNDP in West Bank and Gaza

123. Over the period from 2011 to 2015, the UNDP conducted three surveys⁴¹ in West Bank and Gaza under its SAWASYA Project (the UNDP/UN Women joint program, "Strengthening the Rule of Law: Justice and Security of the Palestinian People"). All surveys used similar questionnaires and methodologies and aimed at capturing respondents' perceptions of the quality of services offered and received, as well as the level of contact with the targeted justice and security institutions. All three were face-to-face household surveys with relatively large samples (8,000 households in 2015 of which 6,823 agreed to answer the questionnaire). All three reports included discussions of survey methodology, noting that the samples were randomly selected through a multi-stage cluster technique. Thus, for the 2015 survey, both the 320 primary (geographic) sampling units and the 25 households within each were selected at random. The final stage was to select one male or female member over 18 in each household as the respondent. The selection process here was not explained except that there was an effort to maintain a gender balance. All surveys included both urban and rural areas and "contained 16 strata representing all the governorates in the west Bank and Gaza strip." Refugee camps were also included.

124. The 2011 survey showed relatively little citizen contact with sector agencies within the prior 12 months and very negative perceptions of services, much of which the UNDP believed were based on earlier experiences. Although not designed as "user surveys" all three did capture the views of those with

⁴¹ The surveys were done in 2011, 2012, and 2015. The repetition of the 2011 survey one year later is not a recommended practice, but the 3-year gap between the second and third exercises is reasonable.

contact, either during the target period or earlier. Over the four years contact increased although only to 28.7 percent of the households sampled while perceptions appeared to have improved. By 2015, levels of trust in the justice and security systems rose from 2.79 to 3.19 on a 5-point scale. Also, although women were less likely than men to take their disputes to court, the “gender gap” in satisfaction with service appeared to have been reversed, with women now more satisfied than men.

125. According to its 2015 report, the UNDP intended to continue the surveys, but did not specify the intervals. However, it did plan to include households from the 2015 survey so as to track changes in their perceptions. Unfortunately, there was no information on any further surveys even from knowledgeable sources within West Bank and Gaza. A separate UNDP criminal case file analysis (not a survey) was conducted but never released.

126. Although not specifically mentioned in the final document (State of Palestine 2018), it is likely the surveys influenced the work of the National Legal Aid Commission. The Commission’s National Legal Aid Strategy was completed in 2018 after several years of work and was to be followed by an implementation plan and a draft legal aid law. Unfortunately, neither has been done, and the proposed national legal aid system thus remains as a vision, if a very detailed one. This work occurred under the auspices of SAWASYA II and was funded by the Governments of the Netherlands, Sweden, the Spanish Agency for International Development Cooperation, UNICEF, UN Women, and the UNDP.

World Bank Trust Survey in Morocco

127. As of writing this survey had just concluded, but the final analysis and report were still being done. Consequently, the short description here relies on exchanges with the survey team and some initial documents they provided. The Trust Survey is another Bank tool intended for use in various countries. It measures trust in several public sector institutions, against a series of socio-demographic and experiential variables, the latter including the respondents’ satisfaction with “process,” their treatment, and the outcomes or results. One interesting preliminary finding is that trust was more closely related to satisfaction with process and treatment than with the service outcomes.

128. The survey was offered by the Bank to the Moroccan authorities, who accepted the proposal and moreover, have requested more detailed analysis of several sectors, including justice.⁴² The Government’s interest stemmed in part from its “New Development Model,” in which trust and the social contract are emphasized. The Bank supervised the survey for which a local firm was hired. The firm’s database of phone numbers was used as a sample frame, allowing randomized selection of the over 6,000 respondents with stratification (by geographic location). Since the final analysis has not concluded (and that for justice will take still more time), the Moroccans’ use of the results is unknown for now, but the

⁴² The Bank is working on a more detailed analysis of the responses to the justice questions, as requested by the Moroccan authorities. The initial presentation showed the justice institutions’ relative ranking on trust, but much more could be done with the database to identify the demographic and experiential factors influencing the score.

themes' connection to the new model is promising. This also suggests that even surveys initiated by donors for their own purposes may match local priorities thus augmenting the likelihood of their use.

Various HIL Legal Needs Surveys

129. By now HIL has conducted these surveys in roughly twenty countries; except for that contracted by the UAE, they are typically financed by a donor, usually the Department of Foreign Affairs of the Dutch government. This was the case for those done in Jordan, Lebanon, Morocco, Tunisia, and Yemen. It is worth noting that Jordan's 2011 survey covered much of the same ground, with a larger sample than that used by HIL (which also, as is its practice, combined a non-random selection of districts with randomized sampling within them).

130. HIL's reports on its surveys do not have information on how they were received by the targeted countries, and whether any of their findings influenced sector policies. Interviews in Morocco and Tunisia in conjunction with earlier (2020) World Bank diagnostic work indicated that many sector actors were unaware that the surveys had been done. Other donors working in these countries do seem aware, and selectively use the findings in their work. The surveys contribute a general knowledge base on legal issues in MENA and other regions.

131. It is to be seen how exactly the Dutch government, the usual funder, intends to use the quantity of information these surveys provide. For HIL, the surveys do support its current argument about the need for "people-centered" justice. The findings, nearly everywhere, suggest a limited use of formal institutions, and perhaps more importantly, that the poor have problems the formal system is not designed to address.

Inclusion of MENA countries in international survey-based indices

132. The most important are World Justice Project (WJP), Arab Barometer, Transparency International (TI), the Global Economic Index, and before it was dissolved, the World Bank Group's Doing Business (really not a survey but rather a poll of expert opinions). Only WJP focuses solely and most comprehensively on justice, but it only covers 8 countries (Algeria, Egypt, Iran, Jordan, Lebanon, Morocco, Tunisia, and the United Arab Emirates). Both Transparency International and the Arab Barometer cover 13 countries, but slightly different ones, and Transparency includes Sudan in its list. Neither index focuses on justice, although both occasionally add questions on corruption/trust in judicial institutions. The Global Competitiveness Index, which has the broadest coverage, includes justice (trust, competence) in only two or three of the many questions in its Executive Opinion Survey,⁴³ and the now dissolved Doing Business focused only on a specific type of commercial case, considering only procedures and not "exogenous" factors (like corruption and political interference) that could easily undermine the best designed process. Doing Business has rankings, but as it is not survey-based is only discussed here because of the lessons it provides about policy impacts.

⁴³ See https://www3.weforum.org/docs/WEF_GCR_2019_Appendix_B.pdf for a description of contents and analysis.

133. Since only the Global Index can be conducted out of country, the exclusion of more countries from the others can be explained variously by: government rejection of the proposed survey, country conditions (e.g. civil war) making doing a survey impossible; failure to find a suitable local survey enterprise, and for WJP, the difficulty of adjusting its survey methodology to very small countries.⁴⁴

Since none explained the reasons for the exclusions, it can only be guessed at what they are. In the two survey systems using rankings (WJP and TI⁴⁵) it is interesting that the UAE (limited only to Abu Dhabi, Dubai, and Sharjah) scored highest, followed by Qatar (not in WJP) and then Jordan and Tunisia, but that for the rest, the order varied considerably in the two indices (and not only for exclusions). Part of that variation are because that while WJP looks at 8 dimensions of justice (with corruption as only one), TI exclusively covers corruption in the entire public and private sectors.

134. WJP and TI are arguably the two most important indices for MENA that use surveys (although not exclusively) to rank justice (and for TI, other public sector) institutions. The Arab Barometer, which is entirely survey-based, in the last two years eliminated the few questions on trust in justice institutions, while the Global Competitiveness Index (like Doing Business) only looks at what presumably matters to the enterprises it surveys. Except for WJP, none covers justice as more than a part of its questions and coverage of MENA countries is never complete, especially for the Arab Barometer. Scores and, where used, rankings vary considerably, by year and index, but none suggested improvements were not needed. Except for the now defunct Doing Business, it is not obvious that the results had any impact on country or sector policies. Still, this is not unusual and nearly universally the most frequent question when poor results come in is not “how can we do better,” but rather “what can we do to raise the score?”

135. Good scores, or those indicating improvement often were publicized by the press, and if less, frequently, poor or lowered ratings as well. As with the HILL studies, donors working in the countries do use the findings to support the need for their projects, or at times, their projects’ contributions to any improvements. The only exception to these generalizations is the now defunct Doing Business, whose ranking system featured actionable indicators, meaning that a few countries took the obvious measures to improve their positions – for example, lowering the number of steps or time required to enforce a contract. These were often positive changes. However, given the factors not covered by Doing Business (e.g., corruption, political interference) whether they, as intended, improved the business environment remains an unanswered question. Also, even if business investors profited from the changes, the impact on other users was likely minimal.

Use and impacts of surveys – Were there any?

136. These are difficult questions given the limited transparency among the region’s justice institutions. It is known that among the self-initiated examples, the 2011 Jordanian survey and that conducted in Tunisia in 2013 did produce some positive changes – although those in Tunisia did not come from the survey so much as from the accompanying regional meetings and focus groups. Moreover,

⁴⁴ In its 2021 report, it mentioned this as a difficulty for adding small Caribbean countries.

⁴⁵ Doing Business has rankings, but as it is not survey-based is only discussed here because of the lessons it provides about policy impacts.

whether locally or donor initiated, both were intended to provide ideas on how to improve justice performance. For Morocco's enterprise surveys and the two examples from the UAE it is hard to say whether the exercises motivated any change.

137. Except for Abu Dhabi's online "survey," the non-survey elements in the Tunisia consultation, and Doing Business' actionable indicators, none of these examples aimed at identifying possible improvements, but rather at tapping perceptions and identifying issues. And Abu Dhabi's online questionnaire was really not a survey, but rather a direct attempt at getting feedback on a new system. The problem as discussed earlier, is that the jump from an identified issue to its solution can be vast. Moreover, an effective remedy depends on an understanding of causes, something a survey itself rarely supplies. If for example, the overriding perception is that justice is too slow, nothing much can be done without knowing:

- Whether other data (e.g., statistics, but also from other sources) substantiate this belief, or suggest it is the result of unrealistic expectations.
- Where delays originate – in complex procedures, lawyers' tactics (to string out a case and so increase payments or "buy time" for their client), in the overburdening of judges or other court staff, or simply in a failure to monitor timelines.
- In which cases and for which clients, delay is most egregious, and for those, for which reasons.

138. Similar questions apply to just about any perceived problem and thus require further exploration if a suitable remedy is to be developed. Moreover, as the UNDP emphasized in its first survey in West Bank and Gaza, where people have infrequent contacts with sector agencies, their opinions may well be based on experiences long in the past (or, not mentioned by the UNDP, very recent, well publicized events that do not affect them directly). Much the same also applies to international surveys, the type that rank countries on one or more aspects of performance. Unless specifically designed with "actionable indicators" (and here there are other issues) they only illustrate the problem without defining how it could be resolved.

139. A final issue, and one never mentioned by the surveys' authors, is that answers to surveys and especially to expert questionnaires (used as a complementary addition by most indices and the basis for the Doing Business scores) can be influenced by respondents' fear of repercussions. This should be a consideration for a few countries in any region, including MENA, and may well explain some exclusions – countries where index authors realized all respondents would fear the consequences of any perceived criticism. This does not explain countries' willingness or unwillingness to use the results but can cast into doubt the quality of the latter. Adding to that doubt, while the indices provide fairly detailed explanations of their survey sampling techniques, the selection of expert contributors remains a black box.

140. Still, for a judiciary or other sector agency that is concerned about improving service quality and user (or non-user) perceptions and experience, surveys can provide the motivation to search for remedies even if, of themselves, they cannot identify them. Assuming that they will provide remedies may represent unrealistic expectations on the part of those commissioning or agreeing to the survey. This is most likely in the case of a first experience with surveys, when recipients are not prepared for what they

will receive or how to use it. Where donors are involved, they can help by explaining the nature of the results, their use, and the ways to explore possible solutions for the issues identified.

141. Investments in justice surveys are meaningful. The international ones in which MENA nations figure serve the purposes of their authors, have their own sources of funds, and are useful to donors, investors, international NGOs, research institutes, and individual researchers in deciding where they wish to operate. It is unlikely that the countries included completely ignore the results and over time they may decide they need to address the problems registered. In the meantime, the wide dissemination of these indices is unlikely to escape the attention of all their citizens, some of whom may begin to demand responses. This is bound to be a slow process, given the many additional issues plaguing these countries, but cultural change is rarely rapid.

142. For the few countries taking the initiative or actively participating in exercises proposed by donors, the surveys, if publicized (and one suspects all are not), indicate an interest in public feedback and provide information potentially relevant to any national or sector plans. As with all surveys, the wealth of information gained far exceeds any agency's ability to act on it. So, one or two changes motivated by the contents could justify the investment of time, and often of funds. As for the rest, assuming sufficient dissemination, it could inform/influence thinking about future needs and solutions.

143. Meanwhile, donor-initiated surveys are often required by their organizations and depending on how organized could support continuation of a specific project or indicate where modifications are needed. As they require little if anything from the targeted countries, the only costs are those incurred by donors, and these are calculated as part of their normal operating expenses. Like the external indices, their sponsors/funders/implementers have their own reasons for these exercises, which presumably are satisfied regardless of their impact in the countries they review.

4.3. Recommendations

144. Recommendations for countries/judiciaries/NGOs are divided between those to be introduced "immediately" and "later" with no particular deadline for beginning either. Immediate in any case means start now but may include actions that will only be completed over time, if ever. (Some immediate actions are simply activities that should be repeated continually).

4.4. Recommendations for global and regional action to promote justice surveys

145. Surveys and other diagnostic tools have an image problem – while they are far less expensive than the normal list of reform mechanisms, their purpose and utility are rarely understood. They are often seen as only producing the bad news that few sector actors want to hear. The recommendations listed in the next two sections are addressed to countries interested in such tools as well as donors and others wishing to promote surveys. A few recommended activities for them are this as follows:

Globally raise the profile of feedback/surveys as part of justice improvement

146. Stress the importance of feedback on existing systems to determine where they are falling short of user (and non-user) needs. This can be linked to issues of value to governments, like foreign investment, national reputation, political stability (also of interest to investors) and modernization. A focus on people-centered justice is also useful where governments are willing to embrace it.
147. Emphasize that the modernization inputs (ICT, training, and infrastructure) are only part of what defines a modern judiciary, and that their effective adoption requires empirically based planning and analysis of data from various sources – one of which is surveys. Absent this part of the transition to modernity, many inputs have been found not to produce the promised improvements in services or in user satisfaction.
148. Stress that feedback is also useful in combatting criticism of performance, based on little more than a few “outrageous” cases and “common knowledge.” A good sign from surveys in some countries is that people who used justice services are more positive about them than those who never have.

Regionally

149. Sponsor workshops within the MENA region on the organization and uses of feedback mechanisms, noting the advantages of surveys for this purpose. Invite country authorities from other parts of the world that have used these mechanisms to good effect.
150. Provide funding and/or technical assistance to organize early efforts while ensuring publication (and thus use) of the results.

4.5. Recommendations for Immediate Adoption by Countries Interested in Surveys

Review any surveys already conducted in country or within international indices to capture results and gaps in justice services.

151. **Except for the smallest MENA countries, or those currently experiencing internal conflicts, most have been included in one international index and a few have self-initiated surveys or surveys initiated by others.** These should be a way to focus discussion and identify issues worth further exploration. Ideally, this discussion should be broadly based, to include not only sector actors, but also NGOs and representatives of the wider public. The purpose is not to provide answers but rather to review what have been identified as issues and elicit feedback on their importance and validity.

Organize a workshop or even better a national consultation like that held in Tunisia in 2013 to elicit views from stakeholders and the broader population on the problems they believe should be addressed.

152. **Since this recommendation is directed at countries with an interest in surveys and other types of user feedback, it should elicit some important findings and suggestions.** While its current situation is challenging, Tunisia's 2013 consultation did produce a long list of issues and possible solutions as well as some concrete positive changes. It also revealed how ordinary citizens perceived their justice system. It is an example worth repeating elsewhere assuming a supportive political economy.

Design and implement a first (or follow-up) survey based on the information derived from the first two measures.

153. **This survey should be announced publicly and its results, at least some of them, published and discussed.** If funding or knowhow is an issue, identifying donors that are amenable for assistance and funding is an option. Building on prior discussions means designing the survey to incorporate questions they identified. This does not imply they should be its sole focus, but only that they should not be ignored.

Publish the results and use them to focus discussions on reform priorities and plans

154. **This is the missing step in nearly all prior MENA examples and should not be ignored here.** Low levels of transparency have not been helpful so far. Here the involvement of donors could be important in organizing and possibly financing events, and especially in bringing in representatives of countries that have successfully used surveys to focus reforms. Obviously, a survey is of little use if the step is not taken. Its findings should be part of a broader discussion on short, medium, and long-term plans, revisiting the ideas arising in the initial workshop or consultation. Ideally the discussion should include participation by all stakeholders (including, if present, donor agencies). While they may be complex with so many views represented, they can also identify specific areas for future conversations among the most affected. The results of both the survey and these broader and narrower discussions should be publicized, if not in their entirety then at least in part.

4.6. Recommendations for Later Action

Combine survey results with other performance evaluation data to develop an overall assessment of the state of the justice sector as a basis for longer term reforms

155. **This is reserved for later as many other data sources may still need to be developed.** They include statistical analysis, as well as more extensive interviews, focus groups, process

mapping in addition to a review of any prior evaluative studies. This work could be consigned to an institution with the required skills and a credible objectivity in its performance (e.g., local research institute, sector organization). If undertaken by a sector agency it should go to an internal department with a credible firewall to protect it from repercussions.

Build surveys into justice reform/modernization plans

- 156. Just like statistical analysis, but measuring different things, surveys should be part of any reform program.** They are essential in identifying issues, tracking results of changes, and capturing emerging trends. Absent these inputs (and both surveys and statistics are vital here) a plan is moving blindly. Assuming a good database, statistical analysis can be continual, but surveys should be scheduled periodically, less often than annually (unless narrowly focused on a specific change) but at least every five years to ensure adequate tracking of changes, positive or negative.

Use feedback and user surveys to review treatment of vulnerable groups and so identify areas where common practices negatively affect equitable outcomes.

- 157. Vulnerable groups should be identified and included among survey respondents. They may include refugees, guest workers, women, and minorities.** Their access may be limited not because it is legally denied, but because of other obstacles. Sometimes what authorities believe is a positive change is not perceived as such by the intended beneficiaries – and the only way to know that is to ask them. Here surveys perform a vital function of testing or even predicting the results of reforms and can be complemented by other feedback mechanisms (focus groups, interviews) as well as statistical analysis.

Plan for a series of surveys spaced over time to measure progress, detect new problems, and facilitate discussions with stakeholders and representatives of the public

- 158. One survey provides a baseline or place from which to start.** Over time, perhaps every five years, its repetition (with any necessary additions and modifications) can be used to measure change, find positive developments, and identify emerging or disappearing⁴⁶ trends and issues.

⁴⁶ Often forgotten is the potential for one-time trends disappearing once policies, laws or various exogenous factors disappear. This is another reason for repeating surveys, as well as other analysis, periodically so as to avoid programs addressing issues that no longer exist.

REFERENCES

Al-Zoubi, Muath. 2020. "Legal Aid in Criminal Matters in Jordan." *Journal of Law, Policy, and Globalization*, Vol. 93. Available at <https://pdfs.semanticscholar.org/e5c2/856fecd004428fbe32c28dfe086f6f2b3651.pdf>

Center for Court Innovation. 2020. Can Courts Be More User-Friendly? How Satisfaction Surveys Can Promote Trust and Access to Justice. Available at https://www.courtinnovation.org/sites/default/files/media/document/2020/CCI_FactSheet_Satisfaction_Surveys_04202020.pdf

CEPEJ (European Commission for the Efficiency of Justice). 2010. "Conducting satisfaction surveys of court users in Council of Europe member states." Strasbourg, September 10 available at https://books.google.com/books/about/Handbook_for_Conducting_Satisfaction_Sur.html?id=NKr5mgEACAAJ

CEPEJ (European Commission for the Efficiency of Justice). 2020a. Case Weighting in Judicial Systems. Strasbourg Available at <https://rm.coe.int/study-28-case-weighting-report-en/16809ede97?fbclid=IwAR0-yZpl8tcO5-KtHueBTLGRAMQ9zxSJMMY2zL6HB4KcyvYwaoRV4sCwGus>

CEPEJ (European Commission for the Efficiency of Justice). 2020b. European Judicial Systems. Available at <https://www.coe.int/en/web/cepej>

Check, Joseph and Charles Schutt, 2012. *Research Methods in Education*. Sage.

ENCJ. (European Network of Councils for the Judiciary). 2019. Independence, Accountability and Quality of the Judiciary. Available at <https://pgwrk-websitemedia.s3.eu-west-1.amazonaws.com/production/pwk-web-encj2017-p/2019-06/ENCJ%20IAQ%20report%202018-2019%20adopted%207%20June%202019%20final.pdf>

European Union and Council of Europe. 2018. "Analysis of the Results of the Court Users' and Lawyers' Satisfaction Surveys: Basic Court of Gjakove/Dakovicaa, Basic Court of Prishtine/Pristina, and Basic Court of Prizren." June, Available at <https://rm.coe.int/kosej-analysis-of-the-court-survey-results-eng/16808d3419>

Family Court of Australia, Federal Circuit Court of Australia. 2015. Court User Satisfaction Survey. Available at <http://www.federalcircuitcourt.gov.au/wps/wcm/connect/fccweb/reports-and-publications/reports/2015/>

Genn, Hazel. 1999. *Paths to Justice: What people do and think about going to law*. London: Hart Publishing.

Hamlyn, Becky, Emma Coleman, Susan Purdon, and Mark Sefton. 2015. Civil Court User Survey: Findings from a postal survey of individual claimants and profiling of business Claimants. Ministry of Justice (UK), Analytical Series. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/472483/civil-court-user-survey.pdf

HIIL (Hague Institute for Innovation in Law). 2020. "Justice Needs and Satisfaction in Ethiopia." Available at <https://www.hiil.org/projects/justice-needs-and-satisfaction-survey-in-ethiopia/>

HIIL (Hague Institute for Innovation in Law). 2019a. "Justice Needs and Satisfaction in Fiji." Available at <https://www.hiil.org/wp-content/uploads/2018/07/HiiL-Fiji-JNS-report-web.pdf>

HIIL (Hague Institute for Innovation in Law). 2019b. "Justice Needs and Satisfaction in Morocco." The Hague, Netherlands.

HIIL (Hague Institute for Innovation in Law). 2018. "Justice Needs of Syrian Refugees: Legal Problems in Daily Life." Available at <https://www.hiil.org/wp-content/uploads/2018/09/Justice-Needs-of-Syrian-refugees.pdf>.

HIIL (Hague Institute for Innovation in Law), 2017a. "Justice Needs and Satisfaction in Jordan." The Hague, Netherlands. Available at https://www.hiil.org/wp-content/uploads/2018/07/JNS-Jordan-2017_EN-Online.pdf

HIIL (Hague Institute for Innovation in Law), 2017b. "Justice Needs and Satisfaction in Lebanon." The Hague, Netherlands. Available at <https://www.hiil.org/projects/justice-needs-and-satisfaction-in-lebanon/>

HIIL (Hague Institute for Innovation in Law), 2017c. "Justice Needs in Tunisia." The Hague, Netherlands. Available at <https://www.hiil.org/wp-content/uploads/2018/07/HiiL-Tunisia-JNST-English-web.pdf>

HIIL (Hague Institute for Innovation in Law). 2016. "Justice Needs in the United Arab Emirates: Preliminary Findings." Available at <https://www.hiil.org/wp-content/uploads/2018/07/Justice-Needs-in-UAE-preliminary-findings.pdf>

HIIL (Hague Institute for Innovation in Law). 2014. "Justice Needs of Yemenis: From Problems to Fairness". Available at <https://www.hiil.org/projects/justice-needs-in-yemen-from-problems-to-fairness/>.

HIIL (Hague Institute for Innovation in Law). 2012. "Rule of Law in Yemen: Prospects and Challenges." Available at <https://www.hiil.org/wp-content/uploads/2018/09/Rule-of-Law-in-Yemen.pdf>.

IFCE (International Framework for Court Excellence). 2020. "Global Measures of Court Performance (Third addition). Sydney Australia. Available at "
https://www.courtexcellence.com/__data/assets/pdf_file/0015/53124/The-International-Framework-3E-2020-V2.pdf

IPSOS. 2010. "Collection of Baseline Information on Court and Prosecutorial Performance in the Republic of Serbia." Report prepared for the World Bank, on file with author.

Katz, Diane S., Nathaniel, and Larry Hembroff. 2008. "Understanding Public Opinion Surveys," February. Available at <https://www.mackinac.org/9262>.

Law and Justice Foundation. 2012, "Legal Australia-Wide Survey Legal Need in Australia. Available at <http://www.lawfoundation.net.au/ljf/app/D5FF73DC95E64EA9CA257B5F00168DF3.html>

LCS (Legal Services Corporation, US). N.d. "Cleveland Legal Aid's Use of Outcomes." Available at <https://www.lsc.gov/grants-grantee-resources/civil-legal-outcomes/case-studies/cleveland-legal-aids-use-outcomes>.

LCS (Legal Services Corporation, US). N.d. "Model Practices and Innovations." Available at <https://www.lsc.gov/grants-grantee-resources/model-practices-innovations>

LCS (Legal Services Corporation, US). N.d "Comprehensive Need Assessment and Priority Setting." Available at <https://www.lsc.gov/grants-grantee-resources/resources-topic-type/comprehensive-needs-assessment-priority-setting>.

Morocco, Morocco. Haut Commissariat au Plan. 2019. Enquête Nationale auprès des Entreprises. Available online at https://www.hcp.ma/Enquete-nationale-aupres-des-entreprises-2019_a2405.html

New Zealand Ministry of Justice. 2019. Court User Survey 2019. Conducted by Colmar Brunton. Available at <https://www.justice.govt.nz/assets/Documents/Publications/2019-Court-User-Survey-Report.pdf>

OECD)/Open Society Foundations. 2019. "Legal Needs Surveys and Access to Justice." Available at <https://www.oecd-ilibrary.org/docserver/g2g9a36c-en.pdf?expires=1612807011&id=id&accname=guest&checksum=D22F45C9DD3CC3D7F5C8A90379AC2AE9>

Prettitore, Paul. 2018. "Can Justice Make Poor Women Less Vulnerable." Available at <https://www.brookings.edu/blog/future-development/2018/02/21/can-justice-make-poor-women-less-vulnerable/>

Prettitore, Paul. 2014a. "Building Legal Aid Services from the Ground Up: Learning from Pilot Initiatives in Jordan." World Bank: MENA Knowledge and Learning. Available at <https://openknowledge.worldbank.org/handle/10986/20554>

Prettitore, Paul. 2014b. "Targeting Justice Sector Services to Promote Equity and Inclusion for the Poor in Jordan," in Cissé, Hassane, N. R. Madhava Menon, Marie-Claire Cordonier Segger, and Vincent O. Nmehielle, eds. *The World Bank Legal Review*, Volume 5. Fostering Development through Opportunity, Inclusion, and Equity: 245-262.

Prettitore, Paul. 2013. "Justice Sector Services and the Poor in Jordan: Determining Needs and Priorities." World Bank: MENA Knowledge and Learning. Available at <https://openknowledge.worldbank.org/handle/10986/16120>

Pleasance, Pascoe, Nigel J. Balmer, and Rebecca L. Sandefur. 2013. *Paths to Justice: A past, present and future roadmap*. London: UCL Centre for Empirical Legal Studies. Available at https://www.researchgate.net/publication/271209897_Paths_to_Justice_A_Past_Present_and_Future_Roadmap.

Republic of Tunisia. 2013. *Consultation nationale sur la réforme de la Justice en Tunisie*. December. Available at <https://www.justice.gov.tn/index.php?id=335&L=3>.

Seattle Municipal Court. 2020. *Access and Fairness Survey Results*. Available at <https://www.seattle.gov/Documents/Departments/Court/Community%20Engagement/2020%20SMC%20Access%20Fairness%20Survey%20Results.pdf>

State of Palestine. 2018. "National Legal Aid Strategy: 2019-2022." On file with author.

Transparency International. 2019. "Global Corruption Barometer Middle East and North Africa: Citizens' Views and Experiences of Corruption." Available at <https://www.transparency.org/en/publications/global-corruption-barometer-middle-east-and-north-africa-2019>

UNDP (United Nations development Program). 2017. "Justice and Security Monitor: A review of Palestinian Justice and Security Sector Data 2011-2016." Available at <https://www.pcbs.gov.ps/Downloads/book2382.pdf>

UNDP (United Nations Development Program). 2015. "Perceptions of Palestinian Justice and Security Institutions in 2015." Available at https://www.ps.undp.org/content/papp/en/home/library/democratic_governance/public-perceptions-of-palestinian-justice-and-security-instituti0.html

UNDP (United Nations Development Program). 2012. “Public Perceptions of Palestinian Justice and Security Institutions.” Available at www.ps.undp.org/publications/docs

USAID (United States Agency for International Development). 2020. Quality of Services Provided by Kosovo Basic Courts – as Evaluated by Lawyers. Available at <https://dplus.org/wp-content/uploads/2020/05/03-Report-Lawyers-ENG-11.pdf>

USAID (United States Agency for International Development). 2019. “Rule of Law Assessment of Egypt’s Economic Courts.” Available at https://pdf.usaid.gov/pdf_docs/PA00TSBQ.pdf.

USAID (United States Agency for International Development). 2013. *Assessment and Impact Evaluation of Colombia Justice House Program*. Prepared for USAID/Colombia in coordination with Partners/Colombia. Available at <https://govtribe.com/opportunity/federal-contract-opportunity/access-to-justice-activity-aja-sol51412000001>.

[The] Vida Agency. 2020. “Focus Group Findings and Recommendations.” Available at [Seattle.gov/Documents/Departments/Court/Community%20Engagement/Improving%20Equity,%20Fairness,%20and%20Accessibility%20at%20the%20Seattle%20Municipal%20Court_Sept2020_The%20Vida%20Agency%20\(1\).pdf](http://Seattle.gov/Documents/Departments/Court/Community%20Engagement/Improving%20Equity,%20Fairness,%20and%20Accessibility%20at%20the%20Seattle%20Municipal%20Court_Sept2020_The%20Vida%20Agency%20(1).pdf)

World Bank Group. 2020a. Improving Commercial Justice in BiH in the Face of Covid-19 Crisis: Guidance for Commercial Case Processing. September. In draft, available with author.

World Bank Group. 2020b. Justice Surveys. Power Pt and recorded presentation. Available at <https://www.worldbank.org/en/events/2020/02/11/justice-surveys-why-to-do-them-and-how-to-do-them#1>

World Bank Group. 2019. “Improving Commercial Justice in Bosnia and Herzegovina: Baseline Survey on Perception and Experience on Access to Justice in Bosnia and Herzegovina for Micro, Small and Medium Sized Enterprises.” June.

World Bank Group. 2014. Serbia Judicial Functional Review. Available at <https://openknowledge.worldbank.org/handle/10986/21531?show=full&locale-attribute=es>

World Bank Group. *Doing Business*. 2002 onwards. Various reports available at <https://www.doingbusiness.org/en/doingbusiness>