The authority of international courts and tribunals had already been examined through a variety of lenses during BIIJ 2016, from using a somewhat abstract sociological model to considering what judges themselves can do to improve their everyday functioning as well as bolster the coherence of the legal system in which their institutions operate. Participants now turned to a critical question that is inevitably linked, in the minds of many, with how authoritative international courts and tribunals are perceived to be—how well do they perform? This session was co-led by Richard Goldstone and Christine Van den Wyngaert, Judge of the ICC and former Judge of the ICTY. It used as a launching point a recent “performance self-evaluation” conducted by the ICC.

Van den Wyngaert began her introductory remarks by noting that assessing the performance of international courts and tribunals has become a topical issue in recent years. A crucial question at the outset of any such evaluation is how to define the concept of performance itself. Evaluating an international court’s performance cannot simply be a matter of calculating profit margins, as with private companies, or even conviction rates and workloads, as with national judicial systems. If it is “hard to put a price on justice,” then this applies a fortiori to international justice, she declared. This is why courts have been trying to develop “performance indicators” which will allow them to measure their performance. At the ICC, all the organs of the Court (judges, prosecutors, members of the registry) recently engaged in a court-wide exercise to develop such indicators. Van den Wyngaert observed that this was a fascinating exercise, and one which proved more difficult than expected at the outset. They learned that it is indeed difficult to measure the work of a court.

It was noted that there has been a lot of recent scholarly work on the performance of international courts and tribunals, and different models have been used to analyze their performance. The nature of these models, however, is such that only those aspects of the court’s work that are quantifiable are capable of being assessed in a meaningful way. The most common models for analysis were then discussed and critiqued.

- The compliance rate approach looks at the degree of compliance by parties with court orders made against them. By this token, the higher the compliance rate, the more successful the international court or tribunal. It is one of the few variables that is capable of being quantified. The problem with this model is that compliance may relate to extra-legal factors and may have nothing to do with the quality of the ruling. For criminal courts, a further problem is that they rely almost completely on the cooperation of states at all possible levels: execution of arrest warrants, evidence gathering, the conduct of investigations, and the enforcement of sentences. The ICC has faced the problem of non-compliance, notably in the failure by states to arrest and surrender President Al Bashir from Sudan, as already discussed in Session 2 by Richard Goldstone. Reliance on signatory states for the functioning of the Court therefore attaches too high a price to cooperation and
non-cooperation and may be too narrow a test for measuring the ICC’s performance. The compliance rates approach therefore does not seem to be suitable for all international courts.

b) The usage rates approach refers to the extent of the court’s workload. Applying this model to international courts may be very different from one court to another. For example, the ICJ was hardly used during the Cold War years. At the celebration of the 70th anniversary of the ICJ on 20 April 2016, many speakers praised the fact that, in recent years, the workload had considerably increased, which had made the court more relevant and important as the principal judicial organ of the UN. By comparison, the usage model as applied to the ICC begs the question: usage by whom? This is in view of the triggering mechanisms under the Statute which can be States, the Security Council, but also the Prosecutor acting proprio motu. At the ICC, which is based on the complementarity principle, less “usage” of the court may mean that complementarity is working and that investigations and prosecutions are effectively taking place on the national level. For the ICC, in an ideal world, it should have no work at all. Like the compliance model, the usage model does not seem to be an adequate model for measuring the performance of all international courts and tribunals.

c) The goal-based approach looks at whether and how a court effectively fulfils the mandate set out in its charter. This model was originally applied to private companies and NGOs to measure their effectiveness and is now proposed by one of the leading authors in the field, Yuval Shany, to measure the performance of international courts and tribunals.1 This model allows for taking into account the fact that several goals may exist simultaneously. In doing so, it allows for a more nuanced assessment of the performance of an institution. Interestingly, and perhaps most problematically for international courts, is the fact that, in contrast to private companies or even national (criminal) justice systems, there may be disagreement about the court’s goals.

Thus, before answering the question whether an international court or tribunal is attaining its goals, a preliminary normative question is what should be these goals? There are many problems with this. Some goals, particularly ultimate goals, may be vague and thus open to interpretation. What Shany defines as “goal ambiguity” may therefore leave too much room for irreconcilable views on the perceived goals of the court. The diverse number of stakeholders in any one international court or tribunal means that entirely dichotomous goals may exist with little chance of these being reconciled. Also, the precise time frame for the attainment of goals is especially difficult to gauge. Of particular importance to infant courts is the extent to which the first years of existence of a court can really be used to gauge levels of performance in the long term in light of inevitable “teething problems.”

There can also be a divergence between performance measurement from an internal perspective (i.e. by actors and stakeholders from inside the institution) and performance measurement from an external perspective (i.e. by actors and stakeholders from outside the institution). The ICC suffers from such divergence. Its most important goals from the inside are ending impunity by increasing accountability of state officials for international crimes, deterrence or prevention of these crimes, ensuring international peace and security, enhancing international cooperation in the prosecution of international crimes, and guaranteeing lasting respect for and the enforcement of international justice. However, the common goals projected on the ICC from the outside may be quite different, coming from states parties, accused persons, victims’ groups, and various NGOs. “It is important to consider the question of acceptance in situation countries,” Van den Wyngaert stressed. The fact that the work of the ICTY did not result in a shared narrative about the Balkan conflict in Serbia, Croatia and Bosnia points, she believes, to a failing of the Tribunal. She contrasted this with the South African Truth and Reconciliation Commission (TRC), which succeeded in establishing a common narrative about that country’s apartheid era.
Van den Wyngaert then noted that the divergence between internal and external goals may be more extreme for the ICC and other criminal institutions than for other types of international and regional jurisdictions. However, she believes that it is important for all international courts and tribunals to consider the external perspective when they attempt to measure their own performance. She also applauded the iCourts’ model discussed in Session 1 for utilizing the external impact of ICs as one of the ways to determine their authority.

Goldstone then opened the discussion period of the session with the following two questions: 1) Should judges be involved in assessing the performance of their respective institutions? If yes, how? 2) What are the performance indicators that are relevant for judges’ respective courts or tribunals? A lively and wide-ranging conversation ensued.

Judges from criminal institutions identified with the tension between internal and external performance indicators experienced at the ICC. One participant observed that criminal tribunals operate in an extremely political environment, and that some prosecutors take positions motivated by extraneous considerations, such as the interests of victims or balance between ethnic groups. He recommended the exercise of caution: “If we play to constituencies, we lose the credibility that the international community is willing to accord us. The mandate of a judge must be to decide guilt beyond a reasonable doubt. If we go beyond that, there is a problem.” He contrasted a criminal tribunal to the South African TRC, which could seek other outcomes as it was not a purely judicial entity.

It was pointed out that criminal tribunals are often evaluated by insiders and parent organizations on the expeditiousness and fairness of their proceedings. A participant declared that the ICTY trials represent the “gold standard” of fairness, even though proceedings were often protracted. Another observed that the participation of victims in ICC trials—a development lauded by external stakeholders—will inevitably slow down the proceedings. “That’s where I see a discrepancy between what an internal and external assessment might be,” he said. In other words, pleasing the victim “audience” may compromise the Court’s performance from the point of view of internal stakeholders. A judge with experience in a hybrid criminal institution expressed his view that judges should organize proceedings for efficiency as they see fit, and then assess their own performance against the standards they set. “As for how their work is assessed externally, that is up to scholars to do.”

Judges in interstate dispute resolution institutions shared some of the views of criminal judges. In speaking of the Appellate Body experience, a participant presented several benchmarking goals: first of all, the “prompt and fair resolution of disputes,” but also “avoidance of disputes” that may arise from disagreements around what constitutes compliance with past rulings. Another external stakeholder issue is access to the Appellate Body by some member states. It might cost upwards of $1 million to have a case adjudicated, so there is now assistance for poorer countries to file disputes.

A member of a regional court opined that “performance improves when there are standards.” Judges should then be intimately connected to assessment procedures and undertake reforms if specific milestones are not met. A colleague from another regional court agreed, adding that such assessment should remain confidential. “You don’t want a document appraising what you do, and then have the temptation to make it more appealing to the outside world.” If individual judges are found to be the source of inefficiencies, then this needs to be worked out internally.

Responding to the assertion that speed of adjudication should be a primary indicator of success, an interstate dispute resolution judge noted that this criterion only goes so far. “Let’s not lose sight of the most important theme, which is the substantive quality of our judgments, the rational outcome of our cases.” He admitted that this benchmark is, however, hard to measure. Another participant pointed out that the way individual judges approach
deliberation is also critical to performance. He asked, “Do they keep an open mind till the end?” A judge countered that in criminal jurisdictions, “there is a clash of judicial philosophy—civil law wants the truth, and common law just wants to know if the Prosecutor proved the case.” The need for judgments to be “expressed in a language the [stakeholder] community can understand” was also raised. And once again, it was suggested that external parties be involved in the assessment of judges’ arguments.

Human rights judges then brought in the perspective of their institutions, which are charged with establishing standards across wide and diverse regions. Sometimes their benches must consider what is “realistic” and whether their rulings can be enforced. Returning to the standard of the speedy resolution of cases, it was pointed out that the ECtHR seems to have sacrificed transparency for expeditiousness. Under pressure from the Council of Europe (COE) to decrease its case backlog, most cases are now resolved by a single judge who produces no opinion laying out a legal argument. This makes an assessment of the Court’s output by external actors more difficult. But in any case, a participant observed, the COE seems more interested in the number of cases won than in the legal rationale behind rulings.

Another human rights judge said of his younger institution, “Number and quality should not be the only yardsticks.” He observed that his court is involved in enhancing the rule of law across its broad jurisdiction. “Think of human rights and justice as two beautiful flowers, which cannot develop in poor soil. Our court is fertilizing the soil.” He urged his colleagues not to overlook how their institutions relate to civil society, national judiciaries, and national human rights commissions when assessing their performance.

Over the course of this discussion on performance, participants also brought up several issues of recurring interest and concern to international judges, issues that have been addressed during past Brandeis Institutes. These included: whether there should be a general code of conduct for international judges; methods for removing a problematic judge from an international bench; how term limits and considerations of post-service employment may affect the independence and performance of judges; and the need to vet nominees for international judicial positions and de-politicize their election/appointment so that the most qualified individuals can join the international bench.

Mikael Rask Madsen noted, as the session ended, that BIIJ participants had “come full circle” by linking the notion of how the external community views international courts with how judges themselves assess their work. Both are critical in understanding the authority of international courts and tribunals, and the reception and ultimate impact of their rulings.

NOTE