

Enhancing Effectiveness through Improved Accountability Assessment

Professor John Kirton, Munk School of Global Affairs, University of Toronto

Paper prepared for a panel on “From Commitment to Compliance: The Challenge of Effectiveness — UN Security Council, IFIs and Gx — Global vs. Regional Organizations” at the second annual Princeton Global Governance Conference on “Rivalry and Partnership: The Struggle for a New Global Governance Leadership,” Princeton University, Princeton, New Jersey, January 14-15, 2011. Version of January 7.

Introduction

Just how effective are global governance institutions? How would one measure and monitor what they do and, on this basis, offer recommendations to improve their performance and accountability in solving the pressing global problems of today’s world?

These questions have long been of central concern to both scholars and practitioners for several reasons. Scholars have charted the extraordinary proliferation of intergovernmental institutions over the past century on the assumption that they make some difference on the otherwise autonomous behaviour of even the most powerful states, but still disagree profoundly about how much, how and why (Kirton 2009; Krasner 1983). Citizens regularly criticize the failure of these institutions to solve pressing global problems and to meet their proudly proclaimed, high-profile commitments. Even where the institutions seem to succeed, there is much dispute about which ones did best and who or what caused and can claim credit for such success. Practitioners confront the cruel choice of where to invest increasingly scarce resources among the ever larger array of institutions available to help meet their goals abroad and at home. Now that the reform of international institutions and their overall architecture has risen to the top of the G20 and global governance agenda along with proliferating proposals for reform, a better understanding of which ones are effective and why is needed to arbitrate the competing claims. And in a world now crowded with international institutions, it is important to identify and share what has worked to enhance effectiveness in one body so that others can benefit as well.

The importance of effectiveness assessment is enhanced by additional concerns. In an era of America’s rising relative vulnerability and declining relative capability, it may increasingly need help from others through international institutions and from effective international institutions themselves to address challenges it can no longer solve largely on its own. In this post-September 11 “after vulnerability” more than “after victory moment,” when the United States has pioneered new summit-level international institutions such as the three-member North American Leaders Summit in 2005, the 17-member Major Economies Meeting/Forum in 2007, the G20 summit in 2008 and the 44-member Nuclear Security Summit in 2009, it is especially important to know how they are effective in meeting American, allied and global goals and needs (Ikenberry 2001; Keohane and Nye 1976). This is especially so at a time when investments in international

institutions, starting with the International Monetary Fund (IMF), have become very expensive, just as the U.S. government has fewer funds at home to spend on anything abroad. If others are to bear a bigger burden, as they appear willing to do, they also want to know how and why the institutions work, including in adjusting the actions of America as the world's most powerful country, in response to others and everyone's views and needs. Moreover, improving the accountability of those institutions is important for their legitimacy, as their many stakeholders want to know not only if they have produced results, but also if they are willing to have their performance inclusively and transparently measured, monitored, constructively criticized and corrected, so the many not at the table can understand and contribute to what they do.

This paper thus asks: In think tanks, universities, governments and international institutions themselves, how has "effectiveness" been conceptualized and operationalized? Are the metrics adequate? How can effectiveness best be assessed and improved through accountability processes from all, and for all, with the greatest stake? It offers three answers. First the effectiveness of international institutions has largely been conceptualized and operationalized as first-order implementation by the member governments with their public, written commitments over the short term or their achievement of authorized expenditures and conditions in specific programs and projects. Second, there is a need to extend effectiveness assessments from textual analysis of their public documents to actors' resulting behaviour, from decisional commitments to direction-setting, principled consensus, from implementation to results, and from individual institutions' actions to their interaction, as it affects effectiveness overall. Third, this requires better measurements of consensus, commitment and compliance, better concepts about the institution's overall "mission accomplished," "human lives improved" and "problem solved," and about better mechanisms, by expanding multilateral organizations' mandates, mobilizing members' legislatures and auditors general, establishing an independent accountability institute, foundation and network, and creating a multi-stakeholder compliance consortium for all the assessors to share, compare and improve and be accountable to their peers for what they do.

The Conceptualization and Operationalization of Effectiveness

In the treatment of an international institution's effectiveness, there is a gap between how this term is conceptualized in the ideal, ideational, cost-free world and how it is actually operationalized in the real world where resources are required to measure and monitor what is done. The concept thus runs the full gamut, from having an institution create a new world that only the author or advocate has imagined, through solving individual problems recognized by very few and achieving optimum outcomes never secured before, to making progress in controlling current problems that many accept as real. However, in moving to operationalization and measurement, the vision constricts considerably, to compliance with the commitment by those that made it or their targets outside, their implementation through commitment-consistent action and their progress toward achieving the specified goals. Even more narrowly, there is a considerable concern with whether particular programs or projects mounted by international organizations achieve their intended and authorized results in the organizationally specified way.

There has been much concern with “regime” effectiveness and even legitimacy, and considerable innovation (Breitmeier 2008). Yet scholarly assessments still concentrate on first-order compliance with concrete commitments encoded in public collective communiqués, especially commitments endowed with the high degrees of precision, obligation and delegation that the “legalization” literature highlights (Abbott et al. 2000). The commitments are treated, in the fashion of “strict constructionists,” rather literally, as they are written. Assessment focuses on the manageably measurable first-order compliance of implementing action by those making the commitment, rather than the extended outcomes of what they do. Here a premium is placed on fighting the good fight rather than winning the battle or the broader war. Moreover, the dominant approach of first-order implementation with public, written commitments tends to be done individually by commitment, communiqué, or institution, in isolation, or at best in a comparative as distinct from an interactive mode. Less reliable evidence and analytical consensus have been obtained about the causes of the compliance that comes by institution, communiqué or country. Attention focuses more on compliance over the shorter term, as longer periods make it more difficult to claim that the result has been caused by the commitment made so long ago.

In formal, heavily legalized international organizations such as the multilateral World Trade Organization (WTO) or regional North American Commission for Environmental Cooperation, one data set and analytical procedure flow from the dispute resolution mechanisms these organizations contain. The record of disputes taken to these bodies, taken up by them and processed and ruled on through these mechanisms, offers some systematic, easily available, well-used evidence of which countries are complying with their commitments under the governing treaties. This approach had the practical advantages of the comprehensibility and credibility coming from its legalization and its close relationship with “enforcement” action for redress or deterrence.

Within informal, plurilateral international institutions without a secretariat such as the G8 and G20, as distinct from those that have small ones such as the Commonwealth and the Asia Pacific Economic Cooperation forum, measurement of effectiveness concentrates more broadly on the overall, ongoing commitments by the highest level political authorities that make them. In the case of the G8, after occasional experimentation, there has emerged a process through which a member-owned Accountability Working Group now assesses members’ performance in meeting their commitments in the domain of development and in specific issues areas such as health contained therein. The most recent report uses a standard framework applied across all issue areas and a multi-year medium-term period for assessment of five to ten years (G8 2010).

Independent assessments of G8 and now G20 compliance have, since the start of serious, non-advocacy assessment in the early 1990s, focused on first-order compliance with individual commitments during the year after they were made and immediately before the leaders meet at their next (G8) summit to make new commitments as they choose. Attention has recently turned to compliance with multi-year commitments, to the causes of compliance, including the compliance “catalysts” the leaders can consciously or otherwise embed in their commitments to increase the compliance that comes, and, much

more tentatively, to the impact of interactive commitments by the same or fellow international institutions.

The state of the art is well reflected in the most recent compliance report of the G20 Research Group and the paper by Marina Larionova based on the work of her team at the State University Higher School of Economics (G20 Research Group 2010; Larionova 2010). The highly efficient approach of content analysis of public communiqués can provide systematic, reliable evidence to measure indicators of the performance functions or dimensions of domestic political management (e.g., communiqué compliments to individual members countries), direction setting (e.g., affirmations of democratic, distributional or embedded “liberal” principles and norms), decision making (e.g., commitments) and the development of global governance (e.g., reference to institutions inside and outside the one issuing the communiqué).

The Adequacy of the Existing Metrics

It is well recognized that the existing metrics and assessment processes that put them into practice are inadequate in many ways. Indeed, there is continuing experimentation, intellectually and operationally, to find a better way. To begin with, the written record that is relied upon heavily is but a partial reflection of what the leaders actually discuss and do at their summits, often with much added by lower-level officials without the leaders themselves ever reading it. Much of what the leaders themselves do, including the most important things, remains the modern equivalent of “secret agreements, secretly arrived at.”

The need to go well beyond textual analysis is especially important on the dimension of delivery or compliance, which is arguably the most important one by which the effectiveness, and hence credibility of an institution, can be assessed. It is not the declared intention to comply or implement or even the reflections on how well they have done in this regard that is critical, but the actual commitment-consistent and thus compliant behaviour of the member states whose leaders have bound them to do or not do particular things.

Assessing such state behaviour directly requires a great investment of resources, to assess the compliance of eight or twenty members with the more than 300 commitments that a single Gx summit can make. Thus far systematic compliance assessments have generated a considerable data base dating back to 1975 that has provided comparable evidence across many years, members, countries and the G8 and G20 institutions, as well as the individual commitments with different combinations of compliance catalysts. In the case of the G20, compliance assessments have been completed for 37, or about 10%, of the 372 commitments the first four summits have made. There is thus an understandable resistance to altering the approach to Gx accountability and effectiveness assessment, lest the legacy advantages of comparability be lost and of the great expense of creating anything new that is similarly robust.

In the case of the G8, its intergovernmental self-assessment with a public reporting of the results has several shortcomings (Guebert and Bracht 2010; Kokotsis 2010). It covers

only the policy field of development, but not that of security, which is now the G8's second core competence in a world where G20 summitry has arrived. Five-year assessments, such as the *Muskoka Accountability Report*, make it difficult to measure or use the results on the timely one-year schedule when G8 leaders and their citizens presumably want to know on the eve of the next summit whether the promises made the last time have been kept. And it is difficult to find the data detailing which member countries have complied or not.

In the case of the G20, the institution's dominant approach has been to ask other international organizations to monitor the compliance of G20 members with their specified G20 commitments. This has expanded since the first summit in 2008. However, the admittedly limited data currently available suggest that this approach has not had the desired compliance-enhancing effect. In the case of trade, where compliance data is available, the request of the 2009 London Summit to the WTO in regard to the G20 commitments on anti-protectionism was followed by a G20 compliance score on these commitments of +0.50, higher than the summit's overall compliance average of +0.42. However, its repetition at the Pittsburgh Summit five months later produced compliance of only +0.10, which was lower than the summit's overall compliance average of +0.33. Similarly, its repetition at Toronto in June 2010 produced trade compliance of only +0.15, which was again lower than the summit's overall compliance average of +0.30.

This suggests that the initial apparent success of this approach at London on trade led G20 leaders to expand their reliance upon this approach at future summit, even as it had diminishing value in the trade field itself at subsequent summits. However, the trade trend may be due to the fact that the anti-protectionist commitments became more ambitious, by adding a "redress" component to the core refusal one. More research is required on the compliance effects in other fields of the expansion of this approach.

In the case of accountability self-assessment through an international organization's dispute settlement mechanisms, the evidence relates only to the subset of obligations that are legalized, largely from the founding treaty or treaties, rather than to the many other and often more important commitments the parties have made, especially at subsequent ministerial meetings. Moreover, there is high selectivity and skewing from the particular cases that chosen by parties to take to and through the mechanism. Finally, the resulting record is only that of non-compliance, rather than compliance, and non-compliance is difficult to assess against all the cases of compliance and the potential, relevant universe of cases overall.

Enhancing Effectiveness Assessment and Its Effects

To enhance effectiveness through improved monitoring and accountability, several steps stand out as requiring immediate action, both by the institutions themselves and by those outside with a critical contribution to make in this domain. These steps involve better measures, better concepts and better mechanisms.

Better Measures

In regard to better measures, the first step is to measure compliance with the direction-setting consensus over principles and norms, inspired by John Ruggie's (1983) initial constructivist insight that shared social purpose was as autonomously important as relative power in shaping a regime's life and work, and the subsequent rise of constructivist scholarship. Here it is feasible to start with those presented in preambles or permeating the texts of collectively agreed "outcome" documents from which the decisional commitments themselves are identified, proceed to identify which ones have been directly produced and bought into being by the leaders themselves, and what their shared meaning of them is. Particularly in the case of "soft law" institutions, it may be that compliance with general principled consensus is higher than that with specific commitments on regime "rules" and that such summit soft law institutions do not follow a linear logic where they begin with doing something first for domestic political management and then progressively doing it for direction setting, then decision making, then delivery and, ultimately, developing global governance as time goes by.

A second step is to spend more time on the commitments themselves, assessing whether they are timely, well targeted, significant and ambitious, so the compliance finding can be weighed according to the varying importance of the commitments themselves. Early proof-of-concept work showed that such measurements were feasible, but it is a highly resource-intensive task. Beyond lies the question of whether the commitments made are coherent with the others made in the same communiqué. In all cases it is useful to discover, based on the preparatory and negotiating history, the "intention of the founders" to assess, *inter alia*, if a country's compliance comes primarily with the commitments it has invented and initiated, or with those from the most powerful member of the club.

A third step is to consider the cumulative and interactive context, not just among commitment in the same communiqué, but also from the same institutions over time and above all by different institutions in the same field. Beyond the existing concerns with regime "nests" and countries "forum shopping" (or forum creation), do institutions at work on the same subject come to the same, different or antithetical consensus and commitments, and how do combinations affect the compliance that comes? Depending on the answers, in a world where the demand for global governance exceeds the supply, having both the G8 and the G20 cover the same subjects could enhance the effectiveness of both.

A fourth step is to explore more carefully the causes of compliance, especially those that rest within the leaders' or practitioners' own control. The recent work on compliance catalysts could be extended both empirically and conceptually, to consider whether the addition of accountability of monitoring mechanisms of different sorts enhances or erodes the compliance that comes. More intra-institutional accountability means a potential diversion of resources from performing the main six performance functions and may mean, as the 2010 G8 Muskoka Summit suggested, that the prospect of more accountability monitoring means a reduced desire to make ambitious commitments in the first place.

Better Concepts

In regard to better concepts, the major thrust is to move from a focus on compliance and implementation to a focus on results.

Here the first step is to conceive, in a measurable way of, of the “mission accomplished” as that mission applies to the international institution overall and the core collective purpose it was created to achieve. Here one needs to go beyond specific decisional commitments to core, foundational principles and norms. In the G8 this mission is to promote among its members and globally the values of “open democracy, individual liberty and social advance.” In the G20 it is to promote financial stability and make globalization work for all.

A second step is improved concepts of outcomes, the results obtained against referents ranging from the intent of the creators to the demand for global governance of key problems, as independently assessed. In some cases standard data sets provide sound, over-time data specific to the core of a commitment that measure changing welfare outcomes in the state of the world. These could potentially be converted and aggregated into an overall measure of “human lives enhanced” as a result of what the international institution does. This would require attention to causality, starting with the performance dimensions themselves to ask if the individual institution were responding to the proper domestic political pressures, discussing the right subjects, setting the right directions and so on. Then one would consider international institutions in interaction with one another and how this interaction shapes the observed changes in the outcomes.

A third step is to ask if problems are solved by institutions, or shifted to a lower level or different track as they are redefined or fade away. It raises the larger question of how well today’s international institutions, largely conceived in the hierarchical, static, slow-moving Westphalian world of the 19th century, match the fluid, fast-moving, networked world of the twenty-first century and the complex adaptive system that generates the problems it confronts.

Better Mechanisms

Of the many improved mechanisms that offer the promise of improved accountability assessment and thus effectiveness, four stand out.

Expand the Mandates of Multilateral Organization Mechanisms

The first improvement is available to and appropriate for those major multilateral organizations with a large, professional staff that already regularly assesses member performance on behaviour related to the members’ current, cumulating, highest level commitments to and at the organization itself. Thus the IMF could add to its Article IV consultations and reports on each member a systematic assessment of how that member is complying with the commitments made at the ministerial level at the semi-annual IMF meetings, including at related meetings such as those of the International Monetary and Finance Committee. The WTO could do the same as part of its trade policy review mechanisms’ periodic assessments of its members, with the largest countries most

consequential for the system and its overall level of compliance and “defection dynamic” being reviewed most frequently.

Assess Member Governments’ Extra-Executive Branches

A second improvement is to have the member countries of the international institution use components of their own governments beyond the core executive branch more systematically assess compliance with the countries’ international commitments, just as they do with the operations of their own government at home. Legislatures have the key role here, especially with the well-established legislative components or companions of many international institutions, such as the United Nations, the North Atlantic Treaty Organization, the Summit of the Americas and the Organization of American States, the G8 since 2002 and the G20 since 2010. A second step would be to mobilize national governments’ auditors general, in general or in the environmental field, to assist with this task.

Establish an Autonomous Accountability Institute, Network and Foundation

The third improvement is to establish an institute, foundation and network for independent, third-party compliance and implementation assessment, by those without a specific stake, as owners of the commitments or advocates for further action on the issues they address. In the case of the G20’s use of the WTO as an external accountability assessor on trade, the WTO’s public results were less specific and more favourable than those of the fully independent Global Trade Alert Project, perhaps because the WTO was understandably bound by its own legally defined definitions and sensitive to the potential costs of publicly naming member countries that did not comply. Moreover, it is unreasonable to expect, or arguably even ask, the WTO to assess the compliance of G20 members with their repeated commitment to get the Doha Development Round done, if only because in the case of such delicate ongoing negotiations, publicly identifying the non-compliance of individual partners could be counterproductive to the WTO secretariat’s more important role in achieving the intended result of completing the Doha negotiations. Finally, an independent institute, backed by a foundation, could mobilize more resources than the secretariats of international organizations or intergovernmental working groups themselves would devote to this task, especially if the institute were to be assisted by a global network of academics and others contributing on an ongoing basis to the work.

Such an exercise, with input from advocacy organizations but led and controlled by those without an advocacy state, could have considerable appeal to the international institutions and governments under scrutiny, if only to offset the well-publicized “findings” and criticism that the advocacy organizations understandably tend to make. Furthermore, timely, independent third-party monitoring and reporting are above all valuable to those individuals who made the commitment and who thus most want to know if those commitments are being kept. In the minds of the citizens, such independent assessments would have the credibility inevitably lacking in the institutions, their member governments’ own individual or second-hand self-assessments, and the assessments of advocacy nongovernmental organizations and private sector players. That is why the institutions and governments under assessment, as well as philanthropists and others, can

be expected to contribute to the endowment that supports such an independent institute's assessment work.

Create a Multistakeholder Accountability Consortium

The fourth step is to create an ongoing multi-stakeholder compliance consortium and conference for all the assessors to regularly share, compare and improve and be accountable to their peers for what they do. This should foster some agreement on best practices, appropriate division of labour and mutual understanding of the different purposes and thus methods that guide different parties' work.

References

- Abbott, Kenneth, Robert Keohane, Andrew Moravcsik, Anne-Marie Slaughter and Duncan Snidal (2000). "The Concept of Legalization." *International Organization* vol. 54 (Summer): 401–420.
- Breitmeier, Helmut (2008). *The Legitimacy of International Regimes* (Farnham: Ashgate).
- G8 (2010). *Muskoka Accountability Report: Assessing Action and Results Against Development-Related Commitments*.
<www.g8.utoronto.ca/summit/2010muskoka/accountability/index.html> (January 2011).
- G20 Research Group and G8 Research Centre at the State University Higher School of Economics (2010). *2010 G20 Toronto Summit Final Compliance Report*.
<www.g20.utoronto.ca/analysis/2010toronto-compliance.html> (January 2011).
- Guebert, Jenilee and Caroline Bracht (2010). "Assessing the Results and Impacts of G8 Accountability Reports to Help Pave the Way Forward." Paper prepared for a Conference on "Partnership for Progress: from the 2010 Muskoka-Toronto Summits to Seoul Summit," Moscow, October 27-28.
- Ikenberry, John (2001). *After Victory: Strategic Restraint and the Rebuilding of Order After Major Wars* (Princeton: Princeton University Press).
- Keohane, Robert and Joseph Nye (1976). *Power and Interdependence* (Boston: Harvard University Press).
- Kirton, John, ed. (2009). *International Organization* (Farnham: Ashgate).
- Kirton, John, Marina Larionova and Paolo Savona, eds. (2010). *Making Global Economic Governance Effective: Hard and Soft Law Institutions in a Crowded World* (Farnham: Ashgate).
- Kokotsis, Ella (2010). "The Muskoka Accountability Report: Assessing the Written Record." Paper prepared for a conference on "Partnership for Progress: from the 2010 Muskoka-Toronto Summits to Seoul Summit," Moscow, October 27-28.
- Krasner, Stephen (1983). "Structural Causes and Regime Consequences: Regimes as Intervening Variables," in Stephen Krasner, ed., *International Regimes* (Ithaca: Cornell University Press).
- Larionova, Marina (2010). "Assessing G8 and G20 Effectiveness in Global Governance So Far." Paper prepared for the Princeton Global Governance Conference on "Rivalry and Partnership: The Struggle for a New Global Governance Leadership,"

Princeton University, Princeton, NJ, January 14-15, 2011. Draft of December 24, 2010.

Ruggie, John (1983). "International Regimes, Transactions and Change: Embedded Liberalism in the Postwar Economic Order," in Stephen Krasner, ed., *International Regimes* (Ithaca: Cornell University Press).