

THE INDICATOR FAD:
HOW QUANTIFIABLE MEASUREMENT CAN WORK
HAND-IN-HAND WITH HUMAN RIGHTS—A
RESPONSE TO SALLY ENGLE MERRY’S
THE SEDUCTIONS OF QUANTIFICATION

JOSÉ-MIGUEL BELLO Y VILLARINO & RAMONA VIJEYARASA*

I.	INTRODUCTION	986
	A. <i>Defining the Concepts: What Are Indicators?</i>	992
	B. <i>Distinguishing Five Levels of Analysis</i>	994
II.	THE ONTOLOGY OF INDICATORS: WHAT THEY ARE AND WHY WE NEED THEM	995
	A. <i>Indicators and Their Western Origins</i>	996
III.	HOW ARE INDICATORS CHOSEN AND DEVELOPED? CONSTRUCTING INDICATORS	998
	A. <i>The Voice of the North and Measuring Violence Against Women</i>	998
	B. <i>Expertise Inertia or Learning from Experience?</i> ...	1002
IV.	HOW TO TRANSFORM INFORMATION INTO QUANTITATIVE MEASUREMENTS: UNIVERSAL INDICATORS AND THE DIFFERENCES LOST IN THE CRACKS	1003
V.	THE PRACTICALITIES OF COLLECTING AND PROCESSING THE DATA—THE EXAMPLE OF HUMAN TRAFFICKING	1007
VI.	WHAT INDICATORS SAY ABOUT COMPLIANCE WITH HUMAN RIGHTS: WHERE THE REAL PROBLEM LIES	1011
VII.	ATTACHING CONSEQUENCES TO INDICATORS	1015
VIII.	CONCLUSION	1018

* José-Miguel Bello y Villarino is a member of the Diplomatic Corps of Spain (on leave as of this writing) and a PhD candidate at the University of Sydney, researching corruption and international law. Ramona Vijayarasa is a Chancellor’s Post-Doctoral Research Fellow at the University of Technology Sydney. Her scholarship focuses on international women’s rights law, migration, trafficking, sexual and reproductive rights, and the role of women leaders in promoting gender-responsive legislation. The views expressed in this article are the authors’ views alone and may not represent any of the entities they are associated with. The authors are grateful to the

I. INTRODUCTION

Indicators, understood broadly to mean any type of organized categories that can be individually assessed—quantitatively or qualitatively—have received escalated attention in the twenty-first century, both as a way of understanding legal and social environments and as a method of presenting information for further evaluation. This heightened public interest in what are most commonly institutionally-created numerical rankings or scores has been paralleled by academic attention, with leading scholars, research centers, and even international organizations, showing great concern for—and devoting significant resources to—the question of what can, or should, be measured and how.¹

Among the social aspects which can be both measured and compared, human rights figure very highly. If there are some inalienable and global rights that are linked to the human condition and dignity, measuring the enjoyment of those rights in different parts of the world should allow us to compare the extent to which different States are fulfilling their human rights obligations and assess both improvements and declines over time.

Some may argue that the core human rights standards were drafted at a time when effective monitoring was neither

editorial staff at the *N.Y.U. Journal of International Law and Politics*, especially Yekaterina Fomitcheva for her useful substantive suggestions. As always, all errors are our own.

1. See generally, e.g., Gauthier de Beco, *Human Rights Indicators for Assessing State Compliance with International Human Rights*, 77 *NORDIC J. INT. LAW* 23 (2008); Maria Green, *What We Talk About When We Talk About Indicators: Current Approaches to Human Rights Measurement*, 23 *HUM. RTS. Q.* 1062 (2001); Angela Hawken & Gerardo L. Munck, *Cross-National Indices with Gender-Differentiated Data: What Do They Measure? How Valid Are They?*, 111 *SOC. INDICATORS RES.* 801 (2013); Johannes P. Jütting et al., *Measuring Gender (In)Equality: The OECD Gender, Institutions and Development Data Base*, 9 *J. HUM. DEV.* 65 (2008); Debra J. Liebowitz & Susanne Zwingel, *Gender Equality Oversimplified: Using CEDAW to Counter the Measurement Obsession*, 16 *INT'L STUD. REV.* 362 (2014); AnnJanette Rosga & Margaret L. Satterthwaite, *The Trust in Indicators: Measuring Human Rights*, 27 *BERKELEY J. INT'L L.* 253 (2009); Irene van Staveren, *To Measure is to Know? A Comparative Analysis of Gender Indices*, 71 *REV. SOC. ECON.* 339 (2013); Boris Branisa et al., *New Measures of Gender Inequality: The Social Institutions and Gender Index (SIGI) and Its Subindices* (Courant Res. Ctr., Discussion Paper No. 10, 2009), <http://EconPapers.repec.org/RePEc:got:gotcr:010>;

intended nor achievable.² Yet, monitoring of global progress towards more equal, more just, and more sustainable societies is a core component of the human rights and development sectors today. As a result, methods for counting progress have proliferated, triggering a substantial debate about what constitutes a comparable, somewhat objective, universal standard against which all countries can be measured.

Developing such a universal standard is not a simple endeavor. While global human rights treaties set out agreed principles for what are meant to be universally accepted human rights norms, interpretation is an evolutionary process, responding to changes in societies. This is reflected in the issuing of General Recommendations and General Comments, which the nine core human rights treaty bodies have the power to do.³

Despite this global norm setting, determining what should be measured in each case—as well as how such measurement should occur—was always likely to stimulate substantial debate. The goal of this process may also be contentious. Do we undertake this exercise for mere comparison purposes, or are we serving the broader goal of monitoring human rights towards their universal enjoyment? Should the outcomes of numerical ranking shape policy or, adopting a more extreme approach, be used as the basis for a system of rewards and punishments in response to a State's performance?

Regardless of the answers to these questions, the span of this debate proves that we live in an “indicator-prone culture,”⁴ which, perhaps, has been accepted without a thorough valida-

2. ANNE F. BAYEFSKY, *THE U.N. HUMAN RIGHTS TREATY SYSTEM: UNIVERSALITY AT THE CROSSROADS* 2 (1st ed. 2001).

3. There are nine treaties, but ten treaty bodies. The Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (Dec. 10, 1984, 1465 U.N.T.S. 85) has its own treaty body, as well as the Subcommittee on Prevention of Torture and other Cruel, Inhuman or Degrading Treatment or Punishment (SPT) established pursuant to the Optional Protocol to the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (Dec. 18, 2010, 42 ILM 26).

4. We prefer this term to “indicator culture,” used by Merry, as we believe that there is an inclination to favor information supported or presented as indicators or rankings, but not necessarily to conduct human rights work through indicators. Cf. SALLY ENGLE MERRY, *THE SEDUCTIONS OF QUANTIFICATION: MEASURING HUMAN RIGHTS, GENDER VIOLENCE, AND SEX TRAFFICKING* 9, 220 (2016).

tion of the need for and purpose of measuring human rights as a whole. In this context, Sally Engle Merry's book, *The Seductions of Quantification*, raises valid questions about how and why we have become surrounded by multiple systems of country rankings. For example, there are indices that seek to show us which countries are more corrupt (e.g., the Corruption Perception Index), or less protective of the rights of women (e.g., the Social Institutions and Gender Index, SIGI; or the Gender Inequality Index, GII), or better at guaranteeing our civil liberties (e.g., the Human Freedom Index or the Democracy Index), or overall, simply a better place to live (e.g., the Human Development Index). Such indices have proliferated without the time to step back and adequately assess what these figures really mean, either individually or when placed side by side.

Merry's book offers an anthropological and sociological take on the issue of indicators in relation to the field of human rights. She is not the first, and will certainly not be the last, to discuss the intractable problems with human rights compliance, application, and enforcement—in theory and in practice—of which, we believe, indicators are a part. Yet, we approach Merry's arguments from a legal and practical point of view, as scholars and practitioners in the field of international law.

A reply to her book requires, first, an overview of her work, which we concede is limited for its reductionist character. Merry offers an overarching critique of the indicator culture, both in terms of the process by which indicators are developed and the ways in which they are applied. The three dominant concerns articulated throughout her book are (1) the roles played by *power* and knowledge in the creation, establishment, affirmation, and reiteration of a limited set of indicators for any one topic;⁵ (2) the assumed *universal* nature of indicators and, in turn, their inability to reflect important contextual and cultural differences; and (3) that the choice and design of indicators *define the very problem* that is being measured,⁶ which alone is problematic, but even more so if done in a way that is "decontextualized, homogenized, and remote from local systems of meaning."⁷ Moreover, Merry argues

5. *Id.* at 20.

6. *Id.* at 21.

7. *Id.* at 3.

these indicators are often based on previous indicator projects and hence reinforce the same understanding, *blocking out new notions* or interpretations of the concept at hand.⁸

In *The Seductions of Quantification*, Merry highlights why indicators have become so popular. As articulated in her earlier critical pieces, Merry writes that it is often considered easier to assess compliance with a treaty through numbers.⁹ In concrete terms, indicators simplify the process of evaluating progress and monitoring it over time,¹⁰ and they provide comparability, both for individual countries over the years and between countries. Hence, what has resulted from this “seduction” is numerous indices that rank and compare countries and show gains and losses and winners and losers in each area.

In response to this embrace of indicators, Merry offers three indicator projects as examples to lay out her criticisms. First, Merry dedicates two chapters of her book to the issue of violence against women, discussing the process by which the United Nations Statistical Commission (UNSC) set out to develop a set of indicators and guidelines to collect data on violence against women.¹¹ To make her point, Merry discusses four approaches to the topic of violence: a gender equality approach, a human rights approach, a criminal justice approach, and an approach based on national statistical capacity. By comparing the four approaches, she highlights how while all purport to measure the same idea (i.e., violence against women) in practice, they are “not measuring the same thing at all, even though they are calling it by the same name.”¹²

8. *Id.* at 31.

9. See Sally Engle Merry, *Measuring the World: Indicators, Human Rights, and Global Governance*, 52 CURRENT ANTHROPOLOGY SUPPLEMENT 3, S83, S84 (2011); see also THE QUIET POWER OF INDICATORS: MEASURING GOVERNANCE, CORRUPTION, AND RULE OF LAW 20 (Sally Engle Merry, Kevin E. Davis, & Benedict Kingsbury eds., 2015) (discussing the issue in wider terms beyond human rights).

10. See MERRY, *supra* note 4, at 16–19 (offering a few examples of “successful indicators” as a simplified overview).

11. A brief overview of the contents of the discussions in this process can be found in UN STATISTICAL COMMISSION, REPORT ON THE MEETING OF THE FRIENDS OF THE CHAIR OF THE UNITED NATIONS STATISTICAL COMMISSION ON STATISTICAL INDICATORS ON VIOLENCE AGAINST WOMEN 26 (2010), <https://unstats.un.org/unsd/demographic/meetings/vaw/docs/finalreport.pdf>.

12. MERRY, *supra* note 4, at 107.

While Merry concedes that the effort to quantify violence against women has helped to advance women's rights by bringing greater visibility to the problem, raising awareness, and displacing the belief that violence is a problem for only the "unruly few,"¹³ her biggest gripe with the attempt to quantify the issue is the suggestion that any of these approaches can reveal an "objective truth" about violence.¹⁴ To the contrary, any approach—whether a feminist approach or a human rights one—constructs our understanding of violence in the first place. Merry is concerned that the data may be viewed as the result of "an objective scientific process of data collection and analysis."¹⁵

Merry's second indicator project is the annually-produced United States Trafficking in Persons report ("TIP report"). The centerpiece of the TIP reports is a system for ranking countries in terms of their efforts to prevent and respond to human trafficking. Tier 1 countries are considered the most compliant, Tier 3 the least.¹⁶ A fourth tier, the Tier 2 Watch List was added in 2004.¹⁷ For those countries ranked Tier 3, there is a risk of "negative consequences" in their relations with the United States.¹⁸ Merry identifies numerous problems with the TIP report indicators, including how they have influenced what we understand to be trafficking and lack of objectivity, given the system is heavily intertwined with politics—and the political agenda of the U.S. government in particular.¹⁹ As will be discussed in Part V, Merry is not the first author to highlight the clear agendas behind how trafficking is framed and understood. In our view, the TIP report—heavily critiqued by feminist academics, policy analysts, and other governments—is a parochial evaluation, lacking a strong methodol-

13. *Id.* at 107.

14. *See id.* at 5 ("Rather than revealing truth, indicators create it.").

15. *Id.* at 111.

16. U.S. DEP'T OF STATE, *TRAFFICKING IN PERSONS REPORT: JUNE 2017* (2017), <https://www.state.gov/documents/organization/271339.pdf>.

17. U.S. DEP'T OF STATE, *TRAFFICKING IN PERSONS REPORT: JUNE 2004*, 25 (2004), <https://www.state.gov/documents/organization/34158.pdf>.

18. U.S. DEP'T OF STATE, *supra* note 16 at 29.

19. *Id.* at 114. For a detailed discussion on the issue of politicization see generally Kirstin Roster, *Politicization of the Trafficking in Persons Report; Is Political Proximity to the US Associated with Better Rankings?* (McCourt Sch. Pub. Pol'y, 2016), <https://repository.library.georgetown.edu/handle/10822/1040854>.

ogy and with results that are too heavily influenced by the politics of the day. As such, the TIP report is perhaps a too convenient choice for Merry to demonstrate the problems with quantification. Moreover, we hardly consider the report to be a quantitative indicator, having a more qualitative nature that is subsequently, in a non-transparent manner, transformed into a ranking with tiers.

The third indicator project that Merry discusses is the development, between 2005 and 2012, by the Office of the High Commissioner for Human Rights (OHCHR), of a set of indicators to measure core human rights, such as the right to education and the right to life.²⁰ With this example, Merry's three neat categories become blurred. It is doubtful that Merry intends to suggest that freedom from violence against women is not a human rights concern. A prohibition of gender-based violence has been established as a key principle of customary international law;²¹ and one of ten human rights treaty bodies,²² the Committee on the Elimination of Discrimination against Women (CEDAW), just released a general recommendation on gender-based violence.²³ The first and third indicators projects are thus naturally overlapping.

Similar reasoning could apply to the second and third projects. The set of indicators developed by the OHCHR, for instance, includes the "elimination of forced labour," with human trafficking offered as an example.²⁴ "Violence at work,

20. MERRY, *supra* note 4, at 194–206.

21. Comm. on the Elimination of Discrimination Against Women, *General Recommendation No. 35 on Gender-Based Violence Against Women, Updating General Recommendation No. 19*, U.N. Doc. CEDAW /C/GC/35 ¶2 (2017).

22. The ten human rights treaty bodies are the Committee on Civil and Political Rights; the Committee on Economic, Social and Cultural Rights; the Committee on the Elimination of Discrimination against Women; the Committee on the Elimination of Racial Discrimination; the Committee on the Rights of the Child; the Committee on the Rights of Persons with Disabilities; the Committee against Torture; the Committee on Migrant Workers; and the Sub-committee on the Prevention of Torture and the Committee on Enforced Disappearance. *Cf. supra* note 3.

23. *See* Comm. on the Elimination of Discrimination Against Women, *supra* note 21; *see also* Ramona Vijeyarasa, *A Quarter of a Century Later: Violence Against Women and CEDAW's General Recommendation* (forthcoming 2018) (manuscript at 35) (on file with authors).

24. U.N. High Comm'r for Hum. Rts., *Human Rights Indicators: A Guide to Measurement and Implementation*, at 95, U.N. Doc. HR/Pub/12/5 (2012).

forced labour and trafficking” are specifically addressed as part of the violence against women indicators.²⁵ In the end, all three projects are about measuring compliance with human rights standards, looking into certain rights in more or less detail. In relation to the OHCHR project, Merry’s main concern seems to be a bias towards the global North in the development of those indicators. We will look into the underlying logic of that argument in section II.A.

This article analyzes Merry’s main critiques of the growing interest—Merry might say obsession—with quantifying progress on human rights and development. It assesses whether Merry adequately defends and sustains her critiques through her three examples, and it questions whether these critiques are any different when it comes to other mechanisms for accountability and tracking change within the human rights system at large. It also analyzes Merry’s choice of projects presented in the book as examples: are these particularly weak examples or are the critiques directed at them equally sustained when it comes to other examples of accountability sought through quantification and rankings? We conclude by arguing that Merry’s view, while raising a valid point (i.e. the enthusiastic use of indicators hides the significant shortcomings of quantification) does not offer a balanced evaluation of how to monitor human rights through quantitative approaches, especially in terms of a cost-benefit analysis and the best means to ensure access to the relevant information presented in a particular data set.

A. *Defining the Concepts: What Are Indicators?*

The term indicator is used for many different kinds of numerical representations with no real consistency. Its meaning ranges from something that indicates or points to a fact, such as that a person has been trafficked, to an elaborate combination of data merged into a single rank or score.²⁶ Merry defines quantification as “the use of numbers to describe social phenomena in countable and commensurable terms.”²⁷ Yet, in her book—as in her previous works—Merry focuses on social phenomena largely related to the enjoyment of human rights

25. *Id.* at 99.

26. MERRY, *supra* note 4, at 13.

27. *Id.* at 1.

rather than other measureable phenomena. Unsurprisingly, an expansive number of social phenomena are currently tracked and quantified, many of which have only limited or indirect connections with human rights. Some examples include patterns of wealth, the use of social media, penetration of smart mobile use, and demographic data related to sexual orientation.

Given that a large body of existing indices, and particularly the better known ones, tend to measure the enjoyment or protection of different aspects of human rights, our assessment is founded on an assumption that those indices are often proxies for human rights accountability. In other words, if we measure a population's access to drinkable water, we would be looking at compliance with the right to water and sanitation. If we rank countries on the basis of access to contraceptives for their citizens, we would be trying to assess the enjoyment of sexual and reproductive rights. If we quantify how much corruption is perceived in a country, we would be doing so because we believe people should live free of corruption and under the rule of law. In the end, we are quantifying human rights. And, yet, this says very little about what we mean when we talk about quantification, because this concept covers many different interpretations of complex processes that are analyzed with very different lenses and at different levels.

In general, instead of discussing "indicators" at large, this article, and most of Merry's related work, could be more properly described as scrutinizing "human rights indicators," which have been defined as "information that is presented quantitatively and that can be related to human rights norms and standards; that addresses and reflects the human rights concerns and principles; and that are used to assess and monitor promotion and protection of human rights."²⁸ To help explain what is meant by "quantitatively," in the following section we present five different levels of analysis for how we understand indicators and measurement.

28. Rajeev Malhotra & Nicolas Fasel, *Quantitative Human Rights Indicators: A Survey of Major Initiatives* 3 (Mar. 3, 2005) (unpublished manuscript), <http://gaportal.org/sites/default/files/Quantitative%20Human%20Rights%20Indicators.pdf>.

B. *Distinguishing Five Levels of Analysis*

Indicators can be analyzed from at least five different approaches, or “levels of analysis.”²⁹ First is the “*ontological*” level, which examines their nature of being and becoming a reality. Put simply, this normative analysis looks into whether and why indicators should exist. Further, assuming that indicators add value, at this level, we can also interrogate how indicators can be grouped in order to achieve a specific aim, especially if the goal is to establish some kind of hierarchy or taxonomy for the indicators developed. We can also ask whether indicators and their ranks can be subdivided into smaller groups according to similarities and differences.

If we agree that indicators are useful or necessary, as well as a valid way of assessing compliance with human rights, the second level focuses on *the construction of those indicators*. As a human construction, the choices made in the process of developing an indicator—or a system of indicators—are naturally influenced by cultural backgrounds and related biases. Given that we claim and defend that human rights are by their definition universal, we should be able to measure compliance with them in a globally agreed upon way. At the same time, we should recognize how cultural biases affect those choices. Unlike Merry, who sees indicators as failing to adequately respond to cultural divergences,³⁰ this second level of analysis enables us to see culture embedded in the choice of what gets measured and how.

Third, assuming that an indicator is developed in a globally accepted manner and that it is feasible to measure it (e.g., quantitatively or qualitatively), we can move into the *procedural and technical level*, where we need to understand how the necessary information (i.e., inputs) to determine a performance for that indicator (i.e., output) can be collected, processed, validated and, finally, presented. Again, all such steps towards collecting inputs to develop an output are understood as a way of

29. Merry et al. describe four phases in the life of an indicator: conceptualization, production, uses and effects, and impacts. *Supra* note 9 at 10.

30. MERRY, *supra* note 4, at 168 (citing Russel Lawrence Barsh, *Measuring Human Rights: Problems of Methodology and Purpose*, 15 HUM. RTS. Q. 87 (1993), and Todd Landman, *Measuring Human Rights: Principle, Practice, and Policy*, 26 HUM. RTS. Q. 906 (2004)). Merry’s point is better explained in her own words. See MERRY, *supra* note 4, at 215.

assessing compliance with human rights in a comparable manner, with whatever disclaimers may be necessary.

At the next level, the *evaluation level*, we can assess what those indicators say about compliance with human rights in a given jurisdiction. We must assess the reliability of the information we have obtained from the process. This is a prerequisite to undertaking a normative analysis of those results, such as “is performance x of country A, better than performance y of country B?” or “what is the desirable/ideal performance in relation to indicator N measuring compliance with the right to a fair trial?”

Finally, the fifth level is the *consequential level*. At this stage, we consider whether it is fair and reasonable for performance of a given indicator to result in practical repercussions. One can think of punishments and rewards, or “carrots and sticks” as possible legitimate consequences.

Having set out these five stages—the ontological stage, the construction stage, the procedural and technical stage, the evaluation stage, and finally, the consequential stage—the next sections will assess Merry’s arguments against the stages above. At each of these points, we analyze how Merry defends that indicators are failing to meet their expected effect, how she sustains these criticisms, and what alternatives exist for human rights accountability.

II. THE ONTOLOGY OF INDICATORS: WHAT THEY ARE AND WHY WE NEED THEM

Merry’s book opens with a harsh critique of indicators: we have been fooled as to their appropriateness, believing them to be objective when they are not, relying on them in our fight to address global problems, when in fact, they are not universally applicable.³¹ Her critique is extended not only to the human rights and the development sector communities but also to the media and civil society at large:

Once indicators are established and settled, they are typically portrayed in the media as accurate descriptions of the world. They offer forms of information that satisfy the unease and anxiety of living in a complex and ultimately unknowable world. They address

31. MERRY, *supra* note 4 at 1, 19–21.

a desire for unambiguous knowledge, free of political bias. Statistical information can be used to legitimate political decisions as being scientific and evidence-based in a time when politics is questioned. They are buoyed up by the rise in bureaucracy and faith in solutions to problems that rely on statistical expertise. Such technocratic knowledge seems more reliable than political perspectives in generating solutions to problems, since it appears pragmatic and instrumental rather than ideological. These are the seductions of quantification.³²

While her criticisms are largely substantiated, Merry lacks the background analysis that helps explain the increased interest in quantification, particularly during the past two decades. Moreover, readers are left questioning whether some of Merry's legitimate critiques are equally directed at other systems of accountability. In other words, is the issue the indicators themselves or the reality of how States Parties respond when being held accountable for their observance of human rights? Finally, does Merry merely present a negative appraisal of the *status quo* or offer a potential and feasible alternative? In the next sections, we seek to answer some of these questions and offer a more balanced and nuanced view of the use of indicators for the monitoring of compliance with human rights standards and its current shortcomings.

A. *Indicators and Their Western Origins*

It is hard to disagree with Merry that the interest in and use of indicators is, to a significant extent, a non-inclusive process, or, if collaborative, often driven by representatives of States Parties from Europe and the Americas.³³ The real question here is if this is something particular to the development of indicators created to measure rights and welfare or part of a wider picture in which, as Merry describes it, the “global North”³⁴ countries—generally rich and developed States—are at the origin of the existing international system of human rights and global governance. We could go even further and claim that, in the end, the whole idea of human rights is a

32. MERRY, *supra* note 4, at 4.

33. *Id.* at 180.

34. *Id.* at 204.

Western (or European) driven idea, shaped through the global and regional international institutions created during the twentieth century. These were, to a significant extent, the result of a promotion of an understanding of human rights supported by the global North.³⁵

Based on this reasoning, to offer one example, we can measure political freedom according to certain proxies such as access to information, freedom to manifest political ideas or freedom to vote. In doing so, we adopt a particular way of understanding political freedom. This particular understanding corresponds to the ideas promoted by the bodies entrusted with their monitoring, such as the Human Rights Committee, which monitors the International Covenant on Civil and Political Rights.³⁶ It cannot be surprising that the output of this Committee is rooted in the ways in which global North countries understood these rights in the first place and the role they played in influencing the drafting of the treaties themselves.

In this respect, there is an intrinsic Western bias with the use of indicators because they try to measure understandings of concepts with a heavy Western influence. Yet, does this problem—if it is indeed problematic—justify a shift away from indicators altogether? For Merry, her concerns with the Western origins of indicators appear to be part of a broader critique of the power dynamics inherent in the development of indicators. We therefore turn to the construction stage—the identification of what gets measured and how.

35. See generally Andrew Moravcsik, *The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe*, 54 INT. ORG. 217 (2000) (offering an account of the origins of the institutionalization of human rights regimes in the European context).

36. For example, the link between access to information, freedom to manifest political ideas, and the right to take part in the conduct of public affairs is a recurrent reasoning of the Human Rights Committee, often referring to these concepts as closely connected. See UN HUMAN RIGHTS COMMITTEE (HRC), CCPR GENERAL COMMENT NO. 25: ARTICLE 25 (PARTICIPATION IN PUBLIC AFFAIRS AND THE RIGHT TO VOTE), THE RIGHT TO PARTICIPATE IN PUBLIC AFFAIRS, VOTING RIGHTS AND THE RIGHT OF EQUAL ACCESS TO PUBLIC SERVICE 8 (1996); UN HUMAN RIGHTS COMMITTEE (HRC), CCPR GENERAL COMMENT NO. 34 ARTICLE 19: FREEDOMS OF OPINION AND EXPRESSION 13 (2011), <http://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>.

III. HOW ARE INDICATORS CHOSEN AND DEVELOPED? CONSTRUCTING INDICATORS

Merry draws on Foucault to establish her critique of the tight knit relationship between indicators and power. Power determines what gets measured and the practices of measurement and quantification, which in turn produce knowledge. In the words of Foucault, “power and knowledge directly imply one another.”³⁷ Moreover, once indicators are established, normalized, and readily used, we tend not to return to the initial concern of who played the most central role in their development. Merry argues, “[O]nce decisions about categorization and aggregation are made, the categories may come to seem objective and natural, while the power exerted in creating them disappears.”³⁸ In turn, we stop questioning why a certain measure is being taken—or, probably more relevant here, on whose determination—and the outcome is considered a new norm.

A. *The Voice of the North and Measuring Violence Against Women*

When discussing the first of the three projects—the development of indicators related to violence against women—Merry critiques how the process of developing indicators to measure the gravity of and responses to violence shapes our understanding of the issue: “As the UNSC decided what to measure, how to categorize it, and how to gather data, it defined the phenomenon.”³⁹ Merry in effect is arguing that indicators for violence set the boundaries for what society understands to mean by the term violence.

However, for experts from the field of women’s rights or, specifically, violence suffered by women, Merry’s claim is a stretch. Both rape and intimate-partner violence were unquestionably recognized as forms of violence against women long before this exercise. Her comment clearly speaks to the issues of child marriage, female genital mutilation (i.e., FGM, although Merry uses the term “cutting”), sexual slavery during

37. MERRY, *supra* note 4, at 29 (quoting MICHEL FOUCAULT, *DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON* 27 (1979)).

38. *Id.* at 77.

39. *Id.* at 44.

wartime, and violence while in police custody—issues that tend to be included in broader conceptions of violence, but ignored in a narrower approach that focuses on rape and intimate partner violence.⁴⁰ However, again, it is a stretch to suggest that until this moment in time violence was not associated with these rights violations. We must recall that the development of the indicators that Merry analyzes started in the mid-2000s, long after the Convention on the Elimination of Discrimination against Women (CEDAW) was drafted (1989),⁴¹ the Beijing Declaration and Platform for Action issued (1995),⁴² or the CEDAW Committee's General Recommendations No. 12 (1989)⁴³ and No. 19 (1992)⁴⁴ on violence against women adopted.

Take for instance the case of FGM. FGM has long been recognized as not only a form of physical violence but also emotional violence,⁴⁵ the violence itself closely intertwined with the perverse control over women's sexuality that is exercised as a result of the practice. Merry herself makes this point in the pages that follow:

As the national and international focus on the issue developed, it gradually expanded from a narrow concern with rape and domestic violence to a broader analysis of the way patriarchal social structures, cultural understandings of gender difference, and social

40. *Id.*

41. Convention on the Elimination of All Forms of Discrimination Against Woman, Dec. 18, 1979, 1249 U.N.T.S. 13 (entered into force Sept. 3, 1981).

42. See U.N. Entity for Gender Equality and the Empowerment of Women, Rep. of the Fourth World Conf. on Women, U.N. Doc. A/CONF. 177/20/Add.1 (1995); see also U.N. Entity for Gender Equality and the Empowerment of Women, Beijing Declaration and Platform of Action, Fourth World Conf. on Women (1995), <http://www.un.org/womenwatch/daw/beijing/platform/>.

43. Comm. on the Elimination of Discrimination Against Women [CEDAW], General Recommendation 12 (1989), <http://www.un.org/womenwatch/daw/cedaw/recommendations/recomm.htm#recom12>.

44. Comm. on the Elimination of Discrimination Against Women [CEDAW], General Recommendation 19 (1992), <http://www.un.org/womenwatch/daw/cedaw/recommendations/recomm.htm#recom19>.

45. Ruth F. Lax, *Socially Sanctioned Violence Against Women: Female Genital Mutilation is its Most Brutal Form*, 28 CLINICAL SOC. WORK J.; NEW YORK 403–412, 502 (2000).

and economic inequality contributed to a wide range of forms of violence.⁴⁶

In many respects, for readers, the UNSC's decision to develop a set of indicators on violence seems an odd choice to sustain Merry's critique. In fact, that process evidences how an extensive, well-debated, and open-minded effort to develop a set of indicators can result in a fairly strong outcome. Merry explains how the initial list of indicators to measure violence against women—a set that were subsequently approved without further discussion—was in fact very narrow.⁴⁷ While it included FGM,⁴⁸ it excluded a host of other forms of violence, including state violence by police or the military and harassment in the workplace.

However, after much negotiation, the final guidelines, while recognizing the difficulty of measuring the severity of violence, stretched the categories of acts and consequences. The final list of consequences of violence to be measured under that set of indicators includes injuries, mental health repercussions, miscarriage, violence during pregnancy, and fear of the perpetrator,⁴⁹ despite discussions during their development about the difficulty of measuring fear.⁵⁰ The guidelines also emphasize ethical and safety concerns during interviews, a major concern of the World Health Organization's (WHO) study and feminist activists involved in the project.⁵¹ Overall, there was an expansion in scope in terms of how violence was framed and experienced as a controlling behavior.

As Merry herself states, “[G]ender equality advocates worked hard to bring their perspective to the UNSC indicator project . . . [and] succeeded in expanding the UNSC's initial narrow view of violence against women.”⁵² Their more expansive questionnaire and detailed ethical and safety recommendations were incorporated into the guidelines, facilitating greater disclosure of the scope of the problem.⁵³ Recognizing the “triumph” of the feminists at the negotiating table, it is

46. MERRY, *supra* note 4, at 46.

47. *Id.* at 62.

48. *Id.*

49. *Id.* at 82.

50. *Id.* at 82.

51. *Id.* at 73.

52. *Id.* at 90.

53. *Id.*

confusing when Merry critiques the fact that “[t]hrough the process of deciding how to measure something like severe physical violence, the concept is established and defined.”⁵⁴

To the contrary, Merry’s extensive discussion suggests that the process of developing an agreed set of what can be measured is a negotiation, creates a dialogue about women’s experiences—in this case, of violence and what matters for women who experience violence—and considers the feasibility of undertaking the measurement that is being proposed. Merry’s direct exposure to the process, as described in the book, highlights that there are many players involved in the development of indicators. Perhaps, therefore, it is less a question of critiquing the indicators themselves but more about influencing the expertise of the people at the table—ensuring for women’s rights issues, for instance, that there are feminist perspectives rooted in international human rights standards.

A more interesting approach regarding bias in the development of indicators would be to analyze the possible underlying reasons for the assumptions that serve as the basis of indicators. Merry, for example, shows her own interpretation of the drivers of trafficking when she discusses her second indicator project, the U.S. State Department’s TIP reports.⁵⁵ When referring to the TIP reports, she states that the reports “expose deficiencies in state efforts to tackle rural poverty or oppressive marriage practices, even though both of these also fuel trafficking.”⁵⁶ This assumption contradicts existing research, which challenges the necessary causal relationship between human trafficking and poverty⁵⁷ or trafficking and specific manifestations of gender inequality that are often presumed to be relevant without substantiation.⁵⁸

Merry concedes that “experience in measuring violence against women trumped the skepticism of the statisticians.”⁵⁹ In fact, the preferred model of the feminists—even if a northern-dominated perspective—was the only one annexed to the

54. *Id.* at 85.

55. *See* U.S. DEP’T OF STATE, *supra* note 16.

56. *Id.* at 20.

57. RAMONA VIJAYARASA, *SEX, SLAVERY AND THE TRAFFICKED WOMAN: MYTHS AND MISCONCEPTIONS ABOUT TRAFFICKING AND ITS VICTIMS* (2015).

58. *Id.* at 141-144.

59. MERRY, *supra* note 4, at 74.

UNSC guidelines published in 2013,⁶⁰ thereby giving the suggestion that it should be preferred too globally when it comes to national data collection. That a feminist expertise “won” in the exercise of developing indicators on violence against women reflects positively on the potential of indicators to garner interest in measuring change in pertinent women’s rights issues.

B. *Expertise Inertia or Learning from Experience?*

A final question regarding indicator construction is the appropriateness of Merry’s use of the phrase “expertise inertia.”⁶¹ Merry highlights how expert papers from past meetings often form the basis of future discussions and new lists of indicators. She uses the example of a 2007 discussion on indicators related to violence against women, the substance of which was based on formerly used indicators from the U.N. High Commissioner for Refugees, U.N. Statistical Commission for Latin America and the Caribbean (UNECLAC), the U.N. Economic Commission for Europe (UNECE), the European Union, Canada, the United States, United Kingdom, Italy, and Spain.⁶² Similarly, a 2009 “Friends of the Chair” of the U.N. Statistical Commission (UNSC) meeting in Mexico relied on a preexisting national survey to finalize its list of indicators.⁶³ In short, to Merry, these examples demonstrate the tendency to rely on previously applied indicators rather than generate new ones.

However, it is questionable that a reliance on past lessons is necessarily a bad thing. Moreover, Merry notes that not all indicator projects are equally flawed⁶⁴ and offers several alternative approaches to measuring violence against women. For example, a project initiated by the UNECE and UNECLAC adopted a “feminist vision of violence against women.”⁶⁵ The final survey model included controlling behaviors, emotional abuse, fear, and their effects on everyday activities in addition

60. *Id.* at 73.

61. *Id.* at 6.

62. *Id.* at 71.

63. *Id.*

64. *Id.* at 25, 216.

65. *Id.* at 66.

to physical and sexual violence.⁶⁶ The WHO's study of domestic violence and women's health from 2005 also helped to influence the UNECE survey instrument, which incorporated a public health perspective in its focus on psychological harm and family relationships.⁶⁷

Merry also notes the work of Yakin Ertürk, the second Special Rapporteur on Violence against Women—successor to Radhika Coomaraswamy—who had a broad vision for what should be included when we measure violence against women. As Merry notes, Ertürk proposed three broad indicators: “grave violence,” covering issues like rape, intimate partner violence, sexual harassment; “femicide” (i.e., the murder of women); and third, “social tolerance” (i.e., social and cultural contexts that allow violence to continue).⁶⁸ Merry labels Ertürk's approach as a feminist one, defining violence in its broadest sense, and the battlefield broadly too, to include, for example, violence in the work place.⁶⁹

Readers are left with a strong sense, therefore, that at least in the examples offered on measuring violence against women, we have—particularly with strong feminist expertise at the table—come a long way to a more expansive and gender-sensitive means of measuring experiences of and responses to violence against women and that cumulative processes do not necessarily mean “expertise inertia.”

IV. HOW TO TRANSFORM INFORMATION INTO QUANTITATIVE MEASUREMENTS: UNIVERSAL INDICATORS AND THE DIFFERENCES LOST IN THE CRACKS

Merry shows great concern with how local systems of knowledge “must be squeezed” into global ones.⁷⁰ For example, to measure violence against women, throwing acid in the face of one's wife in Bangladesh is equated to shooting a domestic partner in the United States.⁷¹ There is little scope for

66. *Id.* at 88.

67. *Id.*

68. *Id.* at 64.

69. *Id.*

70. *Id.* at 21.

71. *Id.* at 77-78.

nuanced and local difference. This is what Merry calls the problem of “producing commensurability.”⁷²

Merry’s concerns echo the debates over the universalism of human rights which in part explain the existence of universal indicators. The strength of the normative framework established by the Universal Declaration of Human Rights (UDHR) in 1948 is that it calls on all states to accept that *all* human beings are *universally* entitled to the basic rights and freedoms contained in it.⁷³ Human rights are considered universal. If we accept this, we should accept that we can create universal systems to measure these global principles.

However, some may still argue that an indicator designed to measure a particular human right will only be as universal as the right itself. For instance, we have seen an extensive number of reservations by Muslim States to the Convention on the Elimination of Discrimination against Women (CEDAW), namely in relation to Articles 2, 9, 10 and, particularly, to 16 on family life.⁷⁴ This is despite the fact that CEDAW is one of the most ratified and globally accepted treaties.⁷⁵ One could argue that any indicator trying to assess compliance with those articles would not respect the cultural differences of certain countries. Yet, the problem would then lie not in the indicator but in the (contended) “cultural insensitivity” of the right itself.

A related but different issue appears when the right itself is not under debate, but instead the proxy used to measure the right does not adequately capture cultural or social differences.⁷⁶ This valid point—which Merry covers succinctly through examples—has been explored in the context of indicators used in the development-sector to measure well-being

72. *Id.* at 212.

73. G.A. Res. 217 (III) A, pmbl. & art. 2, Universal Declaration of Human Rights (Dec. 10, 1948).

74. *Declarations, Reservations and Objections to CEDAW*, U.N. WOMEN, <http://www.un.org/womenwatch/daw/cedaw/reservations-country.htm> (last visited Mar. 17, 2018); see Judith Resnik, *Comparative (In)equalities: CEDAW, the Jurisdiction of Gender, and the Heterogeneity of Transnational Law Production*, 10 INT’L J. CONST. L. 531, 547 (2012).

75. See *Convention on the Elimination of All Forms of Discrimination Against Women: Status*, UNITED NATIONS TREATY COLLECTION, https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-8&chapter=4&lang=en (last updated Mar. 31, 2018) (detailing ratification status).

76. MERRY, *supra* note 4, at 77.

and enjoyment of fundamental human rights. For example, former Millennium Development Goal (MDG) Three—on the promotion of gender equality and the empowerment of women⁷⁷—has been criticized for its lack of global applicability.⁷⁸ One of its indicators was gender parity in education (i.e., Target 3A) at the primary and secondary level.⁷⁹ If we apply it to the Latin American context, this region would be assessed as a champion of gender equality, as disparity in school enrollment and completion between boys and girls has been almost eliminated.⁸⁰ Yet, gender inequality remains rife, to varying degrees and across a range of issues, across the region.⁸¹ This would suggest the need for a Latin America-specific target on gender equality that does not focus on girls in schools.

A third problem is developing an indicator that ignores local manifestations of broader categories. When discussing her violence against women indicator project, Merry details the difficulty in developing a global measure of violence against women based on severity. In this context, she points out that there are “national and local differences in what is interpreted as a severe act,”⁸² giving the example of Bangladesh. At a “Friends of the Chair” meeting, Bangladesh suggested adding to the list of acts of physical violence additional examples such as acid throwing, dropping from a higher place, and a kick in the abdomen; and to the list of psychological violence, adding pressure for dowry, not paying attention to the children, and rebuking the giving of birth to a female child.⁸³ All examples, Merry points out, are particularly rele-

77. G.A. Res. 55/2, ¶ 20, United Nations Millennium Declaration (Sept. 18, 2000); see also *Promote Gender Equality and Empower Women*, UNITED NATIONS DEV. PROGRAMME, http://www.undp.org/content/undp/en/home/sdgoverview/mdg_goals/mdg3 (last visited Mar. 17, 2018).

78. Ceri Hayes, *Out of the Margins: The MDGs Through a CEDAW Lens*, 13 GEND. & DEV. 67, 68 (2005).

79. UNITED NATIONS & 2015 TIME FOR GLOB. ACTION, THE MILLENNIUM GOALS DEVELOPMENT REPORT 2015, 28 (2015), http://www.undp.org/content/dam/undp/library/MDG/english/UNDP_MDG_Report_2015.pdf.

80. See generally SUZANNE DURYEA ET AL., THE EDUCATIONAL GENDER GAP IN LATIN AMERICA AND THE CARIBBEAN (Social Science Research Network) (2007), <https://papers.ssrn.com/abstract=1820870> (last visited Apr. 10, 2018); see also HAYES, *supra* note 78, at 68.

81. HAYES, *supra* note 78, at 68.

82. MERRY, *supra* note 4, at 77.

83. *Id.* at 77–78.

vant to Bangladesh and South Asia in general,⁸⁴ and not including them in an indicator means not measuring an egregious violation of the human rights of women in the region. This is a valid criticism—although it seems that the final version of the indicator covered these situations—but the real question is if non-quantitative existing human rights mechanisms would result in more encompassing, culturally-sensitive measurements.

At the same time, accommodating nuance and local difference, instead of relying on broader categories, can be a problem. Merry does not offer this critique, but a fourth aspect of the issue of commensurability—and an equally problematic one—occurs when an indicator is developed for a country or region-specific problem and then assessed on a global scale. For example, in 2009 the Social Institutions and Gender Index—introduced in 2007 by the Organisation for Economic Cooperation and Development (OECD) and only applicable to countries in the global South—contained only two measures for barriers to women’s enjoyment of their civil liberties, one of which was an obligation to wear a veil or burqa to cover parts of the body in public.⁸⁵ Such an indicator reflects a bias from the outset targeting the Muslim religion and is an indicator that makes no sense when applied on a global scale. Similarly, the choice to include a measure on the prevalence of FGM—unquestionably an egregious violation of the rights of women and girls—is also open to critique as this too is not a globally prevalent phenomenon. Moreover, is measuring the prevalence of FGM more justified than measuring rates of correctional rape of lesbian women or acid attacks on women?⁸⁶

In practice, indicators designed primarily for the purpose of human rights accountability are no more flawed in their at-

84. *Id.* at 78.

85. Liebowitz & Zwingel, *supra* note 1, at 375. The other measure of civil liberties under the Social Institutions and Gender Index was the freedom of women to move outside of the home. See generally *The OECD Social Institutions and Gender Index*, ORG. FOR ECON. CO-OPERATION & DEV. [OECD], <http://www.oecd.org/dev/development-gender/theoecdsocialinstitutionsandgenderindex.htm#results> (last visited Mar. 17, 2018).

86. See Ramona Vijayarasa, *In Defence of CEDAW: The Need to Return to the Foundations of Women’s Rights Law to Measure Progress on Gender Equality* (forthcoming 2018) (manuscript at 19) (on file with authors).

tempt to present a universal measure than the human rights system itself. Having accepted the cons and many pros of a universal human rights system—and we are of the view that human rights are truly universal and must be accepted as such if progress on their enjoyment is to succeed—we must accept that no global measure will be able to account for the many micro nuances among countries. This is the space for country-based indicators, which should be fully able to measure human rights accountability but within a national context.⁸⁷

Merry's critique is both passionate and understandable: "the complexity of relationships—their histories, the interplay of love and fear, the role of kinship, and residence patterns—disappear."⁸⁸ Problematically, however, the only way to collect this type of data is qualitatively, and it is hard to imagine how countries can collect national data on the scale intended by the UNSC, or the data necessary to allow for any kind of commensurability, if all subjective, qualitative considerations are to be included. It is here that we turn to the next step of collecting and processing data and related practicalities.

V. THE PRACTICALITIES OF COLLECTING AND PROCESSING THE DATA—THE EXAMPLE OF HUMAN TRAFFICKING

As discussed in Part I, it is not enough to accept that an indicator is a valid way to acquire and present information—at the ontological level—and agree on what we are measuring and how—at the construction level. Before creating an indicator, it is also important to consider how the necessary information (i.e., inputs) to determine a performance for that indicator (i.e., output) can be collected, processed, validated, and, finally, presented. At this procedural and technical level, the purpose is to produce a measure that can be compared horizontally—among jurisdictions—or sequentially—in the same jurisdiction, but over a particular number of years.

Here, there are two separate issues worth examining. First, there is the possible existence of an intrinsic bias, where the process of quantification is "adjusted" to potentially achieve a more "palatable" outcome for those who initiated—

87. *Id.* Ramona Vijeyarasa proposes a quantitative accountability system, grounded in CEDAW, where each article is translated into a measurable indicator with direct domestic relevance.

88. MERRY, *supra* note 4, at 62.

or controlled or even paid for—the quantification. Second, there is a more technical aspect, which looks into the methodology and transparency of the process itself. To consider these two aspects, we will use Merry’s project on the TIP reports produced by the U.S. Government.

It is important to recall our previous comments on the particular choice of this indicator and its reliability.⁸⁹ Despite this, the TIP report is a frequently cited source, precisely because the U.S. Government is the only one that produces an annual ranking of other governments’ actions, something that other reports and organizations operating in the field of trafficking choose not to do.⁹⁰ However, considered by some a moral crusade,⁹¹ significant questions have been raised as to whether the U.S. government has access to sufficient data or the political independence to make an accurate assessment.⁹² In fact, we suggest that Merry is wrong when she lists other systems that measure trafficking, including that of the ILO, UNODC, and IOM but nonetheless concludes that, “[a]t least for now, trafficking is what the *TIP Reports* measure.”⁹³ Many of these other organizations are more respected for their contributions to trafficking and more readily turned to than the U.S. government.⁹⁴

Moreover, it is problematic to discuss a ranking process—calling it an indicator is probably excessive—produced by one actor about the performance on trafficking in persons of its allies and enemies. Merry’s view is that overtime the United States TIP reports’ method of data collection has improved and “it has acquired a more settled and legitimate status.”⁹⁵ This is, at best, questionable. As a result, Merry’s decision to dedicate an extensive number of pages to critiquing what is

89. See *supra* Part I.

90. VIJAYARASA, *supra* note 57, at 27.

91. See Laura Agustín, *TIP: Trafficking in Persons, the No Methodology Report*, *NAKED ANTHROPOLOGIST* (June 26, 2009), <https://www.lauraagustin.com/tip-trafficking-in-persons-the-no-methodology-report>.

92. VIJAYARASA, *supra* note 57, at 8.

93. MERRY, *supra* note 4, at 138.

94. See, e.g., Liz Kelly, “You Can Find Anything You Want”: A Critical Reflection on Research on Trafficking in Persons within and into Europe, 43 *INT’L MIGRATION* 235–265, 254 (2005).

95. MERRY, *supra* note 4, at 114.

already a heavily critiqued system of rankings is surprising.⁹⁶ We therefore respond to Merry's arguments concerning the development of indicators to measure human trafficking but with reservation: the TIP reports are a poor example, an odd choice because it is somewhat self-serving to prove the shortcomings of indicators at large.

The agenda of the U.S. government necessarily permeates the whole exercise. Yet, this is not particular to the TIP reports or other indicators used to measure human trafficking. There are clear agendas behind how trafficking is framed and understood,⁹⁷ and as such, there will also be agendas behind any process designed to measure trafficking. If we undertake a qualitative analysis of trafficking across a number of countries, the reports of different State Parties, the accounts provided by different people consulted for the exercise, and the interpretation of a particular country context—again, based on a set of consulted experts—will always have an element of subjectivity. The best way to address the impact of such subjectivity in the final assessment is to promote transparency by different actors about their potential interests in a particular outcome.

Yet in the case of the TIP reports, this subjectivity engulfs their utility. The politics behind the tier system of ranking countries is very clear and has been critiqued elsewhere.⁹⁸ In 2012, for example, Kosovo, Moldova, and Mexico were ranked Tier 2, China was ranked Tier 2 Watch List, whereas Cuba was ranked Tier 3.⁹⁹ Some authors note that the report is “fused with the US’ foreign policy agenda” and flawed by “messy slip-pages with the use of numbers and statistics.”¹⁰⁰ Moreover, as

96. To her credit, Merry makes a valid argument that the measurement of human trafficking has proven particularly elusive. Trafficking is indeed poorly understood and not well defined. These factors make it harder to successfully develop a system of measurement. MERRY, *supra* note 4, at 129.

97. VIJEYARASA, *supra* note 57, at 19–39. While many stakeholders have a legitimate interest in countering trafficking, we have also seen how human trafficking has been presented in a sensationalized or biased light to obtain particular gains. This spans a host of agendas, from a desire for abolition of prostitution to the financial and readership goals of profit-driven media to government goals around regulated or closed borders.

98. VIJEYARASA, *supra* note 57, at 8.

99. U.S. DEP'T OF STATE, *TRAFFICKING IN PERSONS REPORT: JUNE 2012*, 53 (2012), <https://www.state.gov/documents/organization/192587.pdf>.

100. Pardis Mahdavi, *Just the “TIP” of the Iceberg: The 2011 Trafficking in Persons Report (TIP) Falls Short of Expectations*, HUFFINGTON POST, <https://>

Merry points out, the Tier system is used to determine the U.S. government's spending in this area, which is not insignificant. Merry notes that by 2005, the U.S. government had spent \$295 million combatting trafficking in 120 countries.¹⁰¹

Therefore, when Merry critiques the TIP report as being “an apparently objective way of understanding the world,”¹⁰² two responses seem evident. First, the TIP reports are not actually believed to be objective. This is suggested by the fact that quite extensive critiques of the report continue after the noted methodological improvements, even by several of the authors that Merry herself draws upon for her analysis. It is hard to believe that any of the experts working in the field of trafficking would see these reports as objective. Perhaps it is only the U.S. government that presents the report as being objective, but many readers know better.

It is also not new to suggest, as Merry does, that there is a tendency within trafficking circles to repeat the same stories and statistics.¹⁰³ There is in fact very limited space for researchers and practitioners working on trafficking to debate outside of an “echo chamber” of repeated ideas.¹⁰⁴ There have also been challenges to the key players in the trafficking movement for defining the very phenomenon under discussion: the key stakeholders involved in trafficking “not only analyse the phenomenon of human trafficking and debate among themselves about the causes and consequences, but in fact establish the lens through which society understands human trafficking.”¹⁰⁵ Moreover, “[s]uch assumptions easily become self-fulfilling: where Siddharth Kara searches for ‘slaves,’ he finds slaves; if we searched for trafficked women who fit the victim archetype, we would indeed find them.”¹⁰⁶

www.huffingtonpost.com/pardis-mahdavi/just-the-tip-of-the-icebe_1_b_888618.html (last visited Apr. 10, 2018); *see generally* Roster, *supra* note 19 (focusing specifically on the politicization of global rankings).

101. MERRY, *supra* note 4, at 149. Ramona Vijayarasa has elsewhere noted the role that the U.S. government has played as a donor in distorting global priorities when it comes to aid and anti-trafficking programs. VIJAYARASA, *supra* note 57, at 27–28.

102. MERRY, *supra* note 4, at 138.

103. *Id.* at 125.

104. VIJAYARASA, *supra* note 57, at xvii.

105. *Id.* at 19.

106. *Id.* at 20.

We are therefore left unsure about how to respond to Merry's discussion of the TIP reports. This project is a poor choice, particularly compared to the much richer discussion Merry offers on the violence against women indicator *projects*—and here we emphasize “projects”—or in her discussion of the initiatives to develop human rights indicators presented in chapters seven and eight. Obviously, if all indicators—or even a significant number of them—were developed with the poor methodology and blatant political bias present in the TIP reports, indicators would have never acquired the success¹⁰⁷ or the myth of objectivity¹⁰⁸ that Merry grants them.

However, this is no impediment to acknowledging the value of Merry's critique of indicators at the procedural and technical level. Even some of the most known (i.e., “successful”) and methodologically robust (i.e., “objective”) indicators, when analyzed carefully, show obvious shortcomings. One recent example has been the publication of a compelling criticism of the Corruption Perception Index (CPI),¹⁰⁹ challenging the CPI for its lack of objective grounding, despite a thorough review of its methodology conducted in 2012.¹¹⁰ With this in mind, it is essential to always demand a transparent methodology for any kind of quantification or measurement process, so it can be scrutinized and evaluated against the vested interests of those involved, and for the release of any potential conflict of interest in terms of funding, staff interests, organizational values, and objectives.

VI. WHAT INDICATORS SAY ABOUT COMPLIANCE WITH HUMAN RIGHTS: WHERE THE REAL PROBLEM LIES

At the evaluation level, we assess what a given indicator says about compliance with human rights in a given jurisdiction and ask how indicators have been used to measure human rights compliance and to heighten accountability. Merry ar-

107. MERRY, *supra* note 4, at 16.

108. *Id.* at 19.

109. Alina Mungiu-Pippidi & Ramin Dadašov, *Measuring Control of Corruption by a New Index of Public Integrity*, 22 EUR. J. CRIM. POL'Y & RES. 369, 416, 417 (2016).

110. Transparency International, CPI 2017 TECHNICAL METHODOLOGY NOTE 1 (2017), http://files.transparency.org/content/download/2183/13748/file/CPI_2017_Technical%20Methodology%20Note_EN.pdf.

gues that a “[p]roliferation of indicators . . . springs . . . from the desire for accountability.”¹¹¹ Accepting that accountability is desirable, we must therefore ask whether the non-quantitative human rights mechanisms are any better as monitoring systems.

The first question to consider is whether it is possible to monitor or track human rights through other means. The answer is obviously yes, as it was traditionally done by the different human rights international mechanisms, such as using envoys, commissioners, general reporting, fact finding missions, voluntary transmission of comprehensive reports, sampling, and media tracking. We argue that the issue is not the lack of alternatives for measuring and quantifying a situation, but a cost-benefit analysis of these options in terms of the adequacy of the portrait presented, fairness, costs, accessibility of data, and comparability. This analysis is missing in Merry’s book.

The situation of the human rights mechanisms at the turn of the century described by James Crawford is a perfect illustration of previous attempts to assess human rights compliance. In a well-known text from 2000, Crawford offered a detailed list of problems with the U.N. human rights system, posing to readers the question of whether the system was in crisis. The problems Crawford listed include “corrosive effects” of the backlog in state reporting, linked to delays in present-day reporting, such as a delay between the submission of a report and when it is finally considered;¹¹² resource constraints in terms of personnel, and financial and technological capacities faced by Committees;¹¹³ procedural issues, such as lack of follow-up mechanisms for periodic reports and communications;¹¹⁴ substantive concerns about the composition of committees;¹¹⁵ and limited political support from States.¹¹⁶ Many argue that the introduction of the Universal Periodic Review (UPR) did not adequately address these concerns,¹¹⁷ although

111. MERRY, *supra* note 4, at 3.

112. James Crawford, *The UN Human Right Treaty System: A System in Crisis?*, in *THE FUTURE OF UN HUMAN RIGHTS TREATY MONITORING*, 4–5 (2000).

113. *Id.* at 6–7.

114. *Id.* at 7–8.

115. *Id.* at 9.

116. *Id.* at 10.

117. See generally Allehone Mulugeta Abebe, *Of Shaming and Bargaining: African States and the Universal Periodic Review of the United Nations Human*

UPR has somewhat heightened accountability and recognition of the universality of human rights.¹¹⁸

A bigger dilemma, and the one Merry raises, is the problematic nature of judging compliance based on information provided by the organization being judged. She goes on to highlight the problem of “gaming” the indicator: the ways in which those who are being evaluated present data or change their approach purely to get a better ranking, with no substantial impact or change to the issue at hand. Merry argues: “Rather than revealing truth, indicators create it. However, the result is not simply a fiction but a particular way of dividing up and making known one reality among many possibilities.”¹¹⁹ Data acts to suggest the existence of one particular truth, despite the shortcomings of the data collection itself; it is, therefore, a misleading “truth.” In addition, it then influences development and policy planning.¹²⁰

Merry’s concerns are valid. Data shows only one perspective on a situation, and the concern is heightened because the data collector can present the data in a way that shows it in the best light. When it comes to human rights, however, surely we can shed some of our skepticism. We must recognize those States that genuinely wish to improve the well-being of their residents and who have a real interest in a system of data collection that would enable them to better understand the situations they are grappling with and how they can better invest their own resources to improve enjoyment of human rights. However, no one wants to be seen as a human rights violator, and it is true that some States may abuse quantitative indicators to diminish the problems reflected in their country statistics.

However, in these instances, qualitative data proves similarly problematic when it comes to unsupportive States. Those

Rights Council, 9 HUM. RTS. L. REV. 1, 19-30 (2009); Felice D. Gaer, *A Voice Not an Echo: Universal Periodic Review and the UN Treaty Body System*, 7 HUM. RTS. L. REV. 109, 135-139 (2007).

118. Elvira Domínguez-Redondo, *The Universal Periodic Review—Is There Life Beyond Naming and Shaming in Human Rights Implementation?*, 4 N. Z. L. REV. 67, 694-705 (2012); Rhona Smith, “*To See Themselves as Others See Them*”: *The Five Permanent Members of the Security Council and the Human Rights Council’s Universal Periodic Review*, 35 HUM. RTS. Q. 1, 22-26 (2013).

119. MERRY, *supra* note 4, at 5.

120. *Id.* at 5.

data collectors who are more willing to honestly participate in a qualitative process to monitor their performance are more likely to produce more reliable quantifiable data as well. The difference can probably be seen with those States who are reluctant participants in qualitative evaluation processes of their human rights situations. In a given year, one can mislead others in qualitative and quantitative terms, but when the exercise is repeated year after year, it becomes more difficult to coherently produce false quantitative data.

A separate issue Merry raises—and perhaps a less explored one in the literature on indicators—is the potential dependency of the United Nations on funding by the States subject to evaluation.¹²¹ Yet, one could argue that there are many examples of countries threatening to withdraw funding for the publication of facts or decisions by U.N. bodies that have nothing to do with rankings or quantifications. One can think for example of the finally withdrawn report accusing Israel of imposing an apartheid regime in the Occupied Territories,¹²² or the never-published report on the killing and maiming of children in warfare in Yemen by Saudi Arabia.¹²³ In fact, a more accurate view could be the exact opposite situation. In general, countries are more willing to be subjected to global exercises where there is some naming and shaming for everybody than particular investigations into any one pinpointed domestic context. The very existence of the Universal Periodic Review (UPR)—and the debates leading to its creation—would support this view.¹²⁴

121. *Id.* at 168.

122. See Raoul Wootliff, *End Bias or Else, Israel Threatens UN Ahead of Chief's Visit*, *TIMES ISR.* (Aug. 27, 2017), <https://www.timesofisrael.com/end-bias-or-else-israel-threatens-un-ahead-of-chiefs-visit/>.

123. Colum Lynch, *Saudi Arabia Threatened to Break Relations With U.N. over Human Rights Criticism in Yemen*, *FOREIGN POL'Y* (June 7, 2015), <https://foreignpolicy.com/2016/06/07/saudi-arabia-threatened-to-break-relations-with-un-over-human-rights-criticism-in-yemen/>.

124. This point is expressly made in the Preamble of the Resolution adopted by the General Assembly which created the Human Rights and the UPR as “universality” and “non-selectivity”, as a recognition of the focus on a few situations and countries in the work of the Commission on Human Rights, the predecessor of the Human Rights Council. G.A. Res. 60/251, Preamble, U.N. Doc. A/RES/60/251 (Apr. 3, 2006), <https://undocs.org/en/A/RES/60/251>

VII. ATTACHING CONSEQUENCES TO INDICATORS

Many of the criticisms Merry highlights are not particular to indicators but relate to the international system we have created to protect human rights. Perhaps with the exception of some regional systems—namely the European Court of Human Rights and the Inter-American Court of Human Rights—treaty bodies using indicators lack penalties or rewards for good or bad performance beyond issuing public comments of praise or critique.¹²⁵ The same could be said for those that do not use indicators. The U.N. Trafficking Protocol for instance was enacted in Palermo in 2000,¹²⁶ and having defined the concept of trafficking,¹²⁷ was arguably the origin of the U.S. TIP report.¹²⁸ The Protocol itself suffers from a weak implementation mechanism, operating under a “loose” oversight of a working group of States Parties attached to the broader Conference on Parties to the U.N. Convention against Transnational Organized Crime.¹²⁹ In other words, the lack of consequences for non-compliance is a problem stemming from the broader system of international human rights law, or international law, and cannot be solely attributed to indicators.

It is true as well that efforts to promote indicators of human rights compliance face resistance from countries concerned about mechanisms that compare and rank them on the basis of their compliance.¹³⁰ And yet, this is no different from those reporting mechanisms that do not use indicators or collect aggregated data. For example, country-specific reports on their fulfilment of conventions against torture often remained

125. MERRY, *supra* note 4, at 167.

126. Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the United Nations Convention Against Transnational Organized Crime, *opened for signature* Nov. 15, 2000, 2237 U.N.T.S. 319.

127. *Id.* art. 3(a).

128. U.S. DEP’T OF STATE, TRAFFICKING IN PERSONS REPORT: JULY 2001, 8 (2001), <https://www.state.gov/documents/organization/4107.pdf>.

129. Anne T. Gallagher, *Two Cheers for the Trafficking Protocol*, 4 ANTI-TRAFFICKING REV. 14, 22 (2015).

130. MERRY, *supra* note 4, at 167.

unpublished due to the reluctance of those Governments to accept them and make them public.¹³¹

To the contrary, indicators can offer individual states a baseline. Current performance can be checked against previous performance, based on general agreed-upon standards. This would allow countries to be judged only against themselves, eliminating the common claim that developing countries are judged against developed countries' standards. Rewarding improvements and punishing deteriorations would then seem fairer than attaching those consequences to reporting on concrete incidents that may not be comparable. This could also be a way of implementing the evolving and improving nature of human rights,¹³² making it obvious if a country goes backwards in its protection of human rights.

All of this is a different scenario from the one Merry describes on the consequences that the U.S. Government attaches to its TIP rankings. In fact, the U.S. Government—like all governments—is entitled to spend its donor aid as it considers appropriate.¹³³ If the government feels more confident to do this based on a report that it has produced itself, there is no legitimate reason to stand in the way of such an approach. Any country could do the same, basing its decisions on quantitative or qualitative appraisals conducted by them or on their behalf. All this proves that this issue is not linked to quantification mechanisms or rankings, but rather the gathering of data in order to make certain decisions which are political in nature. It is the politics behind the decision that should be the subject of the critique.

It is also important to note that governments and fellow States Parties are far from the only bodies that use data as the

131. For example, the Assembly of the Council of Europe has recently called on the European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment (CPT) to make public any unpublished reports concerning its member States. Eur. Consult. Ass., *Recommendation 2123 (2018) Provisional Version: Strengthening International Regulations against Trade in Goods used for Torture and the Death Penalty*, Doc. No. 14454, art. 10(6) (2018).

132. See generally GENDER, NUTRITION, AND THE HUMAN RIGHT TO ADEQUATE FOOD: TOWARD AN INCLUSIVE FRAMEWORK (Anne C. Bellows et al. eds., 2016) (arguing for a more expansive notion of human rights encompassing an affirmative right to adequate food).

133. For example, the United States funds abolitionists groups, but not NGOs favouring the legalization of sex work. MERRY, *supra* note 4, at 153.

basis for establishing consequences for other entities. The media is also a massive consumer of data. In this respect, it is important to distinguish between how governments or the United Nations, for instance, may attach consequences to a particular system of ranking or performance based on quantitative data and the way in which the media may consume and report on such data. Merry fails to adequately recognize such a distinction.

For example, when she discusses the quantification of corruption, she refers to the use of “expert opinion surveys,” correctly arguing that collecting data is expensive and expert opinion surveys are a cheaper alternative.¹³⁴ In this case, she questions the idea of sending questionnaires to local experts to gather their experiences of corruption, arguing that while this is clearly less expensive, it is “also less comprehensive and accurate.”¹³⁵ However, Merry does not accurately represent this example. The index she refers to measures “perception of corruption” and not “corruption.”¹³⁶ The methodology is designed to survey people on their perception of corruption and Transparency International (TI), the entity which created that index, presents the results as such. In fact, TI has an alternative index—the Global Corruption Barometer—which surveys people about their corruption experiences in the last twelve months,¹³⁷ trying to *directly* assess the prevalence of corruption in a given jurisdiction. This allows for data on corruption prevalence to exist alongside data on *perceptions* of the levels of corruption.

The real issue here is how the media—and particularly non-expert media—reports on data developed from an index versus what the actual index is purporting to measure. This problem would be magnified if the reporting of those indices and quantification formed the basis of policy decisions or was used to apply the “evidence-based governance” that Merry de-

134. MERRY, *supra* note 4, at 7.

135. *Id.*

136. *Corruption Perceptions Index 2016*, TRANSPARENCY INT’L (Jan. 2017), https://www.transparency.org/news/feature/corruption_perceptions_index_2016.

137. *People and Corruption: Citizens’ Voices from Around the World*, TRANSPARENCY INT’L (Nov. 14, 2017), https://www.transparency.org/_view/publication/8064.

scribes at the beginning of her book.¹³⁸ Yet, Merry fails to provide an example of such a problem in practice.

VIII. CONCLUSION

As a general observation, quantification is not intrinsically better or worse than other systems of accountability in assessing compliance with human rights, but it does require a more robust methodology than other mechanisms to be useful. Here, the term “useful” denotes both a process and an end result that is a fair representation of a body of data and where that data can be presented in a manageable format, allowing consumers of that data to assess a situation at a particular moment in time but also changes over time. Such rigor may not be as demanding for some non-quantitative approaches.

Yet, there are some obvious advantages of quantitative measurement. Merry is correct that it cannot all be quantitative, but this does not justify a widespread questioning of the substantial interest in the quantification of human rights indicators that exists today. There are bad apples and good apples, including within Merry’s examples. In the same way that the TIP report is an example of a bad apple, the outputs of the UNECE Project,¹³⁹ the UNSC Project,¹⁴⁰ and the work of Yakin Ertürk¹⁴¹ in the context of violence against women are examples of good apples in terms of how they have defined the issue to be measured and against which governments and other actors will be held to account. The very process of developing such indicators involved an open and honest debate about what would get measured as part of a reasonably global effort towards heightened accountability in relation to violence suffered by women.¹⁴²

138. MERRY, *supra* note 4, at 10.

139. *Id.* at 65.

140. *Id.* at 54.

141. *Id.* at 63, 64.

142. *Id.* at 63-66. Merry’s analysis sheds positive light on the process of indicator development under the U.N. bodies for European (UNECE) initiative, one that “relief on social science scholars with extensive knowledge of gender-based violence and survey techniques developed for measuring it,” was clearly “concerned with women who experience violence” and adopted “a more feminist political stance and an interest in maximizing disclosure.” *Id.* at 67.

As a second conclusion, Merry is also correct in stating that, “some human rights are more amenable to quantification than others,”¹⁴³ but this does not mean that quantification should be excluded from monitoring compliance for certain rights. In those cases, the methodology should be defined to align as closely as possible with the specific right. When proxies are used, these choices should be transparent and any necessary simplification clearly understood by those collecting the data and later by those using it.

According to Merry, proxy measures, giving the example, among others, of using investment in prison and police reform to assess compliance with civil and political rights, “provide far greater specificity to treaty obligations but only focus on a few measures while neglecting many other dimensions of a given right.”¹⁴⁴ This is true, but needs to be compared with the alternative systems available to monitor those “given rights,” especially in terms of costs and benefits. A thorough and systematic reporting on the use of unnecessary force and torture by police officials would be a much better way to track compliance with the relevant articles of the Convention Against Torture, but it is unlikely to be feasible in the majority of developing countries, nor possible in the majority of countries globally on a systematic and regular basis, which is what is required for ongoing accountability.

In connection to this point, we disagree with Merry when she argues that “treaty bodies see their task as evaluating compliance with legal standards based on the data presented, not on checking the quality of the data themselves.”¹⁴⁵ This may be the case in some treaty monitoring bodies, but it is definitely not the case for some others. In fact, in some instances, treaty bodies concluding comments address very specifically the process of collecting information and contrast the data presented in States Parties’ reports with submissions and reports from NGOs, local and international, which allow members to challenge the data with concrete examples.¹⁴⁶ It is true

143. *Id.* at 172.

144. *Id.* at 171.

145. *Id.* at 170.

146. Cossette D. Creamer & Beth A. Simmons, *Ratification, Reporting, and Rights: Quality of Participation in the Convention Against Torture*, 37 HUMAN RIGHTS Q. 579, 585 (2015).

that “[b]y using apparently objective, scientifically produced statistical data, treaty bodies can enhance their authority in rendering judgments about human rights,”¹⁴⁷ but as far as there has been an adequate check of the underlying data, we do not perceive this as a problem but rather as a legitimate option in light of the alternative—governments showing unwillingness to follow concluding comments or denying their legitimacy because such data is lacking.

Finally, Merry does not sufficiently explore the difference between quantification and how the media reports on organized data. At some point, it is necessary to distinguish between the use of quantification by experts in the field of international human rights law and its use by the general public. Therefore, concerns with what and how something is measured must be separated from concerns about how such data is interpreted and consumed by end users.

Indicators may appear objective from a distance, but when we look at them closely, it is easy to see their flaws. Merry’s book can be praised as an effort to make this clear to non-experts who come across her work. Regretfully, this point is well-known to specialized practitioners who value indicators for their practicality in helping to measure similarities and differences, advancement, and gaps in our progress towards development and well-being. Merry has evidently invested an enormous amount of work in analyzing the three indicator projects that are the subject of *The Seductions of Quantification*. As academics and practitioners in international law, we would have hoped for a deeper study of the trade-offs between the different alternatives (i.e., the non-quantitative ones) to assessing compliance with human rights and, potentially, a practical insight on more robust and useful ways to select and develop indicators.

147. MERRY, *supra* note 4, at 170.