

U.S. Department of Labor
Bureau of International Labor Affairs

Office of Trade and Labor Affairs
Contract Research Program

Download this and other papers at
<http://www.dol.gov/ilab/media/reports/otla/>

The views expressed here are those of the author(s) and do not necessarily represent the views or official positions of the U.S. Government or the U.S. Department of Labor.

EXECUTIVE SUMMARY

Formulating and Aggregating Indicators of Labor Rights Compliance

for

Research project: Refining the NAS Matrix

DOL099RP20744

Submitted to

Bureau of International Labor Affairs
Department of Labor

March 10, 2011

Mark Barenberg
Professor of Law
Columbia University
New York

Executive Summary

The Research Objective: Formulating Indicators for International Labor Rights

This paper formulates three bodies of Indicators to measure the compliance of U.S. trading partners with the labor rights provisions of U.S. trade legislation and trade agreements.¹ Two bodies of Indicators serve the purpose of preliminary screening of countries; the third body is for comprehensive evaluation.² The paper also canvasses alternative strategies for weighting and aggregating Indicators,³ and proposes a process for the Bureau of International Labor Affairs (ILAB) to apply Indicators, create new sub-Indicators, and accumulate country-specific data sources in successive rounds of appraisal.

The ultimate authority to make determinations of trading partners' compliance or noncompliance is vested in the United State Trade Representative (USTR) and the President. ILAB may apply the three bodies of Indicators to inform its own decisions about allocating resources for technical assistance and for research, and to inform the USTR's and President's determinations under the trade legislation and agreements.

The three bodies of Indicators rest on eleven proposed revisions of the Indicator methodology developed for ILAB by the National Academy of Sciences (NAS) in 2004.⁴ This Executive Summary summarizes the NAS methodology, sets out the eleven proposed revisions, and concludes with a Flow Chart showing how each Indicator is constructed using the revised methodology.

¹ The labor rights treated in the paper are: (1) freedom of association, rights to organize, and rights to bargain collectively; (2) rights against employment discrimination; and (3) acceptable conditions of work with respect to minimum wages, hours of work, and occupational safety and health. U.S. trade legislation and treaties also include rights against forced labor and child labor, but these two rights are not within the scope of the research proposal.

² The three bodies of Indicators are set out in Appendices A, B, and C to the paper, along with annotations explaining individual Indicators.

³ In order to lay the groundwork for creating valid composite Indicators, the paper undertakes extended conceptual analysis of the different types of Indicators that are optimal for measurement of various aspects of the labor rights in question. The attached Flow Chart shows the relationships among the types of Indicators.

⁴ Since the policy goal is to measure compliance with legal instruments (legislation and treaties), the paper's core methodology is legal and regulatory analysis.

The NAS Methodology and Pilot-Testing

The NAS methodology has three components. First, the NAS formulated three categories of Indicators: Legal Framework, Government Performance, and Overall Outcomes. There are between 5 and 20 Indicators in each category for each labor right. Second, the NAS constructed a 3 by 3 matrix for scoring each Indicator for a given country. One axis of the matrix measures the degree of the “problem” with the trading partner’s compliance with the Indicator: “some problems,” “more extensive problems,” or “severe problems.” The other axis measures the direction of change in compliance: “improving,” “steady state,” or “worsening.” Third, the NAS constructed an online database for analysts to use in scoring the Indicators.

The NAS methodology was pilot-tested in 2009 by social scientists at the University of Michigan. The pilot test showed substantial grounds for refining the methodology. There were very high rates of variance and non-assessment among analysts in the scoring of particular Indicators. The analysts attributed the variance to many factors, especially the difficulty in interpreting ambiguous Indicators, inconsistent terminology, the lack of guidance in determining whether problems were “more extensive” or “severe,”⁵ and limitations in the online database.⁶

FIRST PROPOSAL: Three Bodies of Indicators

The first of the eleven proposed revisions has already been mentioned: In addition to revising the body of Assessment Indicators formulated by the NAS, the paper constructs two new bodies of Indicators.

The two new bodies – Probative Indicators and Diagnostic Indicators – identify trading partners whose compliance records are likely to be very weak. These two bodies therefore

⁵ In addition, many Indicators are double-barreled (asking for multiple responses to a single Indicator) and binary (asking for yes/no responses rather than graduated, three-tier responses). The pilot analysts therefore created a *de facto* 4 by 4 matrix, including the category of “non-assessment” along both axes.

⁶ Searches by country and Indicator typically direct analysts to the home pages of other websites.

comprise triage Indicators, prioritizing countries for comprehensive assessment, for technical assistance, and for specific research initiatives. The Probative Indicators are a short list for initial screening. The Diagnostic Indicators are a somewhat longer list for more intensive screening.

Since Probative and Diagnostic Indicators are not evaluative, they need not be comprehensive nor meet strict standards of statistical validity. As befits their purpose, they are more idiosyncratic and probing. They include direct and indirect measures, relying on reasonably reliable data that are available in a data-poor environment. Some salient violations – such as failures to prosecute known murderers of trade unionists – fall into this category. But such measures may yield false negatives. There are many countries, for example, that pervasively deny workers’ right to organize in ways other than severe violence. Hence, the Probative and Diagnostic Indicators, like their medical namesakes, also test for less salient symptoms of deeper pathologies, such as a high proportion of employment contracts that are temporary – indicating a large informal sector, diminished enforceability of wages, hours, and safety and health, and heightened vulnerability of union supporters.

The third body of Indicators – Assessment Indicators – is for comprehensive evaluation of the government’s compliance. The revised Assessment Indicators (1) are relevant to the labor rights provisions in legislation and treaties, (2) precisely define those legal rights, (3) use consistent terminology throughout the body of Indicators, and (4) systematically follow the hierarchies in labor law to ensure that all important aspects of rights are covered.

SECOND PROPOSAL: Capacity-Building Indicators

Second, the paper proposes a new category of Indicators applicable to each labor right, called Capacity-Building Indicators. For clarification, it also renames the other three categories: Substantive Law Indicators, Enforcement Indicators, and Outcome Indicators.

The Capacity-Building Indicators measure four sub-categories of government activity: (1) systems for collecting and publishing data; (2) the government’s formulation of targets for

improved compliance and the government’s self-evaluation in meeting those targets; (3) the participation of stakeholders in formulating substantive law, in enforcement, in collecting data, and in setting and evaluating targets for improvement; and (4) periodic meetings with peer governments and other international actors to compare the government’s performance with, and to learn from, other governments’ experience.

There are three strong reasons for adding Capacity-Building Indicators. First, international law requires governments to undertake the four activities just mentioned. Second, the terms of U.S. trade legislation match the international standards that impose capacity-building requirements. And, third, collecting data, setting targets, evaluating performance, and securing accountability are elemental components of ensuring that governments have the means and incentive to improve enforcement.

THIRD PROPOSAL: All Indicators are in Binary Form

The third proposal is to reframe all Indicators in the binary (yes/no) form, replacing the 3 by 3 gradations of the NAS matrix. The pilot analysts had difficulty applying the 3 by 3 matrix, which provides no metric for scoring the gradations.⁷

The binary form has the advantage of simplicity – simplicity in formulating, applying, weighting, and communicating the Indicators. Many Indicators are “naturally” binary, evidenced by the fact that many NAS Indicators are themselves binary and therefore unable to fit the 3 by 3 matrix. An example of a binary Substantive Law Indicator is: “Does the law prohibit employers from providing financial support to labor unions?” A binary Enforcement Indicator is: “In the preceding two years, did the labor tribunals issue written opinions in all but a trivial number of final decisions in cases claiming anti-union retaliation?” A binary Outcome Indicator is: “Does the percentage of informal-sector workers who are female exceed the percentage of formal-sector workers who are female?”

⁷ Instead, the NAS set forth four interacting, subjective criteria for scoring gradations, compounding the ambiguities and inconsistencies in the text of each Indicator.

FOURTH PROPOSAL: Four Types of Legal Norms

The fourth proposal is to frame qualitative Indicators in ways that avoid the characteristic deficiencies in traditional compliance methods, which formulate comprehensive sets of highly detailed, fixed bright-line Indicators. Often, such Indicators are easily “gamed” and not well-adapted to country contexts and to changing production systems. In some settings, more general “standards” are better-suited than bright-line rules; and revisable norms may be better-suited than fixed rules. The paper therefore proposes four types of Indicators, suited to different empirical and normative settings. First, fixed bright-line Indicators are optimal where both values and facts are predictable and easily specified, such as the baseline of a 48-hour work week. Second, revisable bright-line Indicators are optimal where values are fixed but factual contexts, though presently specifiable, are likely to change. An Indicator requiring coal mines to provide a gas mask that ensures one hour of protection against carbon monoxide is an example of this type. Third, fixed standards may be optimal where values are clear but the myriad factual contexts to which they apply cannot be specified *ex ante*. An Indicator prohibiting sexual harassment is such a case. Fourth, revisable standards are optimal where both values and facts are in flux. An Indicator prohibiting “psychological coercion” is an example. Only in recent years has that concept become recognized in the law of coercive labor.

FIFTH PROPOSAL: Each Indicator has a Country-Specific Drop-Down Window

The fifth proposal addresses ILAB’s process for revising the body of Indicators and for visualizing and storing country-level data. Whether the Indicator is framed as a bright-line rule or a standard, each Indicator would appear as the heading of a country-specific drop-down window. In each round of assessment or screening, the ILAB analyst would enter in the window new “sub-Indicators” that capture country-specific problems and country-specific data sources related to the window’s heading. The window would also show sub-Indicators that are common to all countries’ compliance with the window’s heading.

SIXTH PROPOSAL: Use of Outcome Indicators for Only Three Purposes

The sixth proposal addresses the relationship between output measures (Outcome Indicators) and input measures (Substantive Law, Enforcement, and Capacity-Building Indicators). The U.S. government's primary concern is to create incentives for governments both to improve the use of enforcement instruments within their control and to produce accurate data about such efforts. If analysts verify enforcement efforts by outcome measures (that is, measures of employer compliance), then the trading partner's incentive to produce data directly measuring enforcement inputs is diminished. In addition, there is a risk of double-counting if the methodology aggregates Indicators that measure both the cause (inputs) and the effects of that cause (outputs).

The bodies of Indicators therefore use outcome measures in only three ways: First, outcome measures are used in Probative and Diagnostic Indicators, which do not call for the same statistical validation as Assessment Indicators and which are applied to information-poor settings in which indirect measures may be necessary and useful. Second, outcome measures are used, not as evidence of inputs, but as elements in the definition of input Indicators. For example, Indicators ask about the marginal decrease in violations (outputs) relative to increased enforcement efforts (inputs). Third, Outcome Indicators are included in the bodies of Indicators, not for assessment purposes, but to facilitate subsequent ILAB research on the question of which input Indicators are effective in achieving outcomes.

SEVENTH PROPOSAL: Normative Metrics, including Longitudinal and Comparative Indicators

Every Indicator must incorporate a normative (values-based) metric, to tell the analyst whether an answer of "yes" indicates positive or negative government performance.

Many Indicators use qualitative metrics. For example, "Does the law prohibit anti-union retaliation?"

Other Indicators use quantitative metrics, including (1) absolute numbers (such as

the legal stipulation of a 48-hour work week), (2) zero-tolerance norms (such as the number of murders of trade unionists that the government failed to investigate), (3) “non-trivial numbers of violations” for nominally absolute requirements, when it would be unreasonable to expect zero violations (such as the number of times a labor tribunal failed to publicly issue a final decision in writing), (4) longitudinal rates of improvement in government performance, and (5) comparison of the government’s performance to that of peer countries. Longitudinal and comparative Indicators are highly reasonable implementations of Congressional intent, when international standards provide no clear qualitative or quantitative yardstick. Longitudinal Indicators measure whether the government is adequately “taking steps” to improve compliance, relative to a baseline of the government’s previous performance. The comparative baseline has a similar advantage of creating incentives for governments to continuously improve, to exceed the performance of others in their peer group (the “race to the top”). In addition, comparative baselines intrinsically reflect international norms embodied in actual cross-country practices.

EIGHTH PROPOSAL: Only One Control Variable – Income per Capita

The eighth proposal addresses the problem of cross-country context. Should we “adjust” either individual or composite Indicators based on such control variables as the country’s level of development, geographic region, political regime type, labor relations system, government capacity, colonial history, or legal origins (e.g., civil vs. common law)?

This paper proposes that we build a minimum of control variables into the definition of the Indicators. Indeed, the proposed Indicators contain only a single control variable: level of economic development, proxied by the country’s quintile of real income per capita. U.S. trade legislation explicitly requires the USTR and the President to control for economic development in their application of internationally recognized worker rights. Why no other control variables? Some constraints that we might take *a priori* as hard constraints may actually be soft constraints – the very factors that we want the government to target with its policy instruments, such as government capacity. As ILAB gathers and analyzes data in

successive rounds of diagnosis and evaluation, it can make judgments about whether to add new control variables or revise old ones.

NINTH PROPOSAL: Incentivizing Data Production via Twin Indicators and Capacity-Building Indicators

The ninth proposal addresses the very serious problem that pertinent data may be unavailable, unreliable, or inconsistent. Rejecting the common response of deleting or downgrading Indicators for which data are scarce, the paper instead proposes explicit, high-powered incentives for data production – on the premise that rich, accurate data are essential to providing the capacity and the incentive for improved enforcement. The paper proposes two means to serve that goal. First, as noted above, Capacity-Building Indicators measure the government’s implementation of *systems* to collect accurate data.

Second, “Twin Indicators” measure the government’s collection of data as to *particular* individual Indicators. As to subject matters for which data collection is especially urgent, there are two Indicators. The first in the pair of Indicators asks whether the government meets the particular substantive standard. (For example, “In the preceding five years, did the government increase the number of labor inspectors per worker at a rate that exceeded the average rate among countries in the same quintile of real income per capita?”) The second Indicator in the pair asks whether the government has convincingly and verifiably demonstrated that it meets the particular substantive standard. (To continue the example: “Has the government convincingly and verifiably demonstrated that, in the preceding five years, it increased the number of labor inspectors per worker at a rate that exceeded the average rate among countries in the same quintile of real income per capita?”) If the government fails to provide convincing data about the substantive subject, it flunks the second Indicator and risks flunking the first Indicator as well (although there is some chance that analysts will independently find sufficient evidence to score the first Indicator positively).

TENTH PROPOSAL: Five Alternative Strategies for Weighting and Aggregating Indicators

The tenth proposal – or, more accurately, analysis – examines five alternative strategies for weighting and aggregating Assessment Indicators: equal weighting; weighting based on the prioritization embedded in the hierarchies of labor law; principal component analysis; *ex ante* multivariate weighting of variables; and *ex post* modeling of variables gathered in successive rounds of screening and evaluation. Equal weighting has the advantage of simplicity, comprehensibility and, perhaps more important, grounding in the co-equal significance of all key aspects of non-derogable, universal rights. Prioritization based on legal hierarchies has similar advantages, apart from a loss of simplicity. Principal component analysis – which weights Indicators based on their contribution to variance among countries, rather than on their intrinsic importance – is not appropriate for measuring legal rights. Multivariate analysis – whether *ex ante* or *ex post* – is not promising in an environment of such economic and political complexity and data scarcity.

ELEVENTH PROPOSAL: An Overarching Process of Iterative Dialogue

The paper proposes that ILAB implement the Indicator methodology through a process of “iterative dialogue.” During each round of applying Indicators, the U.S. government may provide its provisional appraisal to the trading partner and give the latter an opportunity to offer data and argumentation to explain or justify its provisionally poor score. Based on the new data and explanations, U.S. analysts might or might not revise their appraisal, might or might not revise the sub-Indicators in the country-specific drop-down window, and might or might not gain new insight into potential control variables. At this stage, the dialogue between the two governments will be non-transparent, to avert grandstanding and posturing. The process will not be wholly “undemocratic,” however, since the U.S. analysts will also hold “hub and spoke” dialogues with relevant stakeholders – again non-transparent, for the same reasons. After these non-transparent dialogues, a round of transparent public hearings may be held before the U.S. government arrives at its final determination.

CONSTRUCTING A SINGLE INDICATOR: An Example, and a Flow Chart

Figure 1 below is a Flow Chart showing the steps in formulating each Indicator, for all three bodies of Indicators set out in Appendices A, B, and C – that is, Probative Indicators, Diagnostic Indicators, and Assessment Indicators. (Note that every Indicator is in binary form; that feature of the Indicators is therefore not mentioned in the Flow Chart.) For example, consider the following Indicator:

In the preceding two years, did the tribunals that hear cases on employment discrimination publish their final decisions in writing, in all but a trivial number of cases?

Following the Flow Chart in Figure 1, this Indicator was constructed as follows:

Step One: This is an **Assessment** Indicator, since it is intended as one Indicator in the body of comprehensive Indicators used to evaluate whether the country is in full compliance with its obligations under U.S. legislation and treaties. The treaties specifically require tribunals to publish final decisions in writing.

Step Two: This is an Indicator that measures compliance with **rights against employment discrimination**.

Step Three: This is an **Enforcement Indicator**, since it measures the actual functioning of labor tribunals, which are enforcement institutions.

Step Four: This is an **input** Indicator rather than an output measure, since it measures government effort.

Step Five: This is a **fixed bright-line rule**, since it defines the government's obligation with precision – it must publish final decisions in writing – and there is no reason to think that changes in facts or values will alter the international consensus that final decisions must be issued publicly and in writing.

Step Six: This is a **quantitative norm**, using the threshold of “**non-trivial number of violations**,” appropriate to rules or standards imposing requirements that are absolute but that nonetheless cannot reasonably be expected to be satisfied in every instance.

FIGURE 1. How to Formulate an Indicator: A Flow Chart

STEP ONE: Is the Indicator a **PROBATIVE INDICATOR**, a **DIAGNOSTIC INDICATOR**, or an **ASSESSMENT INDICATOR**?

1. If ILAB's purpose is the high-stakes **EVALUATION** of country compliance, then the Indicator is an **ASSESSMENT INDICATOR**.
2. If ILAB's purpose is to **IDENTIFY** countries that likely have **ESPECIALLY POOR COMPLIANCE**, calling for further research, technical assistance, or comprehensive assessment, then the Indicator falls in a short-list of **PROBATIVE INDICATORS** or a long-list of **DIAGNOSTIC INDICATORS**.

STEP TWO: Which **LABOR RIGHT** does the Indicator measure?

1. An Indicator for **FREEDOM OF ASSOCIATION, RIGHTS TO ORGANIZE, and RIGHTS TO BARGAIN COLLECTIVELY**.
2. An Indicator for **RIGHTS AGAINST EMPLOYMENT DISCRIMINATION**.
3. An Indicator for **ACCEPTABLE CONDITIONS** with respect to **MINIMUM WAGES, HOURS OF WORK, and OCCUPATIONAL SAFETY AND HEALTH**.

STEP THREE: Which **PHASE OF COMPLIANCE** does the Indicator measure?

1. **SUBSTANTIVE LAW INDICATOR** for measuring law on the books.
2. **ENFORCEMENT INDICATOR** for measuring institutions, resources, procedures, and the government's actual practices in enforcing law on the books.
3. **CAPACITY-BUILDING INDICATOR** for measuring data collection, policy targets, self-evaluation, stakeholder participation, transparency, and comparisons with peers.
4. **OUTCOME INDICATOR** for measuring actual employer compliance.

STEP FOUR: Make **LIMITED** use of **OUTPUT MEASURES** in formulating **ASSESSMENT INDICATORS**.

NB: **INPUT** Indicators include Substantive Law Indicators, Enforcement Indicators, and Capacity-Building Indicators. Outcome Indicators are **OUTPUT** measures. The primary goals of assessment are to measure inputs – that is, the government’s use of policy instruments within its control – and to incentivize collection of data about such inputs. Using Outcome Indicators as proxies for such inputs will reduce the incentive to collect data about inputs. These rationales are much weaker for Probative and Diagnostic Indicators.

1. Some **INPUT INDICATORS** can be **DEFINED** by the relation between inputs and outputs. For example: “In the preceding five years, what was the ratio of increased employer compliance with minimum wage laws (output) and increased number of labor inspectors (input)?”
2. Some **OUTCOME INDICATORS** are useful for **ILAB RESEARCH** on the actual impact of government enforcement measures.
3. Some **OUTCOME INDICATORS** are useful as **PROBATIVE** or **DIAGNOSTIC INDICATORS**.

STEP FIVE: Every Indicator is framed as one of the following **FOUR “NORM TYPES”**:

NB: An Indicator is a bright-line rule if it defines in detail the precise facts measured by the Indicator. An Indicator is a standard if it states a general principle or policy that the analyst must apply to complex facts that cannot be pre-specified.

1. **FIXED BRIGHT-LINE RULE**, for contexts where facts and values are static.
2. **FIXED STANDARD**, for contexts where facts are complex and values are static.
3. **REVISABLE BRIGHT-LINE RULE**, for contexts where facts are likely to change but values are fixed.
4. **REVISABLE STANDARD**, for contexts where facts and values are fluid.

STEP SIX: Every Indicator must contain one of these **NORMATIVE METRICS**.

NB: **NORMATIVE METRICS** tell analysts whether a “yes” score for an Indicator indicates positive or negative performance by the government.

1. **QUALITATIVE NORM**, asking about the existence of some rule of law, institution, procedure, or practice. For example: “Does the law prohibit the discharge of workers for anti-union reasons?”
2. **QUANTITATIVE NORM**, including:
 - a. **ABSOLUTE NUMBER**. For example: “Does the law require employers to pay at least 125 percent of the regular hourly wage for hours worked in excess of 48 per week?”
 - b. **ZERO TOLERANCE NORM**. For example: “In the preceding year, did the government fail to prosecute any known murderer(s) of trade union supporters?”
 - c. **“TRIVIAL” NON-COMPLIANCE**. For example: “In the preceding year, were all but a trivial number of evidentiary hearings in labor cases open to the public?”
 - d. **LONGITUDINAL METRIC**. For example: “Did the number of labor inspectors increase by at least 5 percent per year in the last five years?”
 - e. **COMPARATIVE METRIC**. For example: “Does the labor administration budget per worker exceed the average among countries in the same quintile of real income per capita?”

STEP SEVEN: Every Indicator is the **HEADING** of a **DROP-DOWN WINDOW** for **EACH COUNTRY**.

The window will show **SUB-INDICATORS**, some of which are country-specific, some of which are common to all countries. The window will also show **DATA SOURCES FOR EACH COUNTRY**. ILAB analysts will enter new sub-Indicators and data sources as they apply Indicators to each country.