Executive Summary

Doing Business (DB), an annual World Bank-IFC publication launched in 2004, is one of the Bank Group's flagship knowledge products. It measures the burden of selected business regulations in 178 countries and ranks the countries on 10 dimensions. The program's stated objective is to advance the World Bank Group's private sector development agenda in four ways: motivate reforms through country benchmarking; inform the design of reforms; enrich international initiatives on development effectiveness; and inform theory.

This independent evaluation of the DB indicators assesses the methods and processes underlying the construction of the indicators; the relevance of the indicators to desired intermediate outcomes; and their use by World Bank Group staff, policy makers, and other stakeholders. It finds that the indicators have been highly effective in drawing attention to the burdens of business regulation, but cannot by themselves capture other key dimensions of a country's business climate, the benefits of regulation, or key related aspects of development effectiveness. Thus, the Bank Group and stakeholders need to consider the DB indicators in a country context and interpret them accordingly.

The Underlying Framework of the DB Indicators

The DB exercise is anchored in research that links characteristics of a country's business environment to firm performance, and thence to macroeconomic outcomes. The regulatory framework—the part of the business environment that DB measures—has been shown to be associated with firm performance, but its association with macroeconomic outcomes is less clear. Many other factors affect macroeconomic outcomes, and the direction of causality between regulation and economic outcomes is very difficult to isolate. Since regulations generate social benefits as well as private costs, what is good for an individual firm is not necessarily good for the economy or society as a whole. Therefore, policy implications are not always clear-cut, and the right level and type of regulation is a matter of policy choice in each country.

The DB exercise reflects the limitations inherent in the underlying research. As an exercise in cross-country comparison, DB is not intended to, and cannot, capture country nuances. Firms' investment decisions also depend on variables not measured by the DB indicators, such as the cost and access to finance and infrastructure, labor skills, and corruption. Different aspects of regulation have varying degrees of economic importance depending on countries' income levels, legal regimes, and other characteristics. Seven of DB's 10 indicators presume that lessening regulation is always desirable, whether a country starts with a little or a lot of regulation. Reform as measured by the DB indicators typically means reducing regulations and their burden, irrespective of their potential benefits.

The evaluation confirmed that the DB indicators primarily measure laws and regulations as they are written. But the relevance of each indicator in a given country depends on the extent to which the law is actually applied, which DB does not aim to measure. Likewise, the pay-off of a particular regulatory reform will depend on how significant a burden the regulation poses in practice. These limitations underscore the need for DB to be interpreted cautiously and used in conjunction with complementary tools such as Investment Climate Assessments.

Overall, the indicators objectively and reliably measure what they set out to measure, with a few qualifications. The controversial *employing workers* indicator is consistent with the letter of relevant International Labor Organization (ILO) conventions, but not always their spirit, insofar as it gives lower scores to countries that have chosen policies for greater job protection. Systematic differences in the country rankings for a few indicators are associated with countries' legal origins in civil or common law, but these patterns have little impact on the overall rankings or the validity of the exercise. The *paying taxes* indicator includes an anomalous subindicator—the *total tax rate*—which does not simply measure administrative burden to firms, but rather reflects a country's overall fiscal policy derived from social preferences. Finally, inaccurate nomenclature and overstated claims of the indicators' explanatory power have provoked considerable criticism from stakeholders.

Methodology and Data Reliability

DB collects its information from expert informants in each country, mostly lawyers, who provide information free of charge. This process can generate reliable data, but three areas of vulnerability need to be addressed.

First, the data are provided by few informants, with some data points for a country generated by just one or two firms. Of particular concern is the *paying taxes* indicator—DB relies exclusively on a single firm to provide both the underlying methodology and the data for 142 countries. The number and diversity of informants for all indicators need to be increased and their information validated more systematically. An increase in the informant base will require a systematic vetting process to reduce self-selection bias. Simplifying the questionnaire may also help to encourage more informants to contribute.

Second, although DB makes available a great deal of information about its data and methods, it remains insufficiently transparent about the number and types of informants for each indicator, the adjustments its staff make to the data received from informants, and the changes made to previously published data and their effects on the rankings. DB needs to adequately explain to users the possibilities for errors and biases.

Third, DB makes much of its country rankings. The rankings entail three weaknesses. First, because most of the indicators presume that less regulation is better, it is difficult to tell whether the top-ranked countries have good and efficient regulations or simply inadequate regulation. Second, the small informant base makes it difficult to measure confidence in the accuracy of the individual indicator values, and thus in the aggregate rankings. Third, changes in a country's ranking depend importantly on where it sits on the distribution: small changes can produce large ratings jumps, and vice versa. These factors contribute to anomalies in the rankings.

These issues alone may not jeopardize the DB indicators' reliability. But the lack of transparency about them undermines DB's credibility and goodwill. DB's documents and presentations should include full explanations and cautions on these points.

Motivating and Designing Reforms

The DB indicators have motivated policy makers to discuss and consider business regulation issues. Its active dissemination in easy-to-understand language permits widespread press coverage and generates interest from businesses, nongovernmental organizations (NGOs), and senior policy makers.

DB has had less influence on the choice, scope, and design of reforms. Most Bank Group staff and country stakeholders interviewed for this evaluation report that they draw on a range of analytical material to determine the nature, sequence, and direction of reforms; the DB indicators have limited use in this regard. As a cross-country benchmarking exercise, DB cannot be expected to capture the country-specific considerations involved in prioritizing, sequencing, and designing policy reforms. Each year DB spotlights countries that have demonstrated the largest gain in the overall ranking and an improvement on at least three indicators. Such an approach, while transparent, does not capture the reforms' relevance and their potential impact on the binding constraints to the investment climate in the country.

IEG did not find evidence that the DB indicators have distorted policy priorities in the countries or in the Bank Group's programs, or that countries have made superficial changes for the sole purpose of improving their rankings.

In summary, DB measures the costs but not the benefits of regulation. Despite its methodological limitations, it has contributed to development by providing countries with a basis for international comparisons of their regulatory regimes. It has helped to catalyze debates and dialogue about investment climate issues in developing countries. For the Bank Group, it is a key global knowledge product. Most of the methodological limitations can and should be addressed promptly, lest they undermine its credibility. Inaccurate nomenclature should be rectified and the DB reports should not overstate claims of causality and the indicators' explanatory power.

Implications for the Bank Group

The evaluation notes two broader implications for the Bank Group.

First, the Bank Group, by prominently recognizing DB's highly ranked countries, may inadvertently be signaling that it values reduced regulatory burdens more than other development goals. The Bank Group's approach entails helping countries achieve a wide range of objectives, yet it has no comparable way of celebrating improvements in other important development outcomes.

Second, the DB exercise has demonstrated that cross-country ranking can be effective in spurring dialogue and motivating interest and action. It could potentially be applied to other development issues—those for which actionable indicators can serve as proxies for the target outcomes and for which the direction of improvement is uniform for all countries.

Recommendations

1. **To improve the credibility and quality of the rankings, the DB team should:**
   a. **Take a strategic approach** to selecting and increasing the number of informants:
      – Establish and disclose selection criteria for informants.
      – Focus on the indicators with fewest informants and on countries with the least reliable information.
      – Formalize the contributions of the supplemental informants by having them fill out the questionnaire.
      – Involve Bank Group staff more actively to help identify informants.

   b. **Be more transparent** about the following aspects of the process:
      – *Informant base*: Disclose the number of informants for each indicator at the country level, differentiating between those who complete questionnaires and those who provide supplemental information.
      – *Changes in data*: Disclose all data corrections and changes as they are made. Explain their effect on the rankings, and, to facilitate research, make available all previously published data sets.
      – *Use of the indicators:* Be clear about the limitations in the use of the indicators for a broader policy dialogue on a country's development priorities.

   c. **Revise the *paying taxes* indicator to include only measures of administrative burden.** Since the tax rate is an important part of the business climate, DB should continue to collect and present simple information on corporate tax rates, but exclude it from the rankings (as it does for information on nonwage labor costs in the *employing workers* indicator). A wider range of informants should also be engaged for the *paying taxes* indicator.

2. **To make its reform analysis more meaningful, the DB team should:**
   a. **Make clear that DB measures improvements to regulatory costs and burdens,** which is only one dimension of any overall reform of the investment climate.
   b. **Trace the impact of DB reforms at the country level.** The DB team should work with country units to analyze the effects of implementing the reforms measured by the DB indicators (such as revised legislation or streamlined processes) on: (i) firm performance, (ii) perceptions of business managers on related regulatory burdens, and (iii) the efficiency of the regulatory environment in the country.

3. **To plan future additions or modification to the indicators, the DB team should:**
   a. **Use Bank analyses to drive the choice of DB indicators.** Business Enterprise Surveys, Investment Climate Assessments, and other work can help determine stakeholders' priorities for domestic private sector growth. The DB team should use such analyses to determine the choice of new indicators and periodically reassess its current set of indicators.
   b. **Pilot and stabilize the methodology before including new indicators in rankings.** Frequent changes in methodology make comparison across time less meaningful. New indicators should be piloted (that is, data collected and published for comment, but not factored into the rankings) until the methodology is validated and stabilized.