

The 2014 Reform Efforts Survey Aggregate Dataset

Released: 2017/01/18

1. Overview

In the summer of 2014, the College of William and Mary's Institute for the Theory and Practice of International Relations (ITPIR) conducted a global elite survey, or the *2014 Reform Efforts Survey*, in partnership with the National Opinion Research Center (NORC) at the University of Chicago. This first-of-its-kind survey was explicitly designed to provide timely, detailed, and accurate data on the trustworthiness, influence, and performance of 100+ Western and non-Western development partners, as observed and experienced by in-country counterparts. The survey ultimately benefited from the participation of nearly 6,750 development policymakers and practitioners in 126 low- and middle-income countries, and analysis of the survey participant sample indicates that it is representative of the broader population of interest on several key dimensions.¹

2. Methodology

Prior to fielding the *2014 Reform Efforts Survey*, we spent nearly five years preparing a sampling frame of approximately 55,000 host government and development partner officials,² civil society leaders, private sector representatives, and independent experts from 126 low- and lower-middle income countries and semi-autonomous territories. While the true global population of development policymakers and practitioners is for all intents and purposes unobservable, we took painstaking efforts to identify a well-defined and observable population of interest. We define this population of interest as including those individuals who are knowledgeable about the formulation and implementation of government policies and programs in low- and lower-middle income countries at any point between 2004 and 2013. See more details on the process of sampling frame construction and the survey questionnaire in the online [Appendix](#) of Custer et al. (2015).

We administered the *2014 Reform Efforts Survey* between May and August 2014.³ Survey implementation was guided by the Weisberg total survey error approach and the Dillman tailored design method. Survey recipients were sent a tailored email invitation to participate in the survey that included a unique link to the online questionnaire. During the course of the survey administration period, survey recipients received up to three different automated

¹ Our survey key findings are discussed in [Parks et al. \(2015\)](#) and [Custer et al. \(2015\)](#). See more discussion on the representativeness of our sample in the online [Appendix](#) of Parks et al. (2015).

² Survey participants who worked at *in-country* offices of their respective development partner organizations were invited to participate in this survey and evaluated *other* partner organizations with which they directly worked with. The survey was designed in such a way that survey participants of development partner organizations *did not* evaluate their own organization but only evaluated other organizations.

³ Parks served as the Principal Investigator. This research was approved by the PHSC of the College of William & Mary under protocol #PHSC-2013-10-17-9041-bcpark.

electronic reminders, as well as some additional tailored reminders. Survey participants were able to take the survey in one of five languages: English, French, Spanish, Portuguese, and Russian.⁴

Of the 54,990 individuals included in the sampling frame, we successfully sent a survey invitation to the email inbox of over 43,427 sampling frame members.⁵ From this cohort of survey recipients, 6,731 participated, yielding an overall, individual-level survey participation rate of approximately 15.5%.⁶

3. Indicators of Development Partner Performance

The *2014 Reform Efforts Survey* aggregate dataset presents four different scores of development partner performance that measure 1) the frequency of communication (based on Question 13 in the questionnaire); 2) the usefulness of policy advice (Question 14); 3) agenda-setting influence (Question 21); and 4) helpfulness in reform implementation (Question 25). These scores intend to summarize our survey respondents' experience-based evaluations of each development partner.⁷ We include three versions of scores in the dataset: 1) unweighted scores; 2) scores that are weighted based on inverse-probability weights⁸; and 3) scores that are weighted equally across policy areas and countries. Scores vary only slightly depending on the weights. Indeed, correlations between weighted and unweighted scores are very high, ranging from 0.90-0.99.

4. Files Included in the 2014 Reform Efforts Survey Aggregate Dataset

The *2014 Reform Efforts Survey* aggregate dataset consists of 8 CSV separated files. Table 1 presents the short description of each data file and Table 2 shows the list of variables included in those data files. Each data file contains the same set of variables but presents them at different levels of aggregation. These variables include weighted or unweighted average scores of 1) frequency of communication, 2) usefulness of policy advice, 3) agenda-setting influence, and 4) helpfulness in reform implementation.

⁴ A professional translation company, Full Circle Translations—as well as several professional freelance translators and native and fluent speakers—conducted translation of the survey materials.

⁵ 25,919 survey recipients are currently—or have previously been—employed by developing country governments.

⁶ This observable figure of 15.5% is almost certainly an underestimate of the true, individual-level participation rate. At the time of survey implementation, we were unable to verify whether an intended survey recipient's email address was currently in-use. It should also be noted that, throughout this report, we employ the terms “participant” and “participation rate” interchangeably with the terms of “respondent” and “response rate.”

⁷ The survey was structured in the way that asked respondents to first identify a set of development partners with which they worked directly and then evaluated each one in the subsequent questions based on their own experiential knowledge.

⁸ See Appendix A for details on how inverse-probability weights are constructed.

Table 1: Files in the 2014 Reform Efforts Survey Dataset

File Names	Description
score_no_wt.csv	Unweighted scores of development partner performance. Bilateral development partner agencies are collapsed by country.
score_inv_prob_wt.csv	Scores of development partner performance weighted based on inverse-probability weights. Bilateral development partner agencies are collapsed by country.
score_country_policydomain_wt.csv	Scores of development partner performance that are weighted equally across policy area and country. Bilateral development partner agencies are collapsed by country.
score_no_wt_agency.csv	Unweighted scores of development partner performance by development partner agency. Bilateral development partner agencies are not collapsed by country.
score_inv_prob_wt_agency.csv	Weighted scores of development partner performance by development partner agency. Bilateral development partner agencies are collapsed by country.
score_country_policydomain_wt_agency.csv	Scores of development partner performance that are weighted equally across policy area and country. Bilateral development partner agencies are not collapsed by country.
score_by_country_no_wt.csv	Country-level scores of development partner performance with no weights.
score_by_country_inv_prob_wt.csv	Country-level scores of development partner performance that are weighted based on inverse-probability weights.

Table 2: Variables Included in Each File

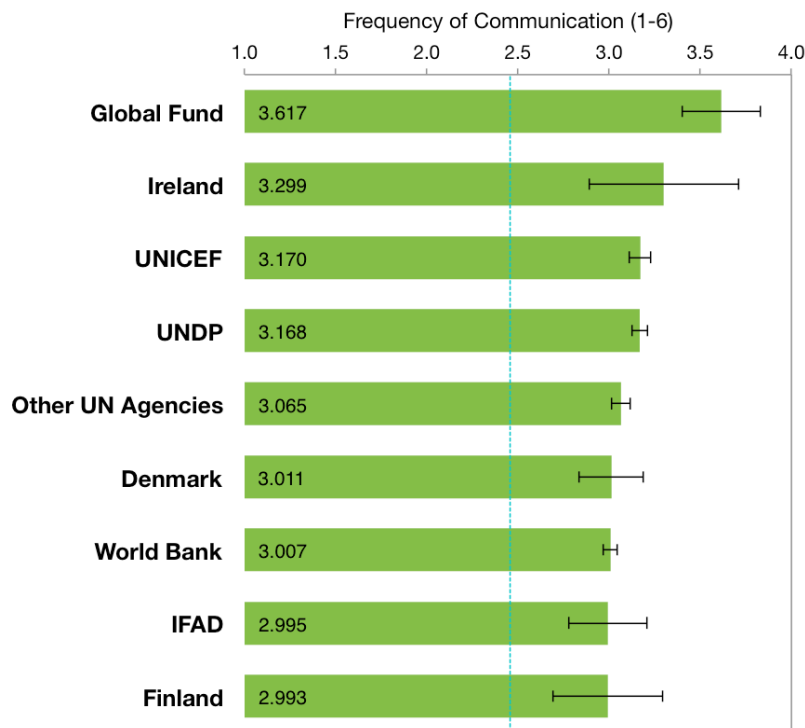
score_q13	Average of responses in Question 13 (frequency of communication)
stderr_q13	Standard error of score_q13
n_q13	Number of observations used to compute score_q13
score_q14	Average of responses in Question 14 (usefulness of policy advice)
stderr_q14	Standard error of score_q14
n_q14	Number of observations used to compute score_q14
score_q21	Average of responses in Question 21 (agenda-setting influence)
stderr_q21	Standard error of score_q21
n_q21	Number of observations used to compute score_q21
score_q25	Average of responses in Question 25 (helpfulness in reform implementation)
stderr_q25	Standard error of score_q25
n_q25	Number of observations used to compute score_q25

5. Scores of Development Partner Performance

5.1. Frequency of Communication

The frequency of communication score is generated based on host government officials' reported frequency of communication with each development partner on an ordinal scale of 1 to 6 where, 1 = "Once a year or less", 2 = "2 or 3 times a year", 3 = "About once a month", 4 = "2 or 3 times a month", 5 = "About once a week", and 6 = "Almost daily." A higher score indicates that a given development partner, on average, communicated more frequently with respondents. Figure 1 shows the top 10 ranking of development partners based on the frequency of communication.

Figure 1: The 10 Most Frequent Communicators



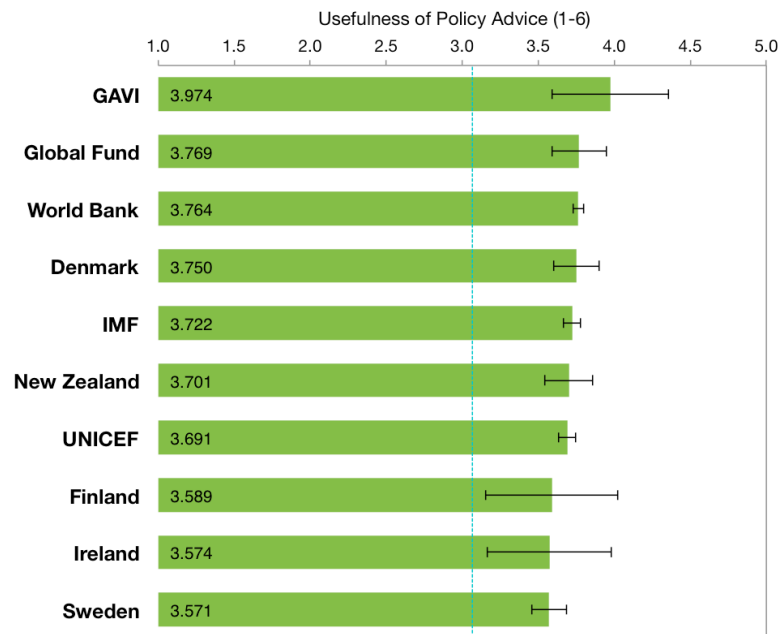
Notes: The vertical dotted line represents the average mean score of all development partners (with at least 10 observations). Error bars correspond to +/- one standard deviation.

5.2 Usefulness of Policy Advice

The 2014 Reform Efforts Survey provided survey participants with an opportunity to give direct feedback on the usefulness of policy advice provided by the development partners they

interacted with between 2004 and 2013. Survey participants were asked to rate the usefulness of each development partner’s advice within their own policy area of expertise on a scale of 1 to 5, with 1 signifying that the advice was *almost never useful* and 5 indicating that the advice was *almost always useful*. The usefulness of policy advice score captures the average frequency with which survey participants found the policy advice of a given development partner to be useful. Figure 2 shows the top 10 ranking of development partners based on the usefulness of policy advice.

Figure 2: The 10 Most Useful Development Partners in Offering Policy Advice



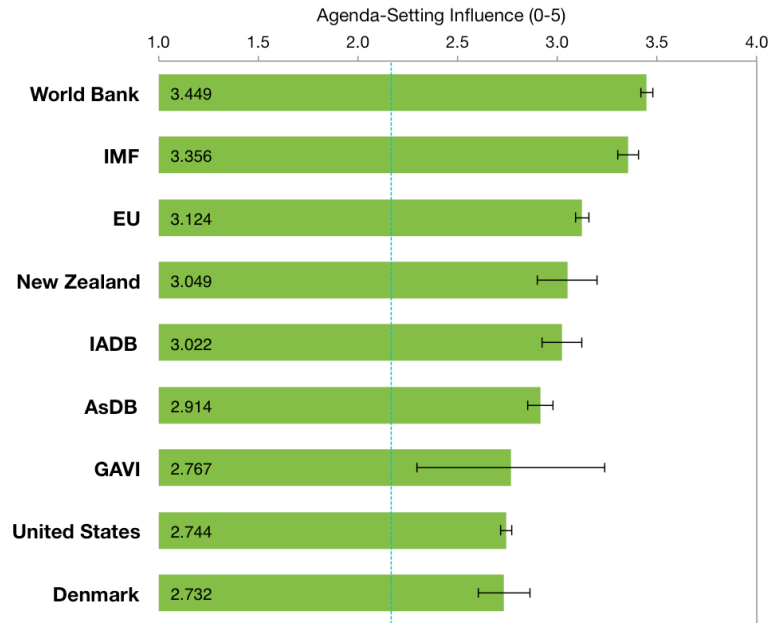
Notes: The vertical dotted line represents the average mean score of all development partners (with at least 10 observations). Error bars correspond to +/- one standard deviation.

5.3 Agenda-Setting Influence

The *2014 Reform Efforts Survey* provided participants with an opportunity to give direct feedback on the agenda-setting influence of the development partners they interacted with between 2004 and 2013. Participants were asked to rate the influence of development partners on their country’s decision to undertake reforms to solve three specific, self-identified policy problems on a scale of 0 to 5, with 0 signifying no influence and 5 indicating maximum influence.⁹ Figure 3 shows the top 10 ranking of development partners based on agenda-setting influence.

⁹ Survey participants identified three policy domain-specific problems that reforms tried to solve in their country (Question 20 in the survey). Subsequently, we asked them about the agenda-setting influence of

Figure 3: The 10 Most Influential Development Partners in Shaping Reform Agendas



Notes: The vertical dotted line represents the average mean score of all development partners (with at least 10 observations). Error bars correspond to +/- one standard deviation.

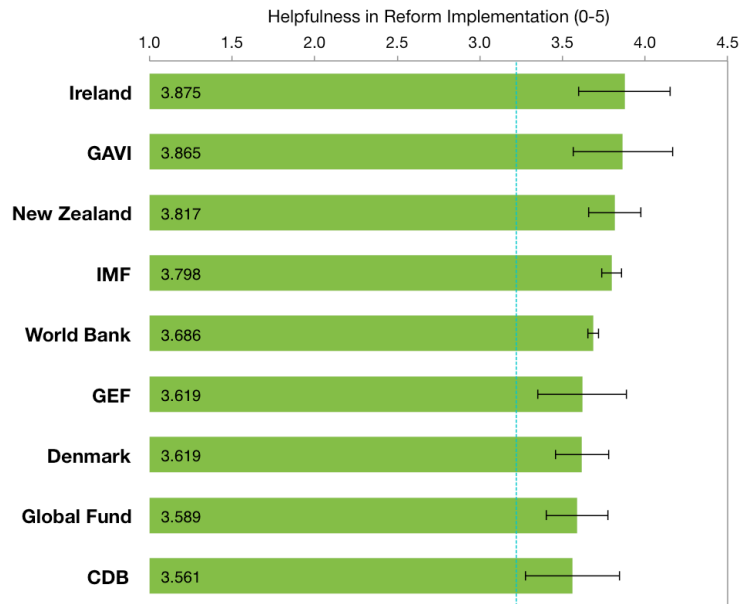
5.4 Helpfulness in Reform Implementation

The *2014 Reform Efforts Survey* provided participants with an opportunity to provide direct feedback on the helpfulness of individual development partners during the reform implementation process in 23 policy domains and 126 countries. Participants were asked to rate the helpfulness of the individual development partners that they identified as being involved in the implementation of reforms between 2004 and 2013 within their domain of expertise (e.g. health, education, anti-corruption). Development partners were rated on a scale of 0 to 5, with 0 indicating that they were not at all helpful in reform implementation and 5 indicating that they were extremely helpful.¹⁰ Figure 4 shows the top 10 ranking of development partners based on helpfulness in reform implementation.

individual development partners in their government’s decision to pursue reforms focused on those problems. See more details on the questionnaire in the online [Appendix](#) of Custer et al. (2015).

¹⁰ To capture host government perceptions of development partner helpfulness during reform implementation, we asked all survey participants to identify all of the development partners involved in the implementation of reforms in their country and policy domain out of a country-specific, fixed list.

Figure 4: The 10 Most Helpful Development Partners in Reform Implementation



Notes: The vertical dotted line represents the average mean score of all development partners (with at least 10 observations). Error bars correspond to +/- one standard deviation

References

Custer, Samantha, Zachary Rice, Takaaki Masaki, Rebecca Latourell and Bradley Parks. 2015. Listening to Leaders: Which Development Partners Do They Prefer and Why? Williamsburg, VA: AidData. <http://aiddata.org/listening-to-leaders>.

Parks, Bradley, Zachary Rice, and Samantha Custer. 2015. Marketplace of Ideas for Policy Change: Who do Developing World Leaders Listen to and Why? Williamsburg, VA: AidData and The College of William and Mary. <http://www.aiddata.org/marketplace-of-ideas-for-policy-change>

Respondents also saw all of their own write-in answers from Question 12 and were provided with the opportunity to identify an additional three development partners in Question 24. See more details on the questionnaire in the online [Appendix](#) of Custer et al. (2015).

Appendix A: Description of Weighting System for Data Aggregation

A.1: Inverse-Probability Weights

The response rate to the *2014 Reform Efforts Survey* was approximately 15%. In light of this relatively low response rate and imperfect information about the representativeness of our sample vis-à-vis the sampling frame (i.e. the population of interest), we employ non-response weights to account for unit non-response (or survey non-response) and generate unbiased and comprehensive aggregate statistics based on the individual respondent-level data. To generate non-response weights, we take the following steps. First, we estimate the probability of survey response by using a logistic regression. For all members of our sampling frame, we have information on their gender, country, institution types (e.g., finance ministry, anti-corruption agency, supreme audit institution) and stakeholder group (e.g., host government officials, development partners). We use all of these predictors to estimate the probability of survey response for each member of the sampling frame (as each of them turns out to be significant in predicting survey response). Second, we take the inverse of the estimated probability to arrive at the final non-response weights used for our analysis. To eliminate extreme weights, all weights above 2.5 were eliminated and replaced with 2.5. This only affected 66 of the 6,731 respondents.

A.2: Weights to Account for the Incompleteness of the Sampling Frame

In order to generate unbiased and comprehensive aggregate statistics, we also experiment with another weighting scheme that gives equal weight to every country-policy area (i.e. economic, governance, social and environmental, and general) pair. As pertains to global performance of individual development partners, unweighted statistics based on raw response data would likely exhibit bias in favor of Western development partners, assistance, and advice and against non-Western development partners, assistance, and advice. This is due to (1) uneven participant counts by country and (2) the construction of the sampling frame itself: non-Western donor staff and officials from closed and autocratic states proved more difficult to identify and contact. We expect that an average survey participant has more interaction and socialization with Western development partners than the overall population, and tends to work in countries and policy areas in which Western development partners have had relatively higher presence and influence.

Pro-Western bias aside, response counts vary greatly between countries and policy areas. A dual purpose of the weighting scheme is to ensure that our global statistics accurately capture (1) the global influence of an individual development partner as measured in an average country and (2) the performance of an average development partner in a single country as measured across *multiple* policy areas.

Here is a specific example. To counteract expected pro-Western bias and provide truly global measures of individual development partner performance, we conduct a separate, two-stage weighting process using data and response counts specific to each development partner. In the first stage, we up-weight all responses so that each country receives equal weight in the calculation of our global statistics. These country-level weights are calculated by finding the inverse proportion of the number of responses from a country against the maximum number of responses found in a single country across all sample countries.

In the second stage, we give equal weight to all policy area responses within each sample country. In-country policy area weights are calculated using the inverse proportion of the number of responses from a policy area within a country against the maximum number of responses found in a single policy area in that same country. In-country policy area weights are then incorporated into global development partner performance statistics via a two-step procedure. First, they are multiplied by the appropriate country-level weights from the first stage of the overall weighting process. Then the product of the two weights is rescaled to ensure that countries still receive equal weight in the global statistics.

These two different weighting schemes seek to address different types of bias, either deriving from unit non-response or from the way contact information was collected in the process of constructing the sampling frame. That said, it is important to note that they produce very similar scores, which are also strongly correlated with unweighted scores, as discussed in Appendix B.

Appendix B: Analysis of Unweighted and Weighted Scores

To test the sensitivity of our scores to our different weighting schemes, Figure B-1 shows correlations between unweighted scores and scores with inverse-probability weights and Figure B-2 shows correlations between unweighted scores and weighted scores that give equal weight to country and policy area. The correlation between unweighted and weighted scores range from 0.92 to 0.99 across four different indicators of development partner performance.

Table B-1: Comparison of Unweighted Scores and Scores with Inverse Probability Weights

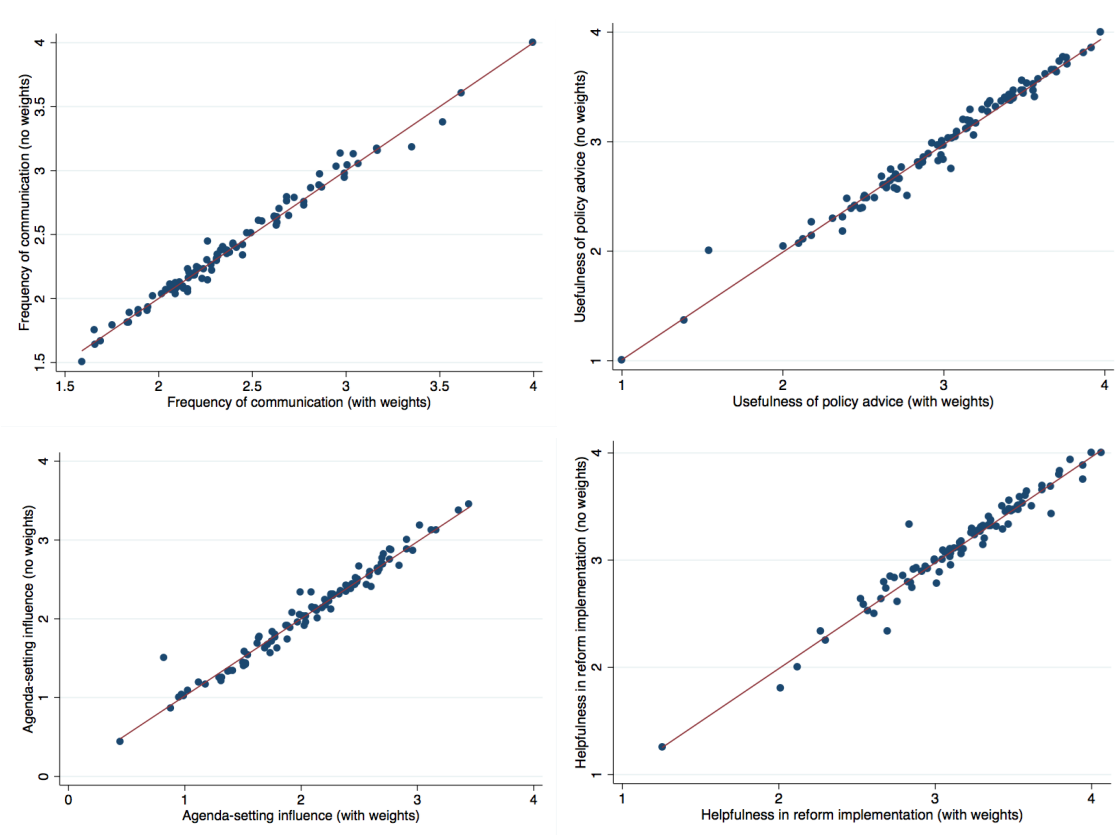


Table B-2: Comparison of Unweighted Scores and Scores with Weights that Give Equal Weight to Policy Area and Country

