

The Marketplace of Ideas for Policy Change

Who do developing world leaders listen to and why?



ACKNOWLEDGMENTS

This publication was prepared by a team led by Brad Parks, Zachary Rice, and Samantha Custer. Rebecca Latourell and Katie Paulson-Smith also made substantial contributions. However, the data and analysis contained in this report is the result of thousands of contributions from individuals and organizations over a five-year period. As such, we have accumulated many debts.

The authors are indebted to Ronald Rapoport, Matt Andrews and Sarah Rose, for their insights on the survey questionnaire, weighting methodology, and early draft report. We extend a special thanks to Michael Tierney for his generous institutional support to facilitate our research from 2010 to 2015. The authors also acknowledge the generous financial support received from the Smith Richardson Foundation, the John Templeton Foundation, the College of William & Mary, the London School of Economics and Political Science, and the World Bank.

The authors wish to thank our outstanding team of student research assistants: Kelvin Abrokwa-Johnson, Melanie Aguilar-Rojas, Arvin Alaigh, Justin Anderson, India Braver, Ben Bunch, Paul Burgess, Megan Coolidge, Anca Cretu, Caroline Davis, Jerusalem Demas, Jake Douglas, Thomas Fringer, Taylor Gaukel, Caitlin Grobarek, Lauren Hoy, Nakul Kadaba, Emily Kelly, Dan Kent, Greg Kirwin, Rachel Kling, Kristin Krejmas, Joe Lahouchuc, Alina Lam, Lisa Lee, Gabi Levi, Ziwei Li, Nic Martinez, Kate McGinnis, Nina Moreno, Robert Mosolgo, Ashley Napier, Rachel Nelson, James O'Leary, James Page, Alyse Prawde, Suzanne Reed, Emily Reeder, Kristin Ritchey, Sarah Sajewski, Salil Singhal, Austin Strange, Chelsea Stresler, Michael Testa, Jessica Tiblin, Marie Tummarello, Linh Vinh, Taylor Wilson, and Diana Winter. We also wish to thank Raneen Ak, Sofia Cevallos, Waly Faye, Gabriel Gaspar, Hagar Hisham, Sophia Kosar, Sarah Mellman, Sarah Parks, Alexander Prokhorov, Daria Prokhorova, Vincent Rampino, and Sarah Wall for their translation work and assistance.

We are also deeply grateful to each of the individuals who helped us pre-test the survey questionnaire and online survey interface: Jean Antonin, Hovhannes Avetisyan, Nicky Bell, Victor Borges, Michael Butler, Raúl Chambote, Eduardo Cruz, Amacodou Diouf, Sonia Flamenco, Francis Gatere, Vanndy Hem, Justin Mosa, Decontee King-Sackie, Nicolas Lemay-Hebert, Gibrill Jalloh, Carlos Melara, Dan Nielson, Ifueko Okauru, Charles Perla, Masroor Qureshi, Rija Ranaivoarison, Henri Ranaivoson, Marius Ratolojanahary, Ivohasina Razafimahefa, Haja Razafinjatovo, Elva Richardson, Phil Roessler, Said Salime, Alpha Sesay, Anna Sevortyan, Pranay Sinha, Ian Walker, Matt Winters, Alex Wooley, and Xolani Zitha.

We are also indebted to the following people for the feedback they provided on the design and content of the *2014 Reform Efforts Survey* questionnaire: Victor Borges, Jeff Chwieroth, Luis Cosenza, Casey Dunning, John Evans, Christian Henning, Sheila Herrling, John Hicklin, Stephen Knack, Eva Krampe, Julia Maciel, Mishkat Moumin, Lucrecia Palacios, Garry Robins, Said Salime, Sarah Jane Staats, Christian Stein, David Wheeler, Mark Buntaine, and Jennifer Widner. For their feedback on an early draft of this report, we are also grateful to Ingo Keilitz, Takaaki Masaki, and Lodewijk Smets.

We could not have successfully gathered the data necessary to produce this report without an exceptional team of researchers and programmers at NORC at the University of Chicago: Renee Hendley, Kate Hobson, David Kordus, Elise Le, Aparna Ramakrishnan, Alex Rigaux, Bhanuj Soni, and Stacy Stonich. We thank them for their many hours of hard work and dedication.

We are also thankful for the assistance of various staff at and friends of the College of William & Mary's Institute for the Theory and Practice of International Relations, including Aneta Baltes, Priscilla Caldwell, Albert Decatur, Joyce Deuley, Josephine Garcia, Marisa LeForge, Sasha Maliniak, and Jason Pully.

Last, but certainly not least, we wish to thank participants in the *2014 Reform Efforts Survey*. Were it not for their willingness to graciously share of their time and experiences, this report would not have been possible.

The authors are solely responsible for any findings and content found in this report; we realize that those who we have acknowledged may hold views that are different from our own.

TABLE OF CONTENTS

1. Introduction	1
1.1. Making Reform Incentives Work for Developing Countries	1
1.2. External Assessments and the Marketplace of Policy Ideas	1
1.3. The Missing Data Problem in Understanding External Assessment Influence	2
2. Methodology	4
2.1. Defining the Population of Interest	5
2.2. Creating the Sampling Frame	5
2.3. Designing the Survey Questionnaire	6
2.4. Survey Implementation	6
2.5. Composition and Representativeness of the Survey Sample	7
3. External Assessment Influence: Global-Level Findings	17
3.1. External assessments can influence reform efforts, but some are far more influential than others.	17
3.2. External assessment influence is strongest at the agenda-setting stage of the policymaking process.	28
3.3. In a crowded marketplace of policy ideas, assessment familiarity delivers an influence dividend.	30
3.4. Influencing reform efforts from the outside is substantially easier in some policy domains than in others.	37
3.5. Geography matters: assessments achieve widely varying levels of uptake across regions and countries.	45
4. External Assessment Influence: Assessment-Level Findings	56
4.1. Do the types of policy problems that external assessments seek to address affect their influence?	56
4.2. External Assessment Characteristics and Uptake Among Developing Countries	64
4.3. Which Assessment Attributes Do Not Appear to Influence Behavior Change?	68
4.4. Alignment with local priorities is a key determinant of external assessment influence.	70
4.5. Survey Participants Speak to Factors Contributing to External Assessment Influence	74
4.6. Position-Driven Perceptions of Contributing Factors to External Assessment Influence	76
5. External Assessment Influence: Country-Level Findings	80
5.1. External assessments are more influential in smaller countries.	80
5.2. External assessments are more influential in countries with democratically elected leaders and a free press.	82
5.3. Countries with effective public sector institutions are more amenable to external analysis and advice.	84
5.4. Reform-minded governments are more likely to draw upon external sources of advice and analysis.	85
5.5. Aid-dependent countries are more influenced by external assessments.	86
5.6. The influence of external assessments may depend upon the training and work experience of political leaders.	87
6. Findings: Intended and Unintended Effects	89
6.1. Assessments have largely positive and rarely negative effects on the policy process.	90
6.2. Governmental and non-governmental stakeholders broadly agree on the effects that assessments have.	91
6.3. External Assessment Effects in Highly Influenced and Weakly Influenced Countries	93
6.4. Influence To What End: Do External Assessments Focus Reforms on the Right Problems?	95
6.5. The Success of Reforms Influenced (or Not) by External Assessments	98
7. Conclusion	115
8. References	117

LIST OF FIGURES AND TABLES

Figure 1. Who Participated in the 2014 Reform Efforts Survey?	8
Figure 2. Countries with the Highest and Lowest Number of Participants	11
Figure 3. Policy Domain Expertise of Survey Participants	13
Figure 4. Roles Played by Survey Participants in the Policy Process	14
Figure 5. External Assessment Influence at the Agenda-Setting Stage	19
Figure 6. External Assessment Influence at the Reform Design Stage	24
Figure 7. Assessment Influence of Development Partners in Agenda Setting and Reform Design	26
Figure 8. The Strong Assessment-Level Correlation between Influence in Agenda Setting and Reform Design	29
Figure 9. The Positive Relationship between Assessment Familiarity and Agenda-Setting Influence	33
Figure 10. The Agenda-Setting Influence of External Assessments, by Policy Domain	42
Figure 11. The Changing Influence of External Assessments by Reform Phase and Policy Domain	44
Figure 12. The Influence of External Assessments on the Decision to Pursue Reforms, by Country	46
Figure 13. The Influence of External Assessments on the Design of Reforms, by Country	47
Figure 14. The Most and Least Influential External Assessments at the Agenda-Setting Stage, by Region	48
Figure 15. The Most and Least Influential External Assessments at the Reform Design Stage, by Region	50
Figure 16. The Agenda-Setting Influence of External Assessments, by Country	51
Figure 17. The Reform Design Influence of External Assessments, by Country	53
Figure 18. The Strong Country-Level Correlation between Assessment Influence in Agenda Setting and Reform Design	54
Figure 19. Substantive Focus of Reforms Influenced by External Assessments	57
Figure 20. Assessment Influence by Policymaking Phase and Problem Type	62
Figure 21. Assessment Influence by Income and Problem Type	63
Figure 22. External Assessment Characteristics and Influence at the Agenda-Setting Stage	65
Figure 23. The Agenda-Setting Influence of Country-Specific Assessments vs. Cross-Country Benchmarking Exercises, by Stakeholder Group	67
Figure 24. The Primary Reason for External Assessment Influence	70
Figure 25. The Primary Reason for External Assessment Influence, by Stakeholder Group	73
Figure 26. Contributing Factors to External Assessment Influence	75
Figure 27. Contributing Factors to Assessment Influence, by Respondent Position Type	77
Figure 28. The Negative Relationship between Country Size and Assessment Influence at the Agenda-Setting Stage	81
Figure 29. Assessment Influence (0-5), by Country Income Level	82
Figure 30. Assessment Influence (0-5), by Regime Type	83
Figure 31. Assessment Influence (0-5), by Country Press Freedom Status	84
Figure 32. External Assessment Influence, by Policy and Institutional Performance at the Country Level	85
Figure 33. The Positive Relationship between CPIA-Based Success in Reform Implementation and Assessment Influence	86
Figure 34. The Importance of Bargaining Power on Agenda-Setting Influence of External Assessments	87
Figure 35. The Positive Relationship between Senior Policymaking Team Socialization and Assessment Influence at the Agenda-Setting Stage	88
Figure 36. The Strength of External Assessment Effects	90
Figure 37. The Strength of External Assessments Effect, by Stakeholder Group	92
Figure 38. The Strength of External Assessment Effects in 10 Most and Least Influenced Countries	94
Figure 39. Success in Reform Implementation, by Country	99
Figure 40. The Positive Relationship Between Net Reform Support and Success in Reform Implementation	104
Figure 41. Net Reform Support, Success in Reform Implementation, and External Assessment Influence	105
Figure 42. External Assessment Influence on More and Less Successful Reform Efforts	107
Figure 43. Assessment Characteristics Associated with Significantly More or Less Success in Reform Implementation	110
Figure 44. The Agenda-Setting Influence of Assessment Suppliers, by Level of Success in Reform Implementation	112
Figure 45. Reform Design Influence of Assessment Supplier Types, by Level of Success in Reform Implementation	113
Table 1. Respondent Familiarity with Assessments: The 3 Most Recognized Assessments in Each Policy Domain	34
Table 2. The 3 Most Influential Assessments in Each Policy Domain (Level of Agenda-Setting Influence)	38
Table 3. The 3 Most Influential Assessments in a Government's Decision to Pursue Reforms, by Specific Problem Type	58
Table 4. The Primary Reason for Assessment Influence in Each Policy Domain	71
Table 5. Assessments with Substantially More Influence Over Reforms Focused on Critical Problems	96
Table 6. Assessments with Roughly the Same Influence Whether or Not Reforms Focused on Critical Problems	97

ACRONYMS AND ABBREVIATIONS

ABC	Agência Brasileira de Cooperação (Brazilian Cooperation Agency)
ACP	African, Caribbean and Pacific Group of States
ADA	Austrian Development Agency
ADFD	Abu Dhabi Fund for Development
AECID	Spanish Agency for International Development Cooperation
AFD	Agence Française de Développement (French Development Agency)
AFESD	Arab Fund for Economic and Social Development
AfDB	African Development Bank
AGOA	African Growth and Opportunity Act
AICD	The Africa Infrastructure Country Diagnostic
ALBA	Bolivarian Alliance for the Peoples of Our America
AMF	Arab Monetary Fund
APRM	African Peer Review Mechanism
AsDB	Asian Development Bank
AU	African Union
AusAID	Australian Agency for International Development
BADEA	Arab Bank for Economic Development in Africa
BANDES	Venezuelan Economic and Social Development Bank
BNDES	Brazilian Development Bank
BTC	Belgian Development Agency
CABEI	Central American Bank for Economic Integration
CAF	Development Bank of Latin America
CDB	Caribbean Development Bank
CIDA	Canadian International Development Agency
CIPE	Country Institutional and Policy Evaluation
CPA	Country Performance Assessment
CPIA	Country Policy and Institutional Assessment
CSO	Civil Society Organization
DAC	Development Assistance Committee
Danida	Danish International Development Agency
DFID	Department for International Development
DIE	German Development Institute
EBRD	European Bank for Reconstruction and Development
EC	European Commission
EITI	Extractive Industries Transparency Initiative
ENPI	European Neighborhood and Partnership Instrument
EU	European Union
FATF	Financial Action Task Force
FDI	Foreign Direct Investment
FIAS	Foreign Investment Advisory Service
FSAP	Financial Sector Assessment Program
GAVI	Global Alliance for Vaccines and Immunization
GEF	Global Environment Facility
GIZ	Deutsche Gesellschaft für Internationale
GOL	Government of Liberia
HIPC	Heavily Indebted Poor Countries Initiative
IADB	Inter-American Development Bank
IBP	International Budget Partnership
IEG	Independent Evaluation Group
IEO	Independent Evaluation Office
IFAD	International Fund for Agricultural Development

IFC	International Finance Corporation
IGO	Inter-Governmental Organization
ILO	International Labor Organization
IMF	International Monetary Fund
IsDB	Islamic Development Bank
ISS	Immunization Services Support
JBIC	Japan Bank for International Cooperation
JICA	Japan International Cooperation Agency
KFAED	Kuwait Fund for Arab Economic Development
KfW	Kreditanstalt für Wiederaufbau (German Development Bank)
LFADA	Libyan Fund for Aid and Development in Africa
LuxDev	Luxembourg Development Cooperation
KOICA	Korea International Cooperation Agency
MCC	Millennium Challenge Corporation
MDGs	Millennium Development Goals
MDRI	Multilateral Debt Relief Initiative
MENA	Middle East and North Africa
NATO	North Atlantic Treaty Organization
NGO	Non-Governmental Organization
NORAD	Norwegian Agency for Development Cooperation
NORFUND	Norwegian Investment Fund for Developing Countries
NZAID	New Zealand Aid Program
OECD	Organization for Economic Co-operation and Development
OFID	OPEC Fund for International Development
PEFA	Public Expenditure and Financial Accountability
PHSC	Protection of Human Subjects Committee
PISA	OECD's Program for International Student Assessment
PRES	Poverty Reduction Effectiveness Situation
SDC	Swiss Agency for Development and Cooperation
SDGs	Sustainable Development Goals
SFD	Saudi Fund for Development
SIDA	Swedish International Development Cooperation Agency
Taiwan ICDF	Taiwan International Cooperation and Development Fund
TI	Transparency International
TIKA	Turkish Cooperation and Coordination Agency
TIP	Trafficking in Persons
UAE	United Arab Emirates
UK	United Kingdom
UN	United Nations
UNDP	United Nations Development Program
UNECA	United Nations Economic Commission for Africa
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
US	United States
USAID	United States Agency for International Development
USG	United States Government
WB	World Bank
WTO	World Trade Organization

1. Introduction

Imagine that you have been appointed to be the Minister of Finance in an aid-dependent developing country. You have limited time, money, and political capital. Interest groups from line ministries, the military, civil society, and the private sector want you to spend scarce public resources on a laundry list of diverse priorities. Meanwhile, dozens of foreign governments, multilateral development finance institutions, and international advocacy groups seek to assess your government's performance in several policy domains—from health and education to anti-corruption and public financial management. Each assessment supplier is attempting to influence your policy decisions and actions in ways that may or may not reflect the values and interests of your domestic constituencies.

Performing well on some of these external assessments may unlock new aid, trade, or investment flows, while others offer no direct financial or material reward but are closely watched by your country's donors, creditors, and investors. Assessments may or may not contain specific and actionable policy recommendations. Some assessments are up-to-date, while others evaluate the performance of the previous administration.

This patchwork quilt of external assessments leaves you in a quandary. Do you use the most accurate and up-to-date assessments to identify your country's most critical policy problems and potential solutions? Or should you focus on improving your government's performance on a small subset of assessments in an effort to signal your credibility to one or more donors, creditors, or investors? You could also choose to ignore these assessments altogether.

1.1. Making Reform Incentives Work for Developing Countries

In 2010, we launched the *Making Reform Incentives Work for Developing Countries* project with a simple goal: to explain how development partners can more effectively support reform efforts in low- and middle-income countries. Specifically, we sought to begin building a body of empirical evidence about whether, when, how, and why external assessments of government performance influence the decisions and actions of policymakers in the developing world. Administered last summer, the *2014 Reform Efforts Survey* collected experiential and perception-based information directly from thousands of development policymakers, practitioners, and experts in these countries. In this report, we examine more than 100 external assessments of government performance across 23 policy domains in 126 developing countries to understand their influence and impact.

1.2. External Assessments and the Marketplace of Policy Ideas

Governments across the developing world face a daunting set of socio-economic, environmental, and governance challenges in the 21st century, and the international community provides hundreds of billions of dollars each year to address these challenges. However, there is a growing recognition among policymakers and scholars that “throwing more resources” at persistent development problems is at best an insufficient—and at worst a counterproductive—response (Grindle 2004; North et al. 2007; and Pritchett et al. 2013). Many of the most deeply rooted problems in developing countries require challenging vested interests, changing deeply ingrained beliefs and behaviors, or otherwise disrupting the status quo (Acemoglu and Robinson 2000; Van de Walle 2001; and North et al. 2007).

To this end, foreign governments, inter-governmental organizations, and non-governmental organizations have created a wide range of instruments to influence policy change, including: conditional aid and debt relief programs, blacklists and watch lists, performance-based organizational accession procedures,

cross-country benchmarking initiatives, and in-depth, country-specific diagnostics. The common denominator across these instruments is that they involve some explicit or implicit assessment of government performance geared towards spurring and sustaining economic, political, social, and environmental reforms in developing countries.

External assessments of government performance are diverse in their content, scope, and sponsorship, but they all seek in some way to change the ideas, interests, and incentives of developing country leaders. Some increase the financial or reputational benefits that policymakers can reasonably expect reforms to yield. Others socially sanction or impose financial penalties on those who postpone or rollback reforms. Still others seek to empower change agents with new sources of evidence, analysis, and advice.¹

Multilateral institutions have been particularly assertive in their efforts to use performance assessment tools to influence the decision-making calculus of reform for policymakers. Consider the WB's annual *Doing Business Report*. This cross-country benchmarking assessment, which is closely followed by donors and private investors, measures the "ease of doing business" in nearly 190 countries, and shines a spotlight on the fastest-reforming governments (IEG 2008; Schueth 2011). The AsDB, the AfDB, the IADB, the GEF, and many other multilateral aid agencies have also developed assessments of their own. Some of these assessments are geared towards helping the authorities in developing countries better understand policy problems and potential responses to these problems, while others are used to support performance-based resource allocation systems and thereby reward and incentivize governments that pursue reform (AsDB 2005; Hout 2007; and Perakis and Savedoff 2015).²

Foreign governments are also actively engaged in the production and promotion of developing country performance assessments. For example, the US State Department publishes an annual *Trafficking in Persons Report* that "names and shames" governments around the world that do not comply with anti-human trafficking standards (Cho et al. 2014; Kelley and Simmons 2014a). The US Government has also created an explicit set of policy performance requirements for countries that seek Millennium Challenge Corporation assistance, which has created an incentive for governments to pursue liberal economic policies, democratic governance, and targeted investments in health, education, and environmental protection (Radelet 2006, 2007; Öhler et al. 2013; and Parks 2014).

1.3. The Missing Data Problem in Understanding External Assessment Influence

Despite the extraordinary amount of time, money and effort that the international community expends to produce and promote these assessments, there is a remarkable lack of understanding about which performance assessments are most and least effective at spurring or sustaining reforms—and why. The absence of reliable data that measure the resonance, uptake, and influence of these assessments is the key constraint that has prevented the accumulation of a credible evidence base.

There is no lack of theory about the potential determinants of external assessment influence; however, very few hypotheses have been subjected to empirical scrutiny. One set of theories focuses on the observable attributes of external assessments themselves—for example, the identity of the supplier and

¹ For more on this topic, see Noland 1997; Jacoby 2006; Pop-Eleches 2009; Krasner 2011; Orbie and Tortell 2009; Sharman 2008, 2009, 2013; Pitlik et al. 2010; Parks and Rice 2013; Molenaers 2012; Del Biondo forthcoming; Molenaers et al. forthcoming; Kelley and Simmons 2014a, 2014b; Parks 2014; and Perakis and Savedoff 2015.

² The European Commission has also developed a number of policy instruments—e.g., the Special Incentive Arrangement for Sustainable Development and Good Governance, the Governance Incentive Tranche, Economic Partnership Agreements, Stabilization and Association Agreements, and various performance-based budget support schemes—to encourage more efficient service delivery and accountable governance (Kelley 2006; Borrman and Busse 2007; Saase 2008; Orbie and Tortell 2009; Kleeman 2010; Zimelis 2011; Molenaers and Nijs 2011; Savedoff 2011; and Del Biondo and Orbie 2014).

sponsor, whether the assessment is country-specific or cross-country in nature, the use of incentives and penalties, and the feasibility of the performance requirements.

Another set of theories focus on how the attributes of target countries may influence the uptake of external analysis and advice by policymakers in developing countries. These country-level characteristics include, but are not limited to: regime type, income, population size, state capacity, social cohesion, political commitment, the presence and professional training of reform-minded technocrats, dependence upon external sources of aid, and the degree to which donors with an in-country presence coordinate their policy messaging.

With this report, we seek to dislodge the assumption that it is not possible to measure the local resonance, uptake, or influence of the assessments that governments and international organizations use to encourage reform in developing countries. By putting forth a set of transparent and rigorous methods that we hope can be replicated and extended by others, we seek to spur a new generation of empirical research on the influence of external assessments.

This report is organized in the following manner. First, we provide an overview of our methodology, including the survey sample. Second, we explore global-level findings of how external assessment influence varies by policy domain, country, and stage of the reform process. Third, we examine variation in the influence of particular types of assessments. Fourth, we test which country-level attributes that contribute to the influence of external assessments. Finally, we examine the drivers of assessment influence as identified by survey participants and the effects of external assessments on downstream reform successes and failures.

2. Methodology

Surveys of elites and “opinion leaders” do not always meet high standards of methodological rigor. Sampling frames are rarely assembled in a transparent, systematic, and replicable manner. Rather than identifying a clear population of interest based upon an explicit set of inclusion and exclusion criteria, professional survey organizations and the organizations that employ them usually rely upon “black box” sampling frames—for example, a “global opinion leader database” (ADB 2009, 2012), a “database of civil society contacts” (IMF 2009: 146) or a “[list of] stakeholders and clients” (WB 2013: 3). Those who conduct elite or opinion leader surveys seldom report or collect demographic information about the respondents and non-respondents in their sampling frames, making it impossible to examine the representativeness of the samples that are drawn (Milner and Tingley 2013).

Some organizations rely upon a “business card approach,” whereby an organization supplies a list of their in-country contacts to a survey implementation firm as the sampling frame. However, this approach poses a significant risk of selection bias and survey findings that are not representative of the population of interest. Other organizations adopt a “positional” approach, whereby one survey respondent is asked to provide authoritative answers on behalf of an entire government or organization (IMF 2009; IATI 2013; and OECD 2015). This approach belies the complexity within developing country institutions since people within a given institution may provide different responses to the same questions.³ Selection bias, non-response bias, and social desirability bias also arise when surveys pose questions about the same organizations that sponsor or commission those surveys (Ganster et al. 1982; Rogelberg and Stanton 2007; and Spitzmüller et al. 2007).

While scholars are more willing to concede that their findings are not representative of the populations about which they seek to draw inferences, their methods are not substantially more rigorous. In many cases, samples are drawn based upon ease of access and are analyzed in the absence of carefully constructed sampling frames deemed too costly, time-consuming or complex to produce (Gabre-Madhin and Haggblade 2001; Bishin et al. 2006; Gray and Slapin 2012; Hafner-Burton et al. 2014; and De Bruyn 2015).⁴

Surveys can improve our understanding of the factors that motivate government officials to pursue or postpone reforms (Gray and Slapin 2012), as well as shed light on the variables that policymakers and practitioners take into account when setting priorities and taking policy decisions (Steen 2003; Taylor 2011; and Jennings and Hall 2011). However, surveys of policymakers and practitioners need to be designed and implemented in ways that better account for issues of bias and representativeness.

To this end, prior to fielding the *2014 Reform Efforts Survey*, our research team spent nearly five years preparing a sampling frame of approximately 55,000 host government and development partner officials, civil society leaders, private sector representatives, and independent experts from 126 low- and lower-middle income countries and semi-autonomous territories.⁵ In this section, we provide an overview of our methodology and describe key attributes of our sampling frame and survey participant sample.

³ The notion that an official organizational position can accurately summarize the perceptions and experiences of the many individuals who work within that organization strains credulity (Spitzmüller et al. 2007).

⁴ For example, a recent study in the prestigious political science journal *International Organization* notes that “we built the sample through professional contacts and then expanded the sample with personal introductions” (Hafner-Burton et al. 2014: 863).

⁵ For the purposes of this study, semi-autonomous territories (e.g., Kurdistan) are treated as separate entities from the countries that contain them (e.g., Iraq). The *2014 Reform Efforts Survey* was framed to ask individuals from semi-autonomous territories about reforms undertaken by the government in that semi-autonomous territory.

2.1. Defining the Population of Interest

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. For more information on sampling frame inclusion criteria, see Appendix D.

In recognition of the need for cross-country comparability and the fact that every government consists of a unique set of institutions and leadership positions, we identified our population of interest by first mapping country-specific public sector institutions (and leadership positions within those institutions) back to an ideal-typical developing country government. This ideal-typical government consisted of 33 institution types, such as a Ministry of Finance, a Supreme Audit Institution, and a National Statistical Office (see Appendix D). We then identified functionally equivalent leadership positions within these institutions, and the specific individuals who held these positions between 2004 and 2013. For the four additional stakeholder groups that we included in our sampling frame (in-country development partners, domestic civil society and non-governmental organizations, private sector associations, and independent experts), we undertook a similar process of first mapping country-specific institutions and positions, and then identifying the individuals who held those positions between 2004 and 2013.

Identifying functional equivalents at the institution- and leadership position-level resulted in a sampling frame that enables comparison across countries. In addition, by clearly defining a population of interest and constructing a master sampling frame that is stratified by country, stakeholder group, and institution type, we managed to overcome one of the most vexing challenges associated with expert panels and opinion leader surveys: the absence of detailed demographic data and the inability to assess the representativeness of findings at various levels. The stratification of our master sampling frame by country, stakeholder group, and institution type makes it possible to generate extremely granular elite survey data that can be published at varying levels of disaggregation without compromising participant confidentiality. It also enables analysis of the factors that influence participation rates as well as the underlying sources of response bias. A more detailed description of the construction of the master sampling frame can be found in Appendix D.

2.2. Creating the Sampling Frame

Our ability to select individuals from the population of interest for inclusion in our final sampling frame was constrained by the availability of individual contact information. We identified the contact information of potential survey participants using publicly available resources, such as organizational websites and directories, international conference records, Who's Who International, and public profiles on LinkedIn, Facebook, and Twitter. We successfully identified the individual contact information of approximately 68% of the population of interest, including 65% of host government officials, 61% of private sector representatives, 77% of development partner staff, and 79% of independent country experts.⁶ We have attempted to correct for potential biases that may result from variation in contact availability, country sample size, and participation rates by implementing the weighting system described in Appendix C.⁷

⁶ We also had an unusually high success rate (70.04%) for finding the contact information of domestic CSO/NGO representatives. This is in part because, given the small size of many CSOs and NGOs, we considered an organizational email address sufficiently specific to include a CSO/NGO individual in our sampling frame. This was not true for members of any other stakeholder group, for whom we required a personal email address.

⁷ This pattern in contact availability was exacerbated by stakeholder-group specific trends in survey receipt and participation rates (see Section 2.5.1).

2.3. Designing the Survey Questionnaire

The methods described in Weisberg (2005) and Dillman et al. (2009) informed the design and evaluation of the questions included in the *2014 Reform Efforts Survey*. We drafted an initial set of questions based on similar surveys previously undertaken by the AsDB, WB, the IMF's IEO, and Princeton University's Task Force on the Changing Nature of Government Service (Volcker 2009; IMF 2009; and AsDB 2010). We then updated these questions according to lessons learned from the *2012 MCA Stakeholder Survey* (Parks and Rice 2013).

We evaluated all questions according to several criteria, including: (1) salience and clarity, (2) consistent interpretation, (3) avoiding oversimplification, (4) sufficient context, and (5) testing for potential response and non-responses biases.⁸ The design of the survey web interface was informed by Couper (2008) and guided by the additional experience of NORC at the University of Chicago. We conducted pre-testing of English-language questionnaire drafts and a web-based survey instrument via in-person cognitive interviews with current and former development policymakers, practitioners, and scholars. Survey translations and the foreign-language online interface were tested in-person and remotely via a personalized hyperlink. For more information on survey questionnaire design, see Appendix E.

2.4. Survey Implementation

We administered the *2014 Reform Efforts Survey* between May and August 2014. Survey implementation was again guided by the Weisberg total survey error approach and the Dillman tailored design method.⁹ 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 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.¹⁰

Of the 54,990 individuals originally included in the sampling frame, we successfully sent a survey invitation to the email inbox of over 43,427 sampling frame members.¹¹ From this cohort of survey recipients, 6,744 participated, yielding an overall, individual-level survey participation rate of approximately 15.53%.¹² To put this participation rate in its appropriate context, it should be noted that:

- Elite survey participation rates are usually not high, particularly when they are conducted online.¹³
- High individual participation rates are difficult to achieve in developing countries due to participation obstacles such as electricity outages, limited Internet access, weak privacy protections, and linguistic heterogeneity (Couper 2000: 474; Stecklov and Weinreb 2010).
- Our goal was to maximize survey coverage across “country-policy domain” pairs rather than to obtain a high individual-level participation rate.¹⁴ Out of 3,024 possible country-policy domain

⁸ For example, early survey questionnaire drafts included a question on the influence of external assessments of government performance on reform *implementation* efforts. After pre-testing and evaluation, it became evident that the idea of “influence in reform implementation” did not resonate with survey recipients.

⁹ 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.

¹⁰ 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.

¹¹ 25,919 survey recipients are currently—or have previously been—employed by developing country governments.

¹² This observable figure of 15.53% 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.”

¹³ Individual-level participation rates to email surveys are typically low (Sheehan 2006; Shih and Fan 2008). Elite survey participation rates also tend to be low, falling in the 10 to 30% range (Gabre-Madhin and Haggblade 2001; Bishin et al. 2006; Jones et al. 2008; Ban and Vandenabeele 2009; Volcker 2009; Gray and Slapin 2012; Ellinas and Suleiman 2012; Pew Research Center 2012; Hafner-Burton et al. 2014; and Avey and Desch 2014).

pairs, the country-policy domain coverage rate was approximately 60%.¹⁵ After collapsing specific policy domains into broader policy areas (i.e., economic policy, governance, social and environmental policy, and general), the country-policy coverage rate increases to above 98% (with at least 1 participant).¹⁶

Please refer to Appendix B for a more detailed discussion of participant attrition and potential sources of bias.

2.5. Composition and Representativeness of the Survey Sample

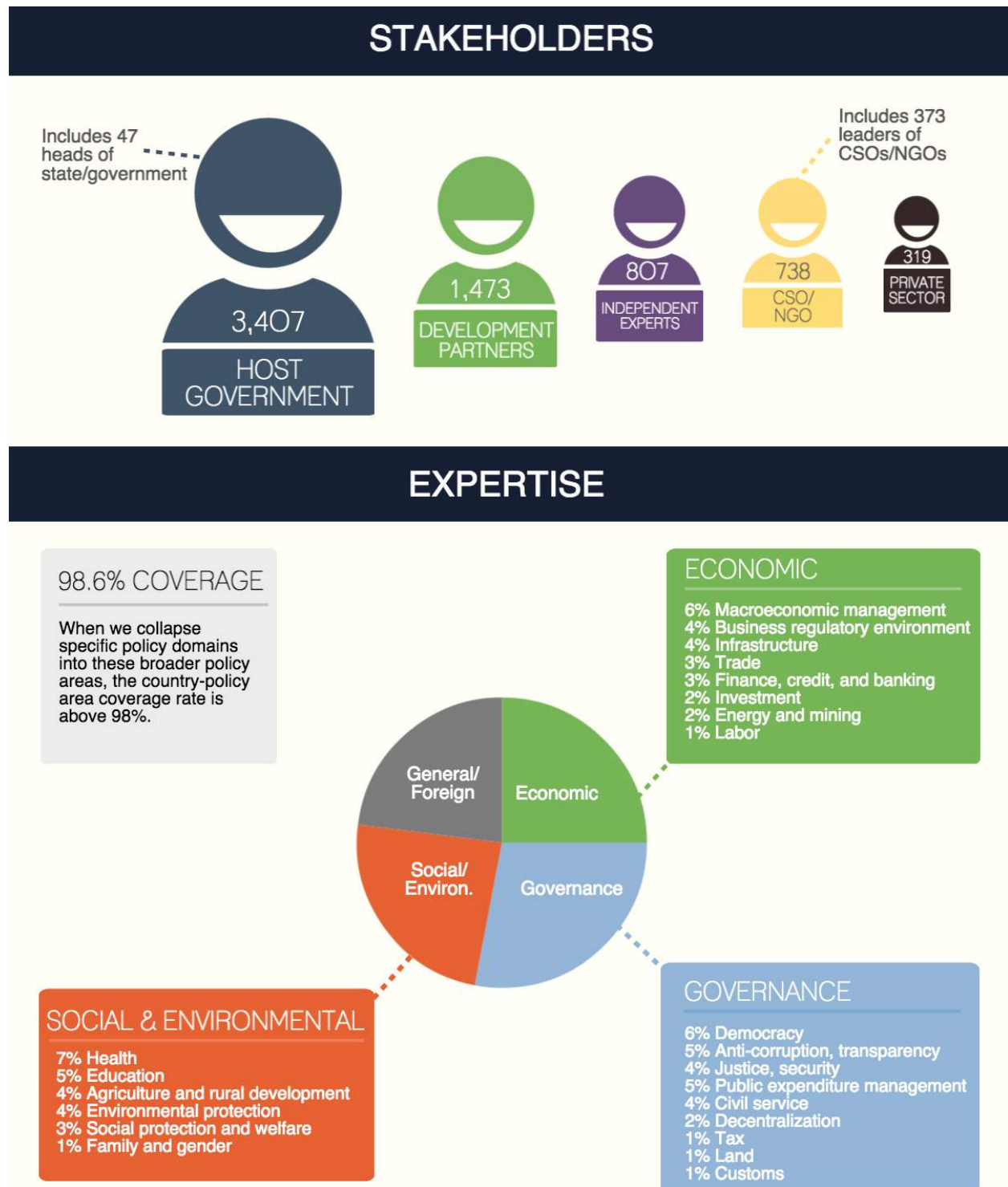
This section describes the characteristics of the individuals who participated in the *2014 Reform Efforts Survey*. For an overview of the attributes of our survey participants, please see Figure 1 below.

¹⁴ By mapping institution types to expected policy domains, we were able to send targeted outreach and reminder messages to survey recipients during the survey activation period. This helped us to maximize our survey coverage, measured by participation at the level of country-policy domain dyad.

¹⁵ This figure includes foreign policy as a 24th policy domain. Foreign policy experts were treated as policy generalists for the purposes of the survey questionnaire.

¹⁶ Survey participants were asked about their experience within in one of 23 policy domains, which were then aggregated into four policy areas: *economic* (macroeconomic management; finance, credit, and banking; trade; business regulatory environment; investment; labor; energy and mining; and infrastructure), *governance* (land, decentralization, anti-corruption and transparency, democracy, civil service, justice and security, tax, customs, and public expenditure management), *social and environmental* (health, education, family and gender, social protection and welfare, environmental protection, and agriculture and rural development), and *general* (foreign policy and general policy).

Figure 1. Who Participated in the 2014 Reform Efforts Survey?



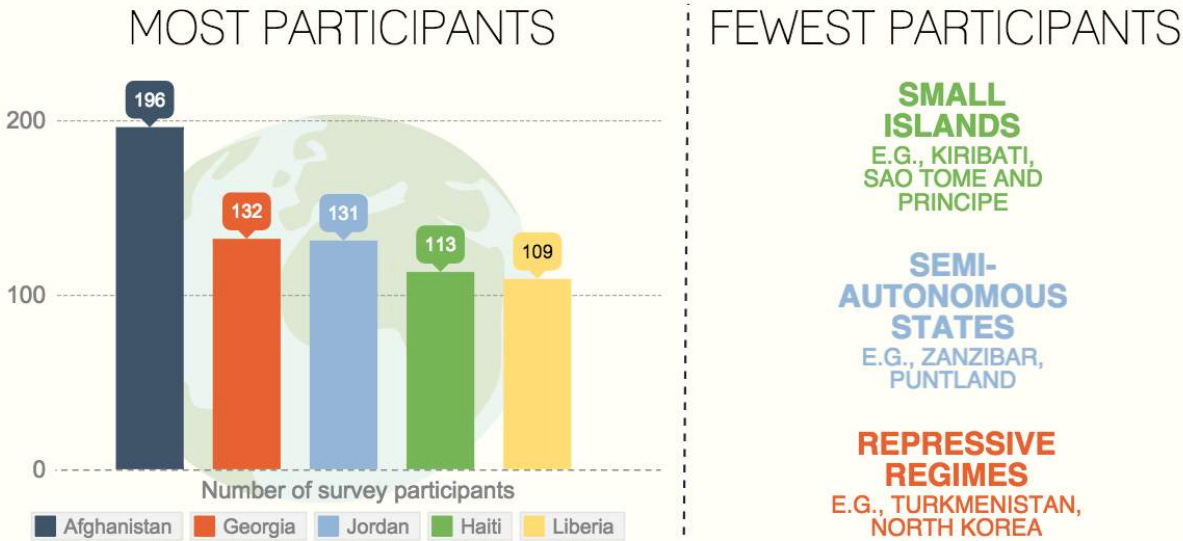
Note: The General/Foreign policy area is comprised both of participants who did not indicate a particular area of focus (17%) and of foreign policy experts (5%), who we treat as generalists. The percentage figure listed next to each specific policy domain indicates the share of all survey participants who selected that policy domain as their primary area of expertise.

SEX



Note: The predominance of male, as opposed to female, officials in both the sampling frame and participant sample is reflective of the male-dominated world of public sector staffing in developing countries. It is also important to note, as discussed in Section 2.5.1, that though there is a statistically significant difference between our male and female survey participation rates, this difference does not result in a substantially different male-to-female participant ratio vis-a-vis the sampling frame or the recipient sample.

COUNTRIES



Note: The countries shown in left-hand "Most Participants" graph: Afghanistan, Georgia, Jordan, Haiti, and Liberia, represent the five countries with the largest shares of survey participants, out of all 126 sample countries.

2.5.1. Survey Participation by Sex, Country, Stakeholder Group, and Institution Type

One of the most vexing challenges in elite survey research is the difficulty of assessing the representativeness of the data that are collected. Elite survey researchers rarely disclose detailed information about their samples and populations of interest (Hoffmann-Lange 2007). This lack of transparency makes it virtually impossible to evaluate the representativeness of the data that are collected. We confronted this challenge by clearly defining a population of interest with a standardized and transparent set of inclusion criteria (see Appendix D). Prior to fielding the *2014 Reform Efforts Survey*, we gathered data on several key attributes for the entire population of interest—that is, the 54,990 individuals in the sampling frame. As such, we have data on the sex, country, stakeholder group, and institution type for both the survey participant sample *and* the population.

Here we briefly summarize our topline findings on the representativeness of the sample vis-à-vis our population.¹⁷ However, in Appendices A and B, we provide a more detailed analysis of the representativeness of our sample.

4,976 males and 1,768 females participated in the *2014 Reform Efforts Survey*, resulting in a male-to-female ratio among participants of 281:100. This ratio is significantly higher than the 101:100 male-to-female ratio in the general world population; however, it reflects the patriarchalism of public sector staffing in developing countries (WB 2011, 2014). The male-to-female ratio among survey participants is only slightly higher than that of both sampling frame members and survey recipients.¹⁸ We can be confident that our sample is largely representative of the population of interest on this dimension.

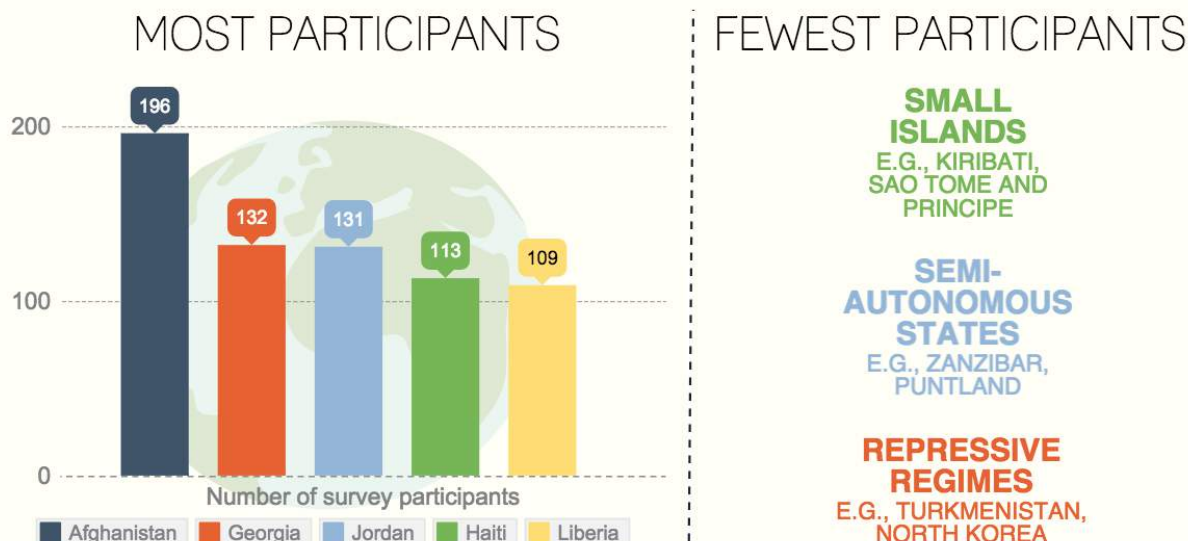
We can also assess the cross-country representativeness of our sample. While there was significant variation in participation rates by country, we did obtain relatively high numbers of responses from nearly all sample countries. The five countries with the largest number of survey participants were Afghanistan, Georgia, Jordan, Haiti, and Liberia, while the countries with the smallest number of survey participants were Zanzibar, Puntland, Equatorial Guinea, Somaliland, and Cuba (see Figure 2). Small islands, semi-autonomous territories, and highly repressive regimes constituted the smallest country strata in the sampling frame, and we find the cross-country composition of our sample largely mirrors this pattern. Perhaps most importantly, we find a strong, positive, and statistically significant correlation between the percent of sampling frame members and the percent of survey participants, which suggests that, despite country-specific deviations, our sample is largely representative of the population.¹⁹

¹⁷ In [Appendices A and B](#), we consider the representativeness of our sample vis-à-vis the sampling frame and the cohort of individuals who received an invitation to participate in the survey.

¹⁸ We do find a statistically significant difference in the participation rates of our male and female survey recipients; however, this is likely an artifact of the exceptionally large size of our survey recipient and participant samples. Additionally, as shown in Figure 1, there is not a substantive difference between our male-to-female ratios of sampling frame members and survey participants.

¹⁹ A Pearson's correlation coefficient is equal to 0.782. We also find a strong, positive, and statistically significant correlation between the number of survey recipients and the number of survey participants by country. See [Appendix A](#) for a more detailed breakdown.

Figure 2. Countries with the Highest and Lowest Number of Participants



Note: The countries shown in left-hand "Most Participants" graph: Afghanistan, Georgia, Jordan, Haiti, and Liberia, represent the five countries with the largest shares of survey participants, out of all 126 sample countries.

Within each country, our population of interest consists of five stakeholder groups: host government officials, development partner staff, CSO/NGO leaders, private sector representatives, and independent country experts. Analysis of the sample vis-à-vis the population finds that we have lower than expected response rates among host government officials and private sector representatives, and higher than expected response rates among development partner staff, CSO/NGO leaders, and independent experts. Thus, the results presented in this report may slightly over-emphasize the views of development partner and CSO/NGO officials and understate those of their host government and private sector counterparts. Nevertheless, the substantive impact of these differences in stakeholder group-specific participation rates on our results and conclusions is likely minimal.²⁰ The survey results consistently suggest the need for an increased role of host governments in external assessment development, measurement, and implementation, a finding that likely would not occur if our results systematically excluded the opinions of host government officials. Instead, our survey results reflect the views of a broadly representative cross-section of survey participants across each of the five stakeholder groups, including host government officials.²¹

Additionally, the systematic and transparent way in which the sampling frame was assembled—in particular, the mapping of functionally-equivalent, country-specific institutions back to generic institution types—makes it possible to examine the extent to which the survey participant sample and sampling frame mirror each other at the institution type level.²² Appendix D provides a detailed breakdown of the

²⁰ As with sex, these statistically significant deviations from expected participation rates by stakeholder group may also be the unintended mathematical result of our exceptionally large recipient and participant samples.

²¹ Independent experts had the highest survey participation rate (25.40%), followed by CSO/NGO representatives (21.95%), development partner staff (17.60%), host government officials (13.14%), and, lastly, private sector representatives (12.22%). Survey attrition followed similar, though not identical, trends by stakeholder group. CSO/NGO representatives registered a 68.56% survey completion rate, followed by independent experts (68.40%), host government officials (62.11%), development partner staff (60.96%), and private sector representatives (50.16%).

²² Readers will recall that our sampling frame was created by mapping country-specific institutions (and leadership positions within those very same institutions) back to an "ideal-typical" developing country government consisting of various "institution types," such as a Ministry of Finance/Economy, a Supreme Audit Institution, and a National Statistical Office.

sample and sampling frame by institution type; across all of the five stakeholder groups, the data that are presented demonstrate that our sample of survey participants is broadly representative of our survey recipients at the (functionally-equivalent) institution type level. Additional analysis within each stakeholder group also reveals that we benefitted from higher than expected participation among traditionally under-surveyed groups— for example, leaders of democracy and anti-corruption NGOs and senior government decision-makers in public procurement agencies, central banks, and aid effectiveness units.

2.5.2 A Balanced and Diverse Group of Development Policymakers and Practitioners

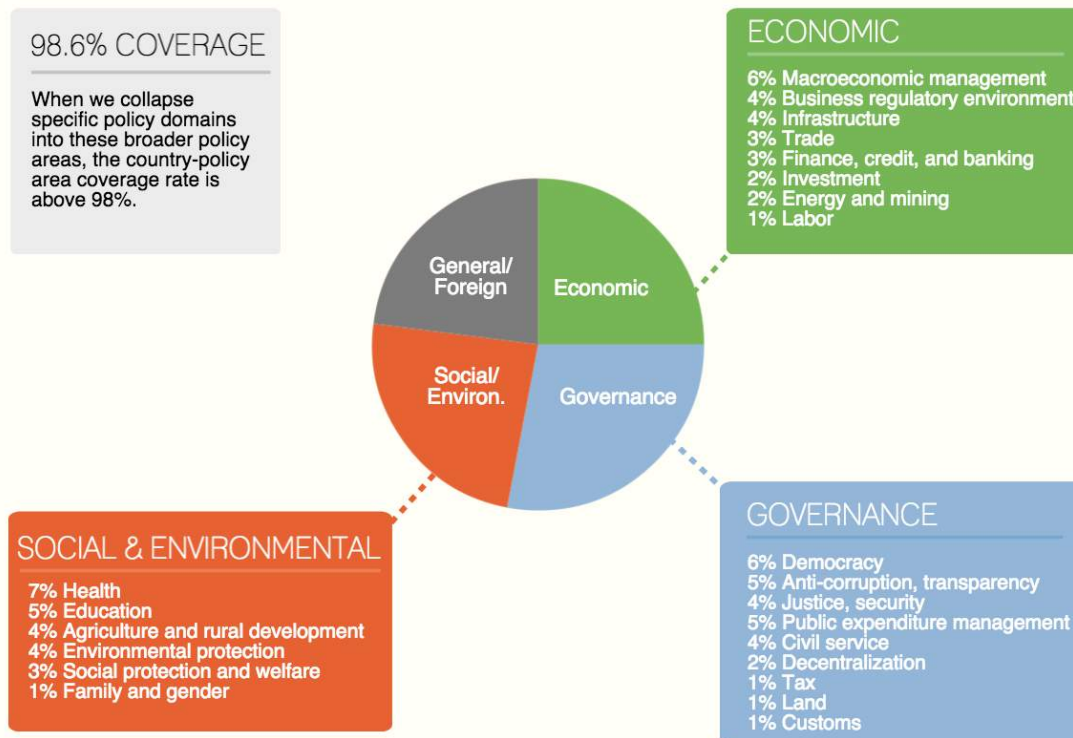
We collected a significant amount of additional demographic data from the individuals who *participated* in the survey, which allows us to explore whether the characteristics of survey participants (e.g., professional and educational backgrounds, areas of policy expertise, roles and responsibilities in the workplace) are skewed in ways that might bias our survey results in an unhelpful manner. We find significant evidence that the survey findings presented in this report reflect the experiences and perspectives of a balanced and diverse set of development policymakers and programmatic experts.

Survey participants were asked about their experience in one of 23 policy domains, which can be aggregated into four broader policy areas:

- *Economic* (including: macroeconomic management; finance, credit, and banking; trade; business regulatory environment; investment; labor; energy and mining; and infrastructure);
- *Governance* (land, decentralization, anti-corruption and transparency; democracy, civil service, justice and security, tax, customs; and public expenditure management);
- *Social and environmental* (health, education, family and gender; social protection and welfare; environmental protection, and agriculture and rural development); and
- *General* (foreign policy and general policy).

While there is significant variation in response counts across the survey participants' policy domain expertise, we observe a roughly even distribution of survey participants across the four aggregated policy areas: economic (25%), governance (28%), and social and environmental (24%), and general (22%). See Figure 3.

Figure 3. Policy Domain Expertise of Survey Participants



Note: The General/Foreign policy area is comprised both of participants who did not indicate a particular area of focus (17%) and of foreign policy experts (5%), who we treat as generalists. The percentage figure listed next to each specific policy domain indicates the share of all survey participants who selected that policy domain as their primary area of expertise.

The *2014 Reform Efforts Survey* benefited from the relatively high participation of senior managers and executives, across all stakeholder groups. We received responses from 47 heads of state or government, 59 chiefs of staff, 249 heads of government ministries and agencies, 139 vice minister-level officials, and 193 secretary general-level officials. Among development partner officials, 242 ambassadors, country directors, and other “heads of mission” participated in the survey (approximately 19% of development partner participation). Within the CSO/NGO and the private sector stakeholder groups, heads of organizations accounted for approximately 58% and 51% of survey participants, respectively.

Yet, this high level of input from political, administrative, and managerial leadership does not seem to compromise the representativeness of our sample of survey participants. Directors and head of technical units and departments, technical specialists, and program managers, accounted for nearly 61% of host government survey participants, while directors, technical specialists, and program managers made up a combined 55% of the development partner participant group.²³

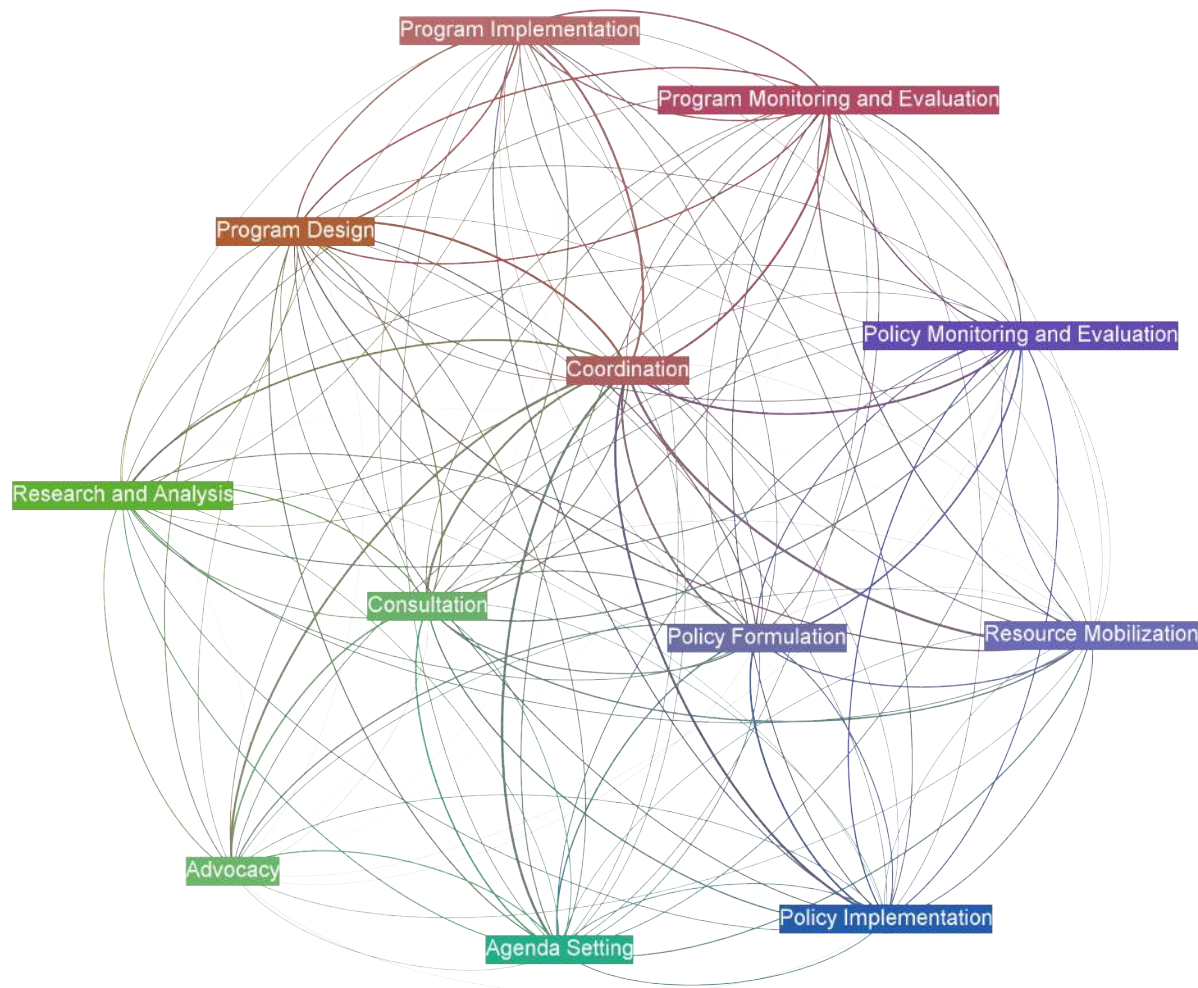
It is also important to note that the official titles and positions of our survey participants do not fully capture the nature of the roles that they played in shaping government policies and programs.²⁴ Figure 4

²³ Only 11.35% of the host government officials, 7.51% of the development partner staff, 9.94% of the CSO/NGO representatives, and 12.94% of the private sector representatives who participated in the *2014 Reform Efforts Survey* indicated position types that fell outside of those used to determine inclusion in the global sampling frame.

²⁴ We asked participants to identify all of the activities they participated in on an average day while holding a specific position of their choice. Each participant was asked to select all applicable activities from a fixed list of 12 options: research and analysis, agenda setting, advocacy, consultation, coordination, resource mobilization, policy formulation, policy implementation, policy monitoring and evaluation, program design, program implementation, and program monitoring and evaluation.

summarizes the various roles played by survey participants and the relationships between those roles. This network of activities has a core-periphery structure, with three central roles played by survey participants: coordination, consultation, and policy formulation.²⁵

Figure 4. Roles Played by Survey Participants in the Policy Process



Note: Nodes/boxes correspond to response option "activities" selected in question 10. A tie/line between two activities represents the relative frequency with which a single survey participant indicated participating in both activities. Ties are directed and weighted by the percentage of participants engaged in one activity who were also engaged in a second, linked activity. Activities that are closer to one another on the graph were selected together more frequently than activities that are farther apart.

Several additional findings merit discussion. Individuals who played a coordination role seem to have been more involved in programmatic activities, while those who played a consultation function were more involved in research and advocacy. Survey participants also appear to have specialized according to the nature of their activities (i.e., policies and programs), but not by the stages of the reform process (i.e.,

²⁵ More than 80% of participants indicated being regularly involved in coordination efforts, 67% engaged consultative activities and 65% played some role in the policy formulation process. The network centrality of individuals who play a coordination or consultation function reinforces the finding that so-called "connectors" and "conveners" exert outsized leadership during both reform design and implementation (Andrews 2013).

agenda-setting, design, and implementation).²⁶ Agenda setting appears to involve a mix of consultation, advocacy, and policymaking activities, while program design is viewed as the application of research and analysis to inform program implementation. Finally, resource mobilization seems to be more central to rolling out new policies than new programs.

2.5.3 Longer-Term Perspectives and Shorter Term-Insights

The *2014 Reform Efforts Survey* benefited from both the longer-term perspectives of host government officials and the shorter-term insights of development partner staff.²⁷ Nearly 38% of host government officials reported having at least twenty years of experience working with or for their country's government and an additional 37% of officials reported having at least ten years of experience. In comparison, participants from development partner institutions have substantially fewer years of experience working on policy and programmatic issues in the country of interest. Forty-three percent of development partner participants reported having fewer than five years of experience working in the country about which they answered survey questions. This pattern likely reflects the high mobility and turnover rates associated with in-country donor agency, embassy, and international organization personnel.

2.5.4 Survey Participant Interactions with Development Partners

Survey participants also reported extensive experience working alongside a large number of Western and non-Western development partner agencies. Large multilateral and bilateral development partners (e.g., the WB, the EU, UNDP, USAID, and the UN) were among the most frequently selected development partner organizations. A significant number of survey participants also indicated having worked with non-Western development partner embassies, including embassies from: China, India, Turkey, Brazil, and South Africa. A smaller number selected non-Western development agencies (e.g., TIKA, China Ex-Im Bank, Kuwait Fund for Arab Economic Development, BADEA, and China Development Bank).²⁸

We also examine the relationship between a participant's *overall* level of interaction with development partners and his or her perception of the average influence of external assessments of government performance: the primary outcome of interest in this study.²⁹ Overall, we find that there is no significant relationship between a participant's level of interaction with development partners as a group and his or her overall perception of external assessment influence.³⁰

²⁶ This lack of participant differentiation between the stages of the policymaking process is a pattern we discuss at various points throughout this report.

²⁷ It is also worth noting the significant country expertise of our private sector representatives, CSO/NGO leaders, and independent country experts. More than 70% of CSO/NGO participants possess ten or more years of experience in the countries about which they answered survey questions. Nearly 60% of private sector participants and nearly 80% of independent country expert participants had comparable levels of country-specific experience.

²⁸ The development-partner interaction reported by our survey participants almost certainly underestimates the true level of first-hand development partner knowledge and experience they possess. Each participant's selection of development partners was bound to a single, specific position, rather than an entire career.

²⁹ Average external assessment influence was calculated at the individual participant level by averaging his or her agenda-setting (or reform design) scores for all of the assessments that he or she evaluated. Interaction with development partners was calculated by taking the natural log of the number of development partners with whom a participant interacted (plus one to avoid multiplication by zero), and multiplying it by the self-reported frequency with which he or she interacted with those same development partners. External assessment influence scores were drawn from questions 31 and 32, and the development partner interaction data were drawn from questions 12 and 13 (see [Appendix E](#) for survey question text).

³⁰ A statistically significant relationship between a participant's level of interaction with development partners and his or her perception of external assessment influence is limited to two small sub-samples: the CSO/NGO and private sector stakeholder groups. This relationship suggests that members of these two stakeholder groups, who are less directly involved in the formulation and implementation of government policies and programs, are more likely to think external assessments of government performance are influential if they have had more exposure to development partners. Nevertheless, these correlations wash out at the overall sample level, so there should not be much cause for concern regarding overall bias on account of development partner exposure. In any case, we view participant interactions with a particular development partner as endogenous to the eventual influence of that development partner's assessments. These interactions can be seen as part of that development partner's outreach efforts, the effects of which are more directly measured through an analysis of assessment familiarity, which we explicitly investigate in this report.

2.5.5 The Development Partner Work Histories of All Survey Participants

We also asked participants to identify, over their entire career, all development partners for which they had worked as a: full-time employee, part-time employee, consultant, and/or in any other capacity. Two patterns are immediately apparent: (1) a very large percentage of survey participants previously worked for Western multilateral and bilateral development agencies; and (2) far fewer survey participants worked for non-Western development partner institutions. The most frequently cited development partners are The WB, UNDP, USAID, the UN, and the EU. Among bilateral donors, the US emerged as the leading source of previous employment. Nearly 30% of survey participants reported having worked for USAID, a US Embassy, or the MCC.

There is remarkable consistency in development partner work history patterns across participants from the five stakeholder groups. Not only do the WB, UNDP, USAID, the EU, and the UN account for five of the top six employers of survey participants in each stakeholder group, but roughly the same percentages of survey participants from each stakeholder group also report having worked for at least one of the five organizations.³¹

For more detailed information about the attributes of the survey participants and discussions of representativeness and bias, please refer to [Appendices A and B](#).

³¹ This uniformity likely reflects the heavy, historical reliance of Western development partners on “local hires”. It raises the question of if and whether Western development partners exert significantly greater “soft power” than non-Western development partners in these countries because they have trained a disproportionate number of in-country policymakers over a long period of many decades (see Dong and Chapman 2008; Chwieroth 2007, 2009b; and Atkinson 2010, Weymouth and Macpherson 2012).

3. External Assessment Influence: Global-Level Findings

The *2014 Reform Efforts Survey* provided each participant with an opportunity to evaluate the influence of a personalized list of external assessments of government performance based upon his or her country, policy domain expertise, and specific years of service in a given position.³² Participants were asked to rate the influence of only those external assessments with which they were familiar on a scale of 0 to 5, with 0 indicating no influence at all and 5 indicating maximum influence.³³ Their evaluations of assessment influence were also restricted to a particular country and a self-identified set of three reform efforts in one of 23 different policy domains (e.g., health, education, and public financial management).³⁴

Our analysis of the survey data on external assessment influence calls attention to **five global-level findings**, which we will preview here and discuss at length in this section:

1. External assessments can influence reform efforts, but some are far more influential than others.
2. External assessment influence is strongest at the agenda-setting stage of the policymaking process.
3. In a crowded marketplace of policy ideas, assessment familiarity delivers an influence dividend.
4. Influencing reform efforts from the outside is substantially easier in some policy domains than in others.
5. Geography matters: assessments achieve widely varying levels of uptake across regions and countries.

3.1. External assessments can influence reform efforts, but some are far more influential than others.

The process of making policy, institutional, legal, regulatory, and administrative changes in the public sector typically consists of different stages—including agenda setting, reform design, reform adoption, reform implementation, and reform institutionalization. External assessments of government performance can have varying levels of influence at these different stages of the process (Grindle 2004; Fox et al. 2007). In this study, we particularly focus on differences and similarities of assessment influence at the agenda-setting and reform design stages.³⁵

3.1.1. External Assessment Influence at the Agenda-Setting Stage

One of the most fundamental ways that external assessments can exert influence is by shaping the government's policy agenda—by influencing which reforms will be prioritized and which reforms will not be prioritized (IEG 2008; McArthur 2013; and Parks 2014). To this end, we asked participants to indicate the degree to which individual assessments influenced a government's decision to pursue a specific set of (participant-identified) reforms.³⁶

³² The process by which external assessments of government performance were identified ex ante and routed to individual participants—based upon their country, policy domain, and specific years of service—is described in [Appendix F](#).

³³ We only included a given assessment in our global analysis if it had sample size of at least 10 observations. At any level of aggregation below the global level, we included assessments with a sample size of at least 5 observations.

³⁴ For example, if a participant worked in Afghanistan from 2005-2007 on health policy issues, he or she was only prompted with questions about external assessments that measured the health policy performance of Afghanistan between 2005 and 2007. Individuals who selected either foreign policy or no area of specialization were asked about assessments that pertained to policy in general or a broad set of policy issues.

³⁵ Survey pre-testing demonstrated that, as far as the role of external assessments is concerned, our sampling frame members do not distinguish between reform design and reform implementation.

³⁶ Participants were asked in an open-ended question (question 20) to describe three specific problems—within their policy domain of expertise—that the government attempted to solve through its reform efforts. Then, participants were asked to evaluate the influence that each assessment in a given set of assessments had on the government's decision to pursue reforms geared towards solving the problems they described.

Some external assessments of government performance proved far more influential in setting the reform agenda than others (see Figure 5). Survey participants broadly agreed that assessments produced by the EU had the greatest influence during the agenda-setting stage—a finding that holds true across the five stakeholder groups included in this study.³⁷ Two EU-produced assessments were among the top 10 most influential assessments, as rated by survey participants: the EU’s Association Agenda³⁸ and the EU’s Stabilization and Association Agreements.³⁹ Both assessments offer countries being assessed significant financial and economic rewards in exchange for domestic reforms (Cadier 2013; Gavriliis 2014; and Gamurari and Ghinea 2014).⁴⁰ Apart from the lure of extrinsic material incentives, these EU-sponsored assessments can exert policy influence through persuasion, emulation, socialization, and competition (Jacoby 2006; Johnson 2008; and Böhmelt and Freyburg 2015).⁴¹

³⁷ We disaggregated these results by stakeholder group in order to reveal differences of opinion within the sample.

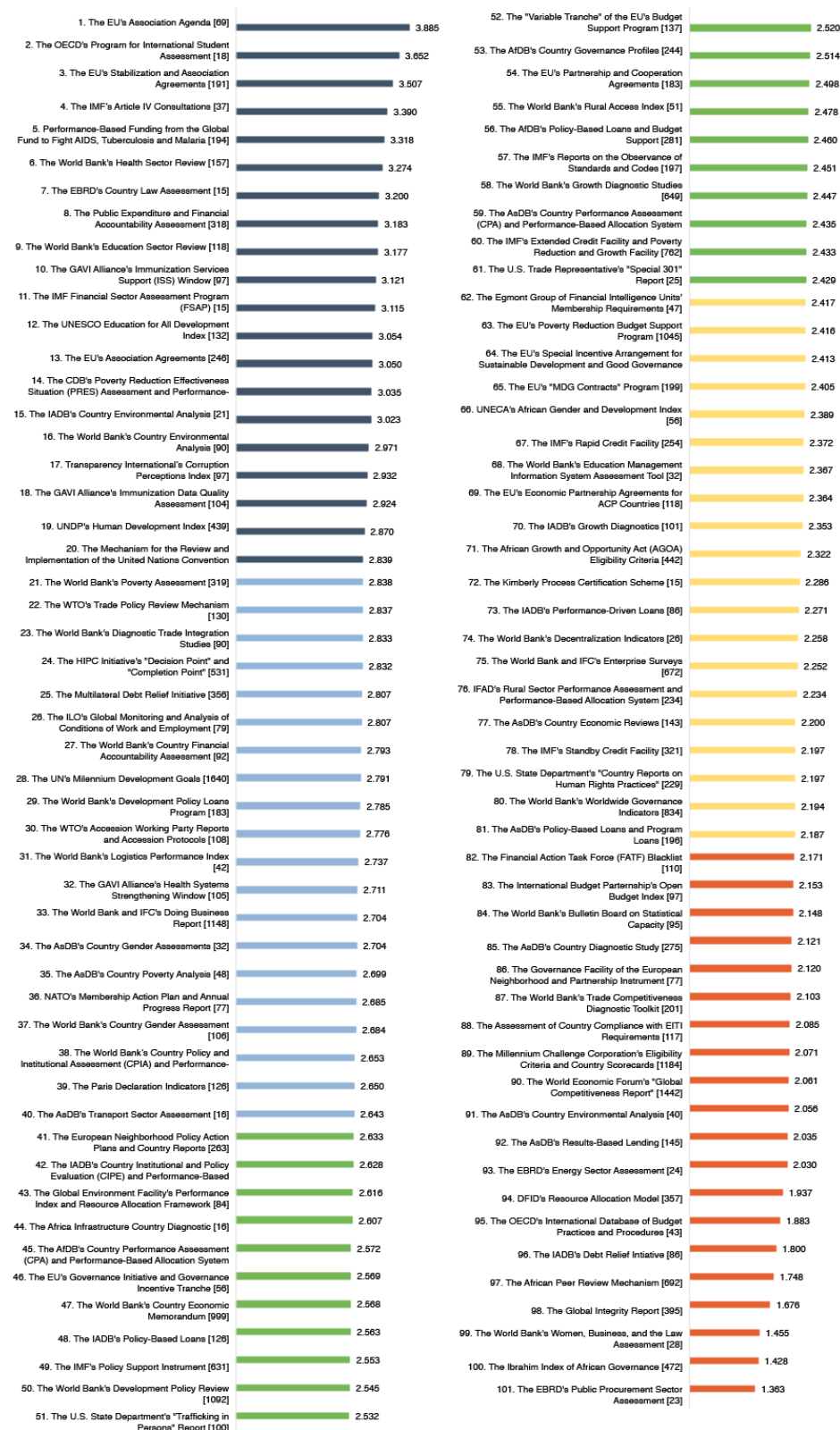
³⁸ The EU’s Association Agenda is, in effect, a variant of the European Neighborhood Policy Action Plan. However, it only applies to a small number of the EU’s “Eastern partners” (Georgia, Ukraine, and Moldova).

³⁹ EU Stabilization and Association Agreements are treaties between the EU and non-EU countries that provide a framework for monitoring the implementation of reforms in a wide variety of policy domains. The non-EU countries that participate in these agreements can receive financial and economic benefits—most notably, access to the EU market—in exchange for the implementation of reforms (Stahl 2011; Vachudova 2013).

⁴⁰ Some of the “external assessments of government performance” are organizational membership accession programs or performance-based aid, trade, or debt relief programs with explicit or implicit assessments of performance.

⁴¹ The geographic scope of these assessments is limited. The calculation of influence scores for the EU’s Association Agenda was limited to three countries (Georgia, Moldova, and Ukraine), while the calculation of influence scores for the EU’s Stabilization and Association Agreements was limited to seven countries (Albania, Bosnia and Herzegovina, Bulgaria, Kosovo, Macedonia, Montenegro, and Serbia).

Figure 5. External Assessment Influence at the Agenda-Setting Stage



Note: In question 31 of the survey questionnaire, participants were asked to evaluate -- on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence" -- the amount of influence that specific external assessments had on a government's decision to pursue reforms focused on three specific problems. These three problems were identified by the survey participants themselves in a prior, open-ended question (question 20). The estimates presented in this chart are presented in rank order by quintile, and the location of an assessment within a specific quintile indicates the relative influence of that assessment vis-a-vis other assessments. The number of observations pertaining to each assessment is in brackets. The following assessments were excluded from analysis because they did not meet a sample size threshold of at least 10 observations: The IADB's Citizen Security Sector Note, The IADB's Education Sector Note, The IADB's Social Protection Sector Note, The IADB's Trade Sector Policy Note, and The IADB's Transport Sector Note.

Assessments produced by several other multilateral organizations also featured prominently in the agenda-setting stage. Topping our list of agenda-setting assessments is the PISA, which benchmarks countries on the educational performance of their fifteen-year old children in the areas of math, science, and reading. While supplier of the assessment—the OECD does not offer the countries being evaluated any financial or economic reward for improved performance; PISA has become “an indicator of the success or failure of education policy” in a given country and has “enable[d] policymakers to make the case for reforms...that would otherwise be contested” (Grek 2009: 26, 35). Two IMF assessments of government performance were also in the top 10 at the agenda setting stage: Article IV Consultations and the Financial Sector Assessment Program (see Box 1).⁴²

Rounding out the top 10 agenda-setting assessments were: WB's Health Sector and Education Sector Reviews; the EBRD's Country Law Assessment; performance-based funding from the Global Fund to Fight AIDS, Tuberculosis, and Malaria; the PEFA's Assessment; and the GAVI's Immunization Services Support Window.

⁴² The FSAP has always been considered an “input” into Article IV surveillance, but in 2010 the IMF made a historic decision to require that all Article IV surveillance of “jurisdictions with systemically important financial sectors” include the financial stability information from a country's FSAP assessment (IMF 2014: 7).

Box 1. External Assessments and the IMF

The outsized agenda-setting influence of the IMF's Article IV Consultations will likely not surprise careful observers of the institution. Article IV Consultations are the primary mechanism by which the IMF conducts policy surveillance of its member states (Simmons 2000; IMF 2006a: 7). The "IMF's Executive Board conducts Article IV consultations with each member country once a year or every two years, based on staff reports that summarize recent developments and discussions with the national authorities" (IMF 2007: 8). The IMF uses this performance information to distribute funding via various programs and award the organization's coveted "good housekeeping seal of approval" (Momani 2007; Dreher et al. 2008; and Lombardi and Woods 2008).

The strong performance of the IMF's FSAP may prove more surprising to those who follow the Fund's activities.⁴³ FSAP is a less well-known program that provides in-depth, country-specific analysis of a country's financial sector. However, its high level of reported agenda-setting influence falls in line with a 2006 evaluation conducted by the Independent Evaluation Office of the IMF, which concluded that "[t]he greatest impact [of the FSAP] has been on within-government dialogue and in supporting the authorities' position in discussions with the legislature". The evaluation further noted that "[our] in-depth examination of the 25 country cases has identified a wide range of cases in which significant [policy] changes did take place subsequent to the FSAP and in which there is some evidence that the FSAP was at least a contributory factor" (IMF 2006b: 35). This finding is also consistent with a 2014 IMF review of the FSAP program, which sought to understand "the degree to which the exercise is seen as useful by the authorities and the effect the FSAP has *in shaping the domestic policy agenda*" (IMF 2014: 23, emphasis added). The review concluded that the FSAP "had been useful, particularly in raising awareness of financial stability issues, encouraging discussion...between government agencies and [with] financial sector stakeholders, and improving the broader policy dialogue with the Fund" (IMF 2014: 23).

The *least* influential assessment at the agenda-setting stage was the EBRD's Procurement Sector Assessment. This relatively new assessment was launched in 2010 and many participants reported low levels of familiarity with it.⁴⁴ Other assessments reported to have low levels of agenda-setting influence, included: the Ibrahim Index of African Governance (see Box 2), the Global Integrity Index, the IBP's Open Budget Index, the African Peer Review Mechanism, and the Governance Facility under the ENPI.⁴⁵ The limited influence of these governance assessments may be partly explained by their substantive focus: developing country leaders may be less interested in introducing reforms that will fundamentally challenge vested interests or disrupt the prevailing political order (Parks and Rice 2013; Smets 2014).

⁴³ While the FSAP is very closely associated with the IMF, it is often conducted in collaboration with the WB. When the IMF and the WB work together on a given FSAP assessment, the IMF typically leads on financial stability component of the assessment, while the WB is responsible for financial development component of the assessment (IMF 2014).

⁴⁴ We measured familiarity with the EBRD's Public Procurement Sector Assessment by calculating a weighted percentage of all participants—who worked at some point between 2010 and 2013 in one of the assessment's 17 target countries and 4 policy domains—who indicated familiarity with the assessment, yielding an estimate of 19.78%.

⁴⁵ The ENPI Governance Facility ties the provision of additional funding from the EC to a country's progress on a (pre-negotiated) governance reform action plan.

Box 2. The Ibrahim Index and the Challenge of Influencing Governance Reform in Africa

The Ibrahim Index of African Governance is an annual assessment of governance in Africa produced by the Mo Ibrahim Foundation. It was conceived as a “diagnostic tool for civil society, donors and governments so that performance can be enhanced and the lives and outcomes of Africans can be strengthened” (Rotberg 2009: 113). The Foundation also sponsors the Ibrahim Prize for Achievement in African Leadership, which provides \$5 million over ten years (and \$200,000 annually thereafter) to former African heads of state and government who were democratically elected, governed their countries justly and effectively and served a constitutionally mandated term (Rheault and Tortora 2011; Ibrahim 2012).

The low agenda-setting influence of the Index reported by participants may be disappointing to many observers, as it was hoped that the assessment would facilitate much-needed governance reforms because of the *personal* financial incentive that the assessment sponsor provided to the leaders of African countries (Rosenberg 2006). There are several potential reasons why the Ibrahim Index has exerted limited influence. Perhaps most importantly, the Index addresses a complex set of issues related to the distribution of power and the nature of political order, and elsewhere in this report we demonstrate that governance assessments are systematically and substantially less likely to influence reform efforts than other types of assessments.

It is also possible that the Mo Ibrahim Foundation may have inadvertently blunted the incentive for leaders to directly address performance shortcomings by sending inconsistent signals about the relationship between the prize and improved performance on the annual Index. The link between performance on the Index and the \$5 million prize has become weaker over time. Rotberg (2009: 123-124) explains that “[a]t first, leaders eligible for the prize had to come from countries with reasonably high ranking on the Index....[which] was a necessary, but not a sufficient, bar over which the leaders were expected to hurdle. By 2008...it was clear that the number of eligible leaders for each year’s prize were so few...that the qualitative deliberations of the prize committee, chaired by Kofi Annan, were sufficient.” Existing research on the credibility of conditional contracts and the determinacy of performance conditions suggests that the decision to “short-circuit” the link between the financial reward and the performance requirements may have diminished the influence of the Index. (Schimmelfennig and Sedelmeier 2004; Stone 2002, 2004; and Öhler et al. 2012). However, it should be noted that we find only weak evidence in this report that financial incentives have a significant impact on external assessment influence.

3.1.2. External Assessment Influence at the Reform Design Stage

External assessments of government performance can also shape the way that governments *design* reforms. Once a government has established a reform priority at the agenda-setting phase of the policymaking process, the authorities need to make decisions about the nature, content, scope, depth, and timing of the changes that will be undertaken. Previous research suggests that external assessments of government performance can directly inform and influence these reform design decisions (IEG 2008; Parks 2014).⁴⁶

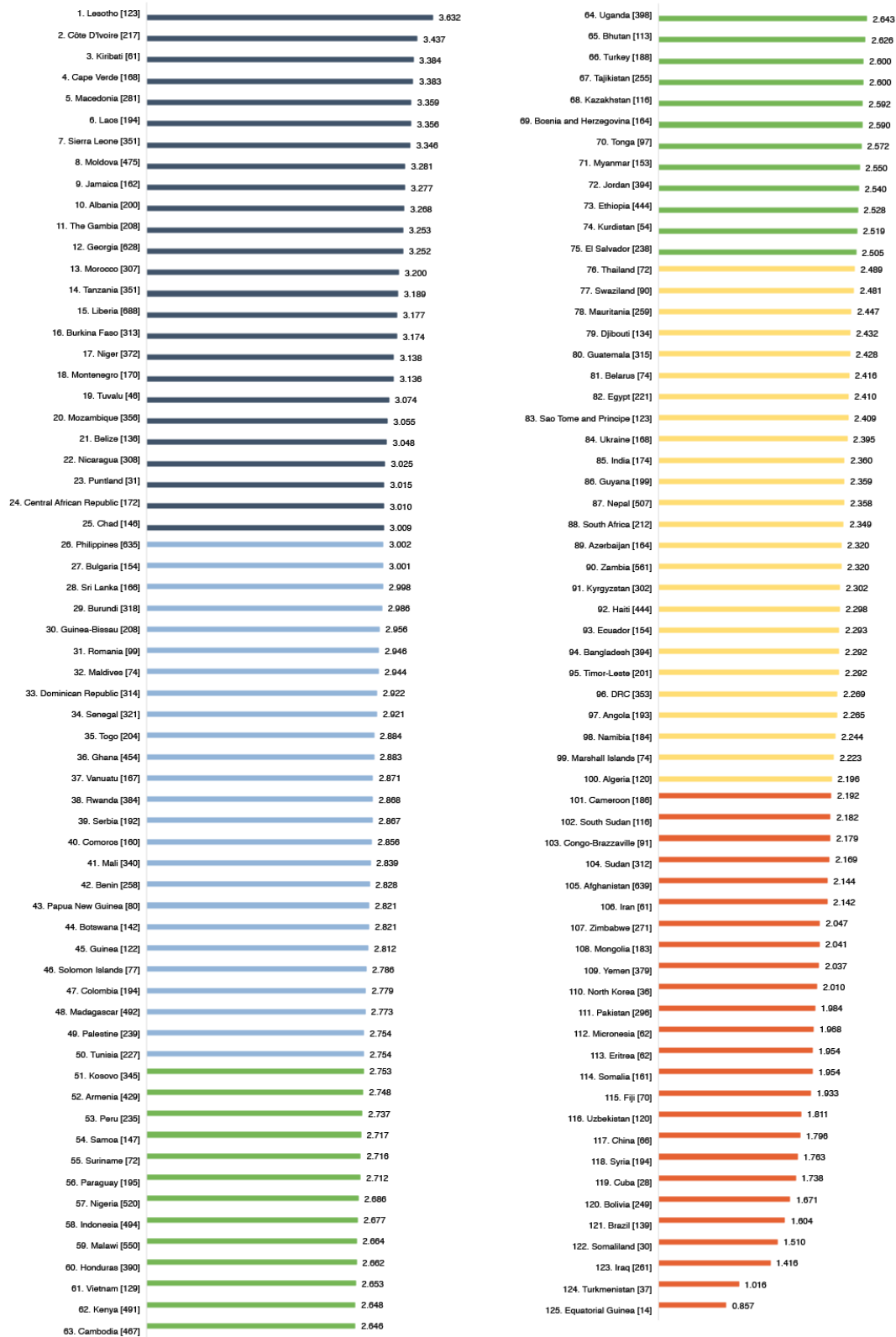
According to the survey participants, the most influential agenda-setting assessments were also among the most influential assessments at the reform design stage. The EU's Association Agenda remains the single most influential assessment at the reform design stage,⁴⁷ while the EU's Stabilization and Association Agreements also registered a high level of influence.⁴⁸ Rounding out the top 10 most influential reform design assessments are: the OECD's PISA, the IMF's Financial Sector Assessment Program and Article IV Consultations, the PEFA assessment (see Box 3), the WB's Health Sector Review, the AICD, the performance-based funding program of the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the EU's Association Agreements.

⁴⁶ For example, once a government has identified public procurement as a reform priority, it might decide among several design alternatives: overhaul its procurement law, blacklist companies guilty of major procurement violations, monitor the income and assets of public procurement officials, or invite civil society organizations to oversee the bidding and award selection processes.

⁴⁷ There is a large literature on the EU's regional impact on policy outcomes in prospective member states and other countries. See, for example, Kelley 2004, 2006; Sasse 2008; Liefferink et al. 2009; Orbie and Tortell 2009; Zimelis 2011; and Böhmelt and Freyburg 2013.

⁴⁸ The data collected on assessment influence by stakeholder group at the reform design stage looks very similar to the results for the agenda-setting stage. With the exception of the CSO/NGO cohort, every stakeholder group reported that assessments produced by the EU had the greatest amount of influence on the reform design decisions and actions of developing country governments. The CSO/NGO stakeholder group reported that the most influential assessment was NATO's Membership Action Plan and Annual Progress Report. This NATO assessment is similar to many EU assessments in that (1) it applies to many of the same "target" countries (e.g., Albania) and (2) organizational membership provides significant benefits to these countries. Like the EU, membership in NATO guarantees military protection and political cooperation between member countries (Wallander 2002).

Figure 6. External Assessment Influence at the Reform Design Stage



Note: Each participant was asked to evaluate the reform design influence of those assessments with which he or she was familiar, on a scale of 0 to 5, where 0 means "No influence at all" and 5 means "Maximum influence". The estimates presented in this chart are presented in rank order by quintile, and the location of an assessment within a specific quintile indicates the relative score of that assessment vis-a-vis other assessments. The number of observations pertaining to each assessment is in brackets. The following assessments were excluded from analysis because they did not meet a sample size threshold of at least 10 observations: The IADB's Citizen Security Sector Note, The IADB's Education Sector Note, The IADB's Social Protection Sector Note, The IADB's Trade Sector Policy Note, and The IADB's Transport Sector Note.

Box 3. Reform Design Influence and the PEFA Assessment

Participants in the 2014 Reform Efforts Survey indicated that the PEFA assessment has significantly influenced the reform design decisions of developing country governments.⁴⁹

The PEFA assessment is not a performance-based aid or debt relief program, but bilateral and multilateral donors condition the provision of budget support and other financial and economic rewards on improved public financial management, as measured by the PEFA assessment (NORAD 2008; Director General for External Policies of the EU 2011; Ministry of Foreign Affairs of Denmark 2013). PEFA is jointly sponsored and published by the World Bank, the IMF, the European Commission, the French Ministry of Foreign Affairs, the Norwegian Ministry of Foreign Affairs, the Swiss State Secretariat for Economic Affairs, and the UK's DFID. The PEFA assessment therefore serves, to some extent, as a coordination mechanism for donors who make their budget support and other forms of aid conditional upon government policy behavior (Clist et al. 2012; Knack 2014).

Host government officials often work together with international experts to produce the PEFA assessment (Reitmaier 2014). This is indicative of a broader pattern we highlight later in the report: that assessments relying on data produced by the target country tend to exhibit higher levels of policy influence in that country.⁵⁰ It also suggests that local "ownership" of an assessment is an underappreciated determinant of its resonance and uptake (Andrews 2011; Swedlund 2013; and Busia 2014).⁵¹

It is notable that the performance-based funding program of the Global Fund to Fight HIV/AIDS, Tuberculosis and Malaria and the PEFA assessment—two assessments that make it possible for governments to reap new and additional financial rewards if they improve their performance—enter the top ten list of most influential assessments of government performance at the reform design stage of the policymaking process. Advocates of performance-based financing argue that it provides reformers with the autonomy and maneuverability needed to determine how best to initiate reforms and shepherd them to completion (Meessen et al. 2011; Perakis and Savedoff 2015).⁵²

These results suggest that external assessments of government performance that make it possible for governments to reap new and additional financial rewards can be influential *in some cases*. However, it is equally important to note that a number of performance-based aid allocation and debt relief programs—

⁴⁹ Our finding that the PEFA assessment significantly influences reform design efforts in developing countries is broadly consistent with the results reported in Mackie and Caprio (2011).

⁵⁰ This finding contradicts conventional wisdom that assessment sponsors and suppliers will achieve greater policy influence if they rely on third-party data collection and policy surveillance. While third-parties may be in a better position to provide unbiased analysis and advice, our results indicate that the most influential assessments are those that rely (at least in part) on host government data collection efforts.

⁵¹ Betley (2008: 34) reports that "[t]he PEFA instrument ... appears to be valued by both governments and development partners." The PEFA website emphasizes that it is committed to "a government-led reform program for which analytical work, reform design, implementation and monitoring reflect country priorities and are integrated into governments' institutional structures" (retrieved from <https://www.pefa.org/fr/node/24> on 22 February 2015).

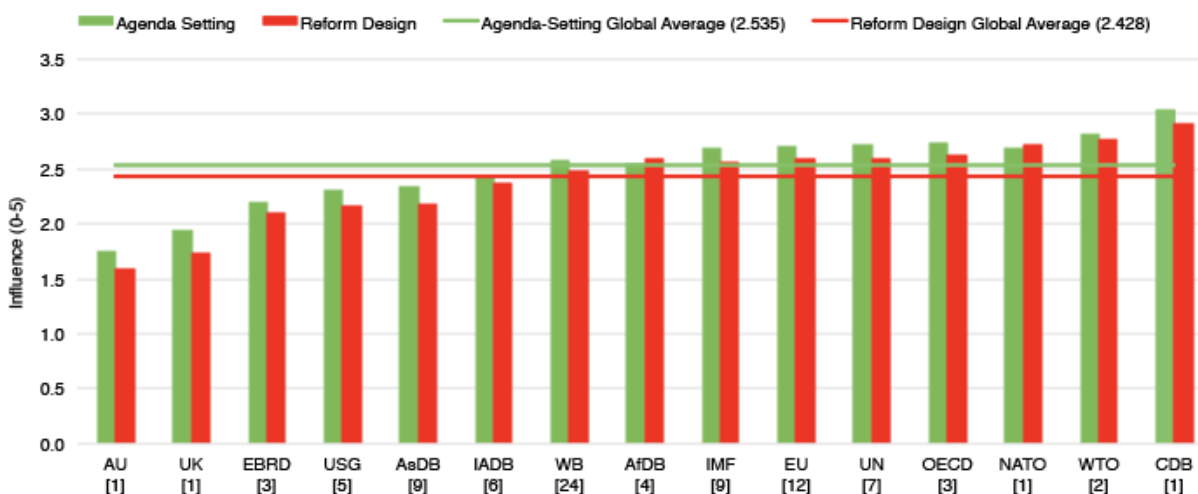
⁵² Another early adopter of performance-based health financing is the GAVI. We see in Figure 6 that its Immunization Service Support Window, Health Systems Strengthening Window, and Immunization Data Quality Assessment ranks 13th, 30th, and 33rd out of 101 external assessments of government performance, respectively. These findings of significant policy influence are broadly consistent with those reported in Lu et al. 2006; Chee et al. 2007; Naimoli and Vergeer 2010; and Perakis and Savedoff 2015.

including the MCC’s eligibility criteria, the IADB’s Debt Relief Initiative, and DFID’s Resource Allocation Model—sit within the bottom quintile of the reform design influence distribution (see Figure 6).⁵³ This pattern in the survey data is surprising and adds nuance and context to the findings that we reported in Parks and Rice (2013). There, we asked 640 development policymakers and practitioners to provide information about the *relative* influence of 18 external assessments of government performance. However, the assessments that participants were asked to evaluate consisted mostly of cross-country benchmarking initiatives and performance-based aid, trade, and debt relief programs. The *2014 Reform Efforts Survey* is different—and arguably more informative—in that it provided survey participants with an opportunity to evaluate a wider array of assessments.⁵⁴

3.1.3. Assessment Influence of Specific Development Partners

Donors and international organizations are actively engaged in the rapidly expanding market of external assessments of government performance, and as such they must compete for the attention of policymakers in developing countries. With this in mind, we sought to examine the influence of assessments supplied or sponsored by specific development partners. Figure 7 presents the influence of the average assessment supplied by specific development partners.

Figure 7. Assessment Influence of Development Partners in Agenda Setting and Reform Design



Note: For both agenda setting and reform design, 0 means “No influence at all” and 5 means “Maximum influence.” The number of assessments included in each development partner’s overall average is indicated in brackets. Only assessments with an overall sample size of at least 10 observations were included in this analysis.

There is significant variation in the average level of influence of the assessments produced by different development partners. The CDB, the WTO, and NATO top the list. The influence of the WTO and NATO is unsurprising, given that membership in these organizations provides strong incentives for candidate countries to achieve higher levels of performance (Sutherland 2008; Tang and Wei 2008; and Wallander 2002).

⁵³ This pattern is also observable on our assessment-level measure of agenda-setting influence (see Figure 8).

⁵⁴ The *2014 Reform Efforts Survey* presented participants with a diverse list of assessments, including many in-depth country-specific diagnostics.

The CDB's high level of assessment influence is unexpected; however, it may be explained by the tailored and targeted nature of the assessment included in our survey. The CDB's PRES Assessment and Performance-Based Allocation System is informed by individualized and detailed analyses of poverty in each member state that include in-depth policy recommendations, as well as strategies for implementing those recommendations (CDB 2008). To the extent that member states use these assessments as a tool for policy design, this may help explain why the CDB is the most influential assessment supplier on average during both agenda setting and reform design.

In contrast, a number of other assessment suppliers, including the IADB and the ASDB, register relatively low influence scores. Some of these suppliers of external assessments have themselves questioned whether the performance requirements of their assessments are sufficiently clear and predictable to incentivize policy change (IADB 2010). The IADB uses a suite of country institutional and policy performance evaluation (CIPE) indicators to determine the size of the concessional resource envelope that its borrowers will be able to access. In 2010, the IADB's Office of Evaluation and Oversight reviewed the WB's use of this performance-based resource allocation framework and concluded that “[w]hile in theory [there] was a possible incentive for countries to improve their CIPE scores, this incentive effect was significantly weakened by the CIPE method itself...the lack of clarity present in the CIPE ratings process...would make difficult a clear understanding of the evaluation criteria...thus limiting the possible incentive effects...from the use of performance-based allocative criteria” (IADB 2010: 4). Thus, assessment suppliers that wield low levels of policy influence and want to be more influential with the makers and shapers of policy in developing countries arguably need to broaden their aperture and think strategically about how they can compete in a fiercely competitive market characterized by high levels of “product differentiation.”⁵⁵

The relatively low levels of agenda-setting and reform design influence exerted by USG-sponsored assessments also merit special attention.⁵⁶ Some scholars have argued that the exceptional nature of the USG will render its assessments of other governments' performance particularly influential.⁵⁷ However, our survey results suggest that these claims about the importance of the identity of this assessment sponsor are either misplaced or overstated. It may be the case that USG assessments have less policy influence than others due to a lack of perceived credibility. If the USG's role as a “ruthless truth-teller” comes into conflict with countervailing (geostrategic, diplomatic, or commercial) interests and pressures, its credibility could be undermined and its influence weakened (Radelet 2003; Stone 2004; and Parks 2014).

Similarly, others have proposed that the AU is a fundamentally different type of assessment sponsor in that it is an intergovernmental body made up of many of the very same states that it assesses (Busia 2014).⁵⁸ It is therefore possible that the local resonance of its diagnostic and advisory content is higher than that of other assessment sponsors (Herbert and Gruzd 2008; Swedlund 2013; and Busia 2014). However, it does not appear that the AU's local resonance translated into high levels of policy influence from the perspective of survey participants.

⁵⁵ See [Appendix G](#) and Section 4 of this report for more details on the different attributes of external assessments of government performance.

⁵⁶ The USG produces a number of external assessments of government performance, including: the State Department's Trafficking in Persons Report, the MCC's annual country scorecards, and US Trade Representative's Special 301 Report, among others.

⁵⁷ Previous research has demonstrated that some USG assessments have effectively instigated reforms by target governments (Noland 1997; Kelley and Simmons 2014a; and Parks 2014). Some have argued that one reason why USG assessments might exert more policy influence than those produced by other assessment suppliers is the USG's large network of embassies, partner organizations, and NGOs. This network may allow the USG to tap into more relevant and credible sources of information (Kelley and Simmons 2014b).

⁵⁸ Relative to other assessments, some of the AU's assessment tools closely resemble self-assessment exercises. Busia (2014: 187) characterizes the AU-sponsored APRM process as “self-evaluation by the participating countries, validation by a group of eminent Africans, and a review of conclusions and findings by African peers.” Busia (2014: 192) also argues that the “APRM methodology can be viewed as a deliberative alternative to external assessments—conducted by donors for internal corporate use—with little or no relevance for recipient partner countries in terms of deepening democratization or accountability between the state and its citizens.”

Finally, while the EU is a relatively influential assessment supplier, it is important to note that its score reflects the average influence of *all* of the assessments it produces. The EU is notable in that some of its assessments are very influential (e.g., the Association Agenda, and the Stabilization and Association Agreements), while others are significantly less influential (e.g., the Governance Facility of the ENPI).

3.2. External assessment influence is strongest at the agenda-setting stage of the policymaking process.

Agenda-setting influence and reform design influence appear to go hand in hand⁵⁹: Assessments that impact government reform priorities also tend to affect the specific design of reforms, with relatively few exceptions.⁶⁰ However, we also find that assessment influence scores at the reform design stage are, on average, lower than at the agenda-setting stage. Between agenda setting and reform design, we observe a 4% decline in the perceived influence score of the average external assessment of government performance. We also find that an overwhelming majority (81%) of external assessments lose influence at the reform design stage.

This decline in influence from agenda setting to reform design does not appear to stem from a failure of assessments to provide specific policy guidance about how to design reforms.⁶¹ Rather, it seems to be the case that, in spite of their advisory content, external assessments exert less influence on the design of reforms than they do on the decisions of governments to prioritize particular reforms. This conclusion is consistent with research on (agricultural) policy reform, which indicates that domestic political conflict and opposition is more likely to arise during discussions about service delivery details than during the identification of policy goals (Krampe and Henning 2011).⁶² It is also consistent with an independent evaluation of the WB's and IFC's Doing Business Report, which found that "The [Doing Business] indicators have motivated policymakers to discuss...business regulation issues," but has "had less influence on the choice, scope, and design of reforms" (IEG 2008: xvi).

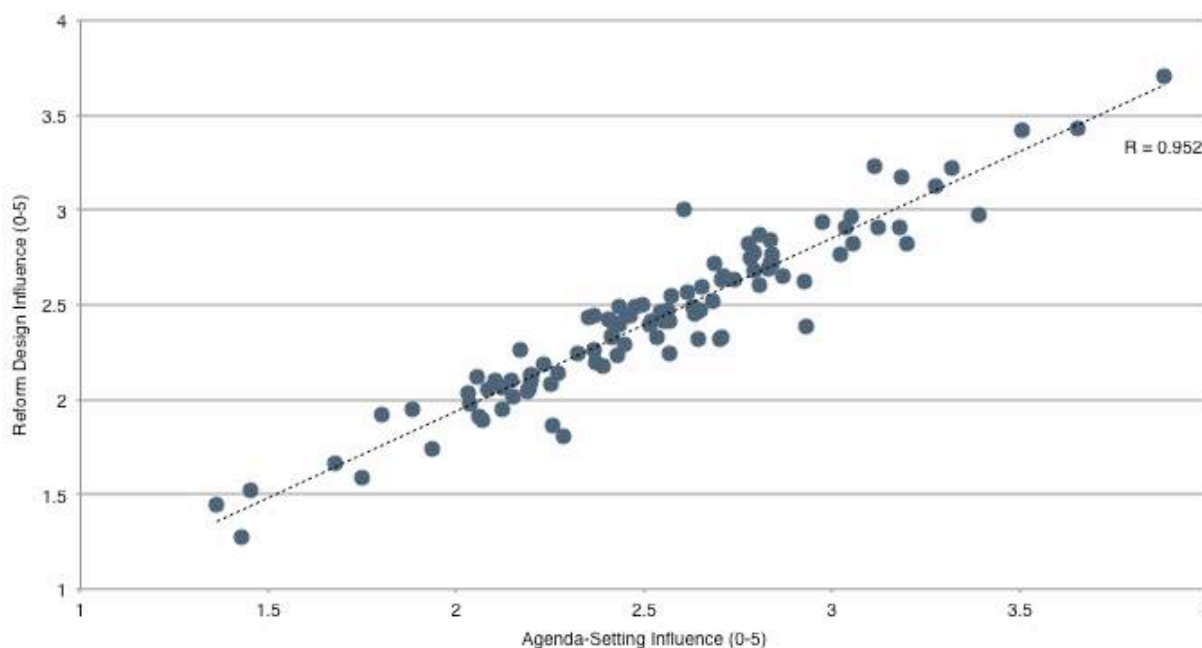
⁵⁹ Figure 8 displays weighted average agenda-setting influence scores on the x-axis and weighted average reform design influence scores on the y-axis, with each point representing an assessment. At the assessment level, there is a strong, positive, and significant correlation (0.952) between an assessment's agenda-setting and reform design influence.

⁶⁰ This strong correlation suggests that (1) participants struggled to differentiate [in retrospective accounts] between an assessment's impact on agenda setting and its role in reform design, and (2) assessment roles at each stage of the policymaking process are overlapping and—more often than not—complementary.

⁶¹ Of the 103 assessments included in the survey questionnaire, 59.22% were prescriptive in nature, providing specific guidance to the government about steps it could take to improve its performance on the assessment. However, the decline in influence of an average assessment does not seem to be impacted by its prescriptive or descriptive nature. While purely descriptive assessments declined in influence by an average of -0.109 on our scale from 0 to 5, assessments with a prescriptive component declined in influence by almost the same magnitude: -0.097.

⁶² One interpretation of diminished assessment influence at the reform design phase might be that intense domestic political conflict over the design of reform creates less policy space for external assessments to exert influence. It is also possible, however, that this finding is an artifact of the predominately technical and programmatic focus of our survey participants, who likely possess more complete—and less overstated—information about the role that assessments play in reform design processes. Nearly half (49.03%) of survey participants reported spending the plurality of their average day working on technical issues, while 20.89% focused on political matters, and only 20.47% on administrative tasks.

Figure 8. The Strong Assessment-Level Correlation between Influence in Agenda Setting and Reform Design



Note: For both agenda setting and reform design, 0 means "No influence at all" and 5 means "Maximum influence". The R-value of 0.952 indicates a very strong, positive relationship between an average assessment's agenda-setting influence and reform design influence.

Notwithstanding this overall trend, we find that all assessments did not experience the same decline in influence between the agenda-setting and reform design stages of the policymaking process. In fact, our results suggest that nineteen assessments actually *gained* influence. The ten assessments that experience the largest increases in influence, include: the Africa Infrastructure Country Diagnostic (see Box 4), the IADB's Growth Diagnostic, the IMF's Financial Sector Assessment Program, and the FATF's blacklist. Interestingly, these all of these assessments are either in-depth, country-specific diagnostics with reform design advisory content or they are associated with tangible forms of technical assistance that will help the host government authorities implement their policy advice.⁶³ This finding is consistent with previous research that suggests the provision of technical assistance can positively influence the design and implementation of reform efforts in developing countries (IEG 2008; Geertson 2010; and Parks 2014).⁶⁴

⁶³ There is also some evidence of mean reversion in assessment influence, whereby assessments that are more influential in agenda setting lose [more] influence in reform design and assessments that are less influential in agenda-setting gain—or lose relatively less—influence in reform design. A negative, bivariate correlation between the distance of an assessment's agenda-setting influence from that of the mean assessment and the change in this distance between the agenda-setting and reform design stages of the policymaking process provides evidence of this mean reversion. The Pearson correlation coefficient is -0.2413 and is statistically significant at $p=0.015$.

⁶⁴ External assessments of government performance vary significantly along this dimension: some are produced and delivered by a small group of experts who do not devote much effort to interacting with or coaching the domestic authorities in target countries, while other assessments are produced and delivered by groups of experts who are willing and able to assist the authorities.

Box 4. Gaining Influence in Reform Design with Specific Policy Advice

The AICD provides a useful example of an assessment that seems to achieve a higher level of influence at the reform design phase of the policymaking process. Responding to a call at the 2005 G8 Summit in Gleneagles for more support and attention to Africa's infrastructure sector, the AICD seeks to "assist donors in designing appropriate support for infrastructure reform... and present and disseminate conclusions and findings to inform development policy and program management activities."⁶⁵

The FATF—another assessment that gains influence at the reform design stage—places governments on a blacklist for taking insufficient action to combat money laundering and terrorist financing. However, these governments are also able to access technical assistance from the IMF, the US Treasury, the WB, and other organizations to design and implement corrective measures (Holder 2003; Ribadu 2010).

3.3. In a crowded marketplace of policy ideas, assessment familiarity delivers an influence dividend.

A potentially critical determinant of assessment influence is the degree to which policymakers and practitioners in the country being assessed are even aware that a given assessment exists. To this end, we asked participants in the *2014 Reform Efforts Survey* to select all of the assessments with which they were familiar out of personalized list of assessments based upon their policy domain, country, and set of years in which they held a particular position.⁶⁶

3.3.1. Participant Familiarity with Specific External Assessments

On the whole, we found participant familiarity with external assessments of government performance to be relatively low, with only 29% of survey participants expressing familiarity with the average assessment. The most recognized assessments were those sponsored and supplied by large development partners.

The UN produces many of the most recognized assessments, including: the MDG, UNDP's Human Development Index, and UNESCO's Education for All Development Index. The WB's Health and Education Sector Reviews and the Doing Business Report (jointly produced with the IFC) also enter the top 10 list of most recognized assessments. Performance-based funding from the Global Fund to Fight AIDS, Tuberculosis, and Malaria and three EU assessments (Economic Partnership Agreements for ACP Countries, Partnership and Cooperation Agreements, and Stabilization and Association Agreements) round out the top ten list.

The fact that participants were most familiar with the assessments produced by large development partners is likely attributable to the fact that these assessment suppliers have a broad global reach, working in many developing countries and interacting with officials in many sectors and institutions (Parks and Rice 2014).⁶⁷ Another potential explanation is that these organizations invest substantial resources in

⁶⁵ Retrieved from <http://www.infrastructureafrica.org/about> on 22 February 2015.

⁶⁶ We measured participant familiarity by calculating a weighted percentage of all participants who saw each individual assessment and indicated familiarity with it. The weighted percentage of participants indicating familiarity with each assessment was calculated one assessment at a time, so that each country and broad policy area (economic, governance, social and environmental, and general) targeted by an assessment is equally represented in the aggregate percentage score for that assessment. Write-in assessments were omitted from our analysis.

⁶⁷ All stakeholder groups tend to be more familiar with the assessments supplied and sponsored by large and high profile development partners. With the exception of independent country experts, all stakeholder groups demonstrated high levels of familiarity (at least 72.71%) with Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis, and Malaria. Similarly, all stakeholder groups—with the exception of CSO/NGO participants—demonstrated a clear familiarity (at least 58.37%)

the broad dissemination of their assessment findings and the creation of strong assessment brands. For example, new research demonstrates that the level of the WB'S engagement with their "clients" (i.e., host government officials) related to specific analytical and advisory products is a strong predictor of whether those informational inputs are eventually used by host government officials and instigate policy or programmatic change (IEG 2012; WB 2010, 2014).

Box 5. The Reach of "Doing Business" and Its Influence on the Reform Process

The WB and IFC's Doing Business Report clearly illustrates the importance of marketing, public dissemination, and engagement with policymakers in "target" countries.⁶⁸ Reportedly the WB Group's best-selling publication, the WB and IFC have put substantial resources into its marketing and distribution (ILO 2007; Biron 2013). A 2008 review by the IEG revealed that each year the WB Group spends "approximately \$1,000,000 for dissemination events...support from a communications team, and maintenance of the [doingbusiness.org] Web site" (IEG 2008: 42).⁶⁹

The review also pointed out that "Doing Business stands out among [World] Bank Group products for the variety and innovativeness of the communications tools" and that "extensive press coverage [of the Doing Business Report] attracts the interest of senior policymakers, government officials, and the business community..." (IEG 2008: 42). In fact, in 2007, doingbusiness.org "was the World Bank Group's most visited online database" (IEG 2008: 42). The same evaluation noted that, in the interest of increasing the policy influence and impact of the Doing Business Report, the IFC's FIAS "created a Doing Business Rapid Response Unit with a mandate to help countries adopt [reform] measures that are "strictly [Doing Business]-related" (IEG 2008: 46).

Interestingly, high levels of familiarity with some of the assessments produced by a given development partner do not always extend to the other assessments the institution produces. Despite strong familiarity with other EU-produced assessments, survey participants were less familiar with the "Variable Tranche" of the EU's Budget Support Program. Similarly, while the Doing Business Report is among the most recognized assessments by our survey participants, another WB Group publication, the Women, Business and Law Report, was reported to be among the least recognized assessments. Limited familiarity with this report, which has only been published since 2011, also points to the potential importance of assessment *longevity*. That is to say, it may be the case that the longer an assessment is active, the higher the likelihood that any given individual will be familiar with it (Parks et al. 2014). We test this hypothesis later in this report.

The UK's DFID Resource Allocation Model and the Rural Sector Performance Assessment and Performance-Based Allocation System produced by the IFAD are also among the least recognized assessments by our survey participants. On the surface, this finding seems counterintuitive. DFID and IFAD have substantial influence on how official development assistance is allocated across countries, and one would expect development policymakers and practitioners to be familiar with these "high stakes" assessments. One explanation for this finding is that, unlike other indicator-based aid allocation systems such as the MCC's annual country eligibility assessments, DFID and IFAD do not appear to devote much time, money, or effort to making their assessments easy for policymakers and practitioners in developing countries to access and understand (Parks 2014).

with the EU's Stabilization and Association Agreements. More information, including region and stakeholder group-specific assessment familiarity tables, is available upon request.

⁶⁸ New research undertaken at the IMF suggests that the level of public outreach associated with external assessments of government performance can also impact real economic outcomes—specifically, stock prices—in non-trivial ways (IMF 2014).

⁶⁹ For more information on the impact of expenditure on user uptake of analytical and advisory products, see Doemeland and Trevino 2014.

Rounding out the 10 least recognized assessments are: the US Trade Representative's "Special 301" Report, the IMF's Rapid Credit Facility, the Egmont Group of Financial Intelligence Unit's Membership Requirements and the IADB's Sector Notes for Social Protection, Citizen Security, and Trade.

3.3.2. The Relationship between Assessment Familiarity and Agenda-Setting Influence

Participant familiarity with an assessment appears to deliver an "influence dividend." We found a positive and statistically significant relationship between assessment familiarity and assessment influence at the agenda-setting stage of the policymaking process (see Figure 9).⁷⁰ This finding underscores the importance of effective assessment marketing, communications, and branding.

Yet, some assessments achieve significantly higher levels of influence than others, even after accounting for the role of participant familiarity. The green observations in Figure 9 represent assessments that, given their levels of participant familiarity, achieved higher than expected agenda-setting influence. These assessments include: the EU's Association Agenda, the EBRD's Country Law Assessment, the OECD's Program for International Student Assessment, the CDB's PRES Assessment and Performance-Based Allocation System, and the IADB's Country Environmental Analysis. In these cases, the salience, perceived quality, and substantive focus of these assessments may have enabled them to "punch above their weight."⁷¹

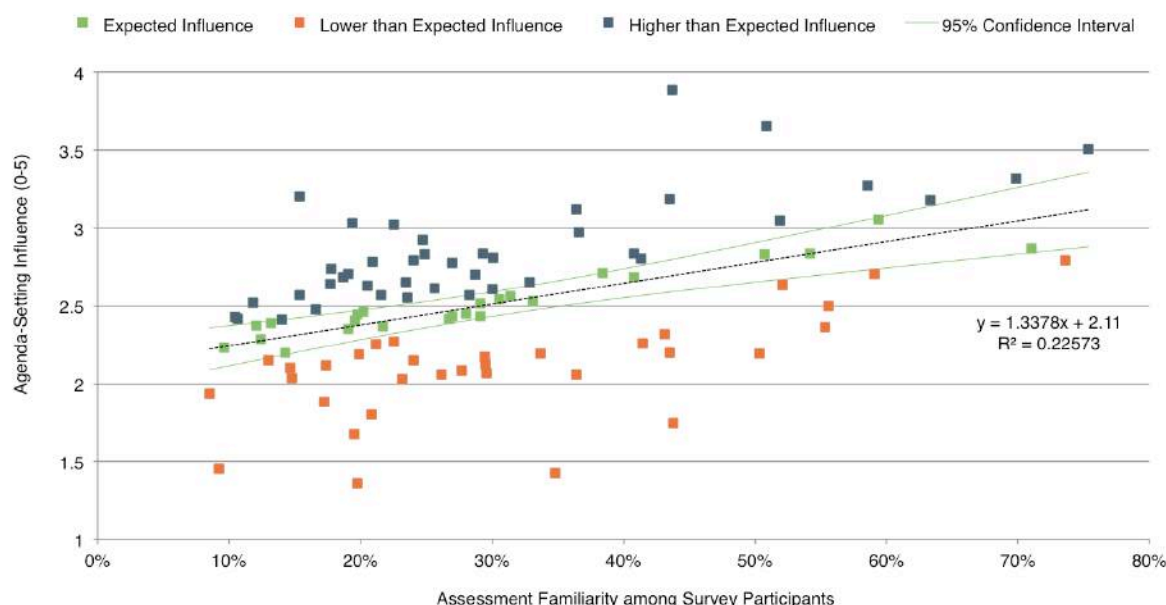
Conversely, assessments can also produce lower levels of influence than expected. The red observations in Figure 9 represent assessments that, given their degree of participant familiarity, achieved lower-than-expected influence at the agenda-setting stage. These assessments include: the Ibrahim Index of African Governance; the EBRD's Public Procurement Sector Assessment; the African Peer Review Mechanism; the WB's Women, Business, and the Law Report; and the Global Integrity Report. It is likely not a coincidence that many of these assessments focus on governance issues, which are often considered by host country authorities to be "red line issues" that fall within their exclusive and sovereign purview (Whitfield 2009; Swedlund 2013; and Busia 2014).⁷²

⁷⁰ The Pearson's correlation coefficient between these two variables at the assessment-level is 0.475.

⁷¹ On the issue of how the salience, perceived quality, and substantive focus of external assessments of government performance shape their policy influence, see Kelley and Simmons 2014b; WB 2014, 2015; and IMF 2014.

⁷² We revisit this issue later in this report.

Figure 9. The Positive Relationship between Assessment Familiarity and Agenda-Setting Influence



Note: Agenda-setting influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". The R-squared value presented in the figure indicates that approximately 22.57% of the variation in assessment agenda-setting influence is explained by variation in the percentage of survey participants familiar with each assessment.

3.3.3. Assessment Familiarity by Policy Domain and Stage of the Reform Process

Assessment familiarity varies significantly across policy domains. Table 1 provides an overview of the three most recognized external assessments of government performance by policy domain. Some assessment suppliers have been far more successful than others in reaching their target audiences. For example, our results suggest that the sponsors of the Doing Business Report—the WB and the IFC—have largely succeeded in reaching their target audiences. Approximately 80% of the participants⁷³ who identified their primary area of specialization as tax, trade, finance, credit and banking, land or business regulatory environment also indicated that they were familiar with the Doing Business Report.⁷⁴

At the same time, high levels of familiarity do not always lead to high levels of influence. The African Peer Review Mechanism, for example, registers very high levels of familiarity in many of the policy domains that fall within the scope of its country assessments (e.g., democracy, decentralization, anti-corruption and transparency). However, it wields substantially less *actual* policy influence than many of its competitor assessments. It is also worth noting that while health and family and gender rank among the policy domains that register surprisingly low levels of overall assessment familiarity, we will demonstrate that these policy domains also appear to be most influenced by external assessments of government performance (see Table 1). Assessment suppliers seeking to expand their policy influence would be well advised to take note of this apparent gap in the market.

⁷³ To be more precise, this 80% figure corresponds to the percentage of participants who (a) were routed into a country-policy domain-year set that we determined *ex ante* was evaluated by the Doing Business Report between the years of 2004 and 2013, and (b) indicated familiarity with the Doing Business Report.

⁷⁴ Our results also show that the WB and the IFC have not had as much success in raising awareness about the diagnostic and advisory content of the Doing Business Report in other policy domains that they seek to influence, such as labor or justice and security.

Table 1. Respondent Familiarity with Assessments: The 3 Most Recognized Assessments in Each Policy Domain⁷⁵

Macroeconomic Management (Average assessment familiarity = 45%)	The EU's Stabilization and Association Agreements (100%, <i>n</i> =7)	The EU's Partnership and Cooperation Agreements (92%, <i>n</i> =9)	The HIPC Initiative's "Decision Point" and "Completion Point" (97%, <i>n</i> =84)
Investment (39%)	The EU's Stabilization and Association Agreements (100%, <i>n</i> =9)	NATO's Membership Action Plan (100%, <i>n</i> =5)	The EU's Economic Partnership Agreements for ACP Countries (68%, <i>n</i> =39)
Democracy (38%)	The European Neighborhood Policy Action Plans and Country Reports (77%, <i>n</i> =47)	The EU's Stabilization and Association Agreements (76%, <i>n</i> =23)	The African Peer Review Mechanism (70%, <i>n</i> =85)
Trade (35%)	The EU's Association Agreements (90%, <i>n</i> =6)	The EU's Economic Partnership Agreements for ACP Countries (83%, <i>n</i> =42)	The World Bank and IFC's Doing Business Report (78%, <i>n</i> =95)
Finance, Credit, and Banking (34%)	The World Bank and IFC's Doing Business Report (67%, <i>n</i> =94)	The IADB's Policy-Based Loans (60%, <i>n</i> =16)	The IMF's Extended Credit Facility and Poverty Reduction and Growth Facility (60%, <i>n</i> =64)
Social Protection and Welfare (31%)	The EU's Association Agreements (80%, <i>n</i> =13)	The UN's Millennium Development Goals (78%, <i>n</i> =116)	UNDP's Human Development Index (76%, <i>n</i> =113)

⁷⁵ Unlike the other analyses conducted in this report and given the complex relationship between sample size and percentages, we used an inclusion criteria threshold here of an assessment being recognized by at least five participants, rather than being simply asked about to five participants. In this table, however, *n* indicates the number of participants from each policy domain routed to the indicated assessment.

Energy and Mining (29%)	The UN's Millennium Development Goals (68%, <i>n</i> =59)	The EU's Association Agreements (58%, <i>n</i> =7)	The EBRD's Energy Sector Assessment (56%, <i>n</i> =11)
Decentralization (29%)	The African Peer Review Mechanism (42%, <i>n</i> =20)	The World Bank's Decentralization Indicators (41%, <i>n</i> =60)	The World Bank's Development Policy Review (40%, <i>n</i> =61)
Business Regulatory Environment (29%)	The EU's Stabilization and Association Agreements (82%, <i>n</i> =16)	The World Bank and IFC's Doing Business Report (80%, <i>n</i> =130)	The EU's Association Agenda (72%, <i>n</i> =7)
Public Expenditure Management (28%)	The EU's Stabilization and Association Agreements (90%, <i>n</i> =8)	The HIPC Initiative's "Decision Point" and "Completion Point" (76%, <i>n</i> =58)	The European Neighborhood Policy Action Plans and Country Reports (68%, <i>n</i> =16)
Agriculture and Rural Development (27%)	The UN's Millennium Development Goals (70%, <i>n</i> =164)	The EU's Partnership and Cooperation Agreements (67%, <i>n</i> =7)	The World Bank's Poverty Assessment (50%, <i>n</i> =163)
Health (27%)	The UN's Millennium Development Goals (80%, <i>n</i> =272)	UNDP's Human Development Index (72%, <i>n</i> =267)	Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria (70%, <i>n</i> =272)
Education (27%)	The EU's Stabilization and Association Agreements (81%, <i>n</i> =11)	The UN's Millennium Development Goals (78%, <i>n</i> =191)	UNDP's Human Development Index (70%, <i>n</i> =188)

Tax (26%)	The World Bank and IFC's Doing Business Report (68%, <i>n</i> =19)	The African Growth and Opportunity Act (AGOA) Eligibility Criteria (45%, <i>n</i> =15)	The IMF's Policy Support Instrument (42%, <i>n</i> =33)
Infrastructure (25%)	The EU's Association Agreements (67%, <i>n</i> =16)	The EU's Partnership and Cooperation Agreements (67%, <i>n</i> =11)	The AsDB's Country Diagnostic Study (56%, <i>n</i> =49)
Labor (24%)	The European Neighborhood Policy Action Plans and Country Reports (72%, <i>n</i> =13)	The ILO's Global Monitoring and Analysis of Conditions of Work and Employment (68%, <i>n</i> =49)	The African Peer Review Mechanism (50%, <i>n</i> =16)
Civil Service (23%)	The EU's Stabilization and Association Agreements (73%, <i>n</i> =11)	The European Neighborhood Policy Action Plans and Country Reports (71%, <i>n</i> =15)	The African Peer Review Mechanism (57%, <i>n</i> =51)
Family and Gender (23%)	The UN's Millennium Development Goals (81%, <i>n</i> =48)	UNDP's Human Development Index (78%, <i>n</i> =48)	The HIPC Initiative's "Decision Point" and "Completion Point" (58%, <i>n</i> =14)
Environmental Protection (22%)	The UN's Millennium Development Goals (84%, <i>n</i> =154)	The EU's Stabilization and Association Agreements (78%, <i>n</i> =9)	The European Neighborhood Policy Action Plans and Country Reports (67%, <i>n</i> =12)
Anti-Corruption and Transparency (21%)	NATO's Membership Action Plan and Annual Progress Report (69%, <i>n</i> =11)	The African Peer Review Mechanism (68%, <i>n</i> =74)	The EU's Stabilization and Association Agreements (67%, <i>n</i> =12)

Customs (20%)	The EU's Economic Partnership Agreements for ACP Countries (74%, <i>n</i> =18)	The World Bank and IFC's Doing Business Report (63%, <i>n</i> =32)	The African Growth and Opportunity Act (AGOA) Eligibility Criteria (60%, <i>n</i> =19)
Justice and Security (17%)	The EU's Stabilization and Association Agreements (88%, <i>n</i> =16)	NATO's Membership Action Plan and Annual Progress Report (74%, <i>n</i> =17)	The African Peer Review Mechanism (55%, <i>n</i> =52)
Land (17%)	The World Bank and IFC's Doing Business Report (63%, <i>n</i> =37)	The African Peer Review Mechanism (43%, <i>n</i> =24)	The Millennium Challenge Corporation's Eligibility Criteria and Country Scorecards (37%, <i>n</i> =36)

3.4. Influencing reform efforts from the outside is substantially easier in some policy domains than in others.

There is an active debate among scholars and policymakers about whether external assessments of government performance exert more influence in some policy domains than others. Some have claimed that assessments may prove more influential in policy domains that lend themselves to quantitative forms of performance measurement (The Economist 2005; OECD and DIE 2014). Others have argued that external assessments will be less influential in policy domains that directly impinge upon the core interests of governing elites (Parks and Rice 2013; Smets 2014). The *2014 Reform Efforts Survey* provides evidence that speaks to this debate. In Table 2, we provide an overview of the three most influential external assessments by policy domain and stage of the reform process.

Table 2. The 3 Most Influential Assessments in Each Policy Domain (Level of Agenda-Setting Influence)⁷⁶

Macroeconomic management	The IMF's Article IV Consultations (4.09, <i>n</i> =17)	The HIPC Initiative's "Decision Point" and "Completion Point" (3.83, <i>n</i> =69)	The EU's Stabilization and Association Agreements (3.70, <i>n</i> =6)
Finance, credit, and banking	The EU's Partnership and Cooperation Agreements (3.42, <i>n</i> =5)	The European Neighborhood Policy Action Plans and Country Reports (3.25, <i>n</i> =5)	The IMF Financial Sector Assessment Program (FSAP) (3.17, <i>n</i> =9)
Trade	The World Bank and IFC's Doing Business Report (3.30, <i>n</i> =73)	The AsDB's Country Performance Assessment (CPA) and Performance-Based Allocation System (3.29, <i>n</i> =8)	The AfDB's Country Performance Assessment (CPA) and Performance-Based Allocation System (3.25, <i>n</i> =7)
Business regulatory environment	The EU's Association Agenda (4.50, <i>n</i> =6)	The EBRD's Country Law Assessment (3.88, <i>n</i> =7)	The World Bank and IFC's Doing Business Report (3.84, <i>n</i> =124)
Investment	The World Bank and IFC's Doing Business Report (4.43, <i>n</i> =17)	The HIPC Initiative's "Decision Point" and "Completion Point" (3.71, <i>n</i> =18)	The Multilateral Debt Relief Initiative (3.41, <i>n</i> =12)
Health	The European Neighborhood Policy Action Plans and Country Reports (3.70, <i>n</i> =6)	The UN's Millennium Development Goals (3.67, <i>n</i> =203)	The EU's Association Agreements (3.38, <i>n</i> =5)
Education	The OECD's Program for International Student Assessment (3.70, <i>n</i> =17)	The Multilateral Debt Relief Initiative (3.32, <i>n</i> =12)	The UN's Millennium Development Goals (3.29, <i>n</i> =135)
Family and gender	The UN's Millennium Development Goals (3.81, <i>n</i> =33)	The Millennium Challenge Corporation's Eligibility Criteria and Country Scorecards (3.73, <i>n</i> =7)	The US State Department's "Country Reports on Human Rights Practices" (3.46, <i>n</i> =13)

⁷⁶ As expected, the number of participants (*n*) evaluating the influence of a given assessment within a specific policy domain can be low. For this reason, we have limited our presentation and analysis to those assessments with at least 5 participants. We assume that, given the specialized knowledge of our survey participants, the narrow scope of some assessments, and our tailored survey design (see [Appendix D](#)), a higher-than-average-influence score as reported by at least 5 policy experts indicates, at least to some degree, consensus about a relatively high level of assessment influence. We urge readers, however, not to treat the listed scores as precise estimates.

Social protection and welfare	The IADB's Policy-Based Loans (3.88, <i>n</i> =5)	The IADB's Country Institutional and Policy Evaluation (CIPE) and Performance-Based Allocation System (3.60, <i>n</i> =6)	The UN's Millennium Development Goals (3.25, <i>n</i> =80)
Labor	The World Economic Forum's "Global Competitiveness Report" (3.06, <i>n</i> =14)	The ILO's Global Monitoring and Analysis of Conditions of Work and Employment (3.01, <i>n</i> =33)	DFID's Resource Allocation Model (2.80, <i>n</i> =5)
Environmental protection	The EU's Stabilization and Association Agreements (4.58, <i>n</i> =6)	The IADB's Country Environmental Analysis (3.71, <i>n</i> =9)	The EU's Association Agreements (3.33, <i>n</i> =7)
Agriculture and rural development	The IMF's Rapid Credit Facility (3.67, <i>n</i> =9)	The World Bank's Development Policy Loans Program (3.25, <i>n</i> =9)	The Multilateral Debt Relief Initiative (3.23, <i>n</i> =18)
Energy and mining	The World Bank's Country Economic Memorandum (3.57, <i>n</i> =14)	The EBRD's Energy Sector Assessment (3.4, <i>n</i> =5)	The Global Environmental Facility's Performance Index and Resource Allocation Framework (3.25, <i>n</i> =8)
Land	The World Bank and IFC's Doing Business Report (3.29, <i>n</i> =20)	The Millennium Challenge Corporation's Eligibility Criteria and Country Scorecards (2.89, <i>n</i> =12)	The World Bank's Development Policy Review (2.69, <i>n</i> =9)
Infrastructure	The AfDB's Country Performance Assessment (CPA) and Performance-Based Allocation System (3.42, <i>n</i> =12)	The Millennium Challenge Corporation's Eligibility Criteria and Country Scorecards (3.21, <i>n</i> =42)	The AfDB's Policy-Based Loans and Budget Support (3.18, <i>n</i> =11)
Decentralization	The EU's Poverty Reduction Budget Support Program (2.37, <i>n</i> =16)	The World Bank's Decentralization Indicators (2.26, <i>n</i> =26)	The World Bank's Country Policy and Institutional Assessment (CPIA) and Performance-Based Allocation System (2.22, <i>n</i> =11)
Anti-corruption and transparency	The EU's Stabilization and Association Agreements (4.50, <i>n</i> =15)	NATO's Membership Action Plan and Annual Progress Report (4.17, <i>n</i> =6)	The EU's Association Agreements (4.00, <i>n</i> =10)
Democracy	The EU's Association Agenda (4.20, <i>n</i> =9)	NATO's Membership Action Plan and Annual Progress	The EU's Stabilization and Association Agreements (3.77,

		Report (4.17, <i>n</i> =7)	<i>n</i> =19)
Civil Service	The EU's Association Agreements (3.87, <i>n</i> =8)	The AfDB's Policy-Based Loans and Budget Support (3.81, <i>n</i> =9)	The EU's Governance Initiative and Governance Incentive Tranche (3.50, <i>n</i> =6)
Justice and security	The European Neighborhood Policy Action Plans and Country Reports (4.22, <i>n</i> =5)	The EU's Association Agreements (3.33, <i>n</i> =10)	NATO's Membership Action Plan and Annual Progress Report (2.83, <i>n</i> =10)
Tax	The EU's Governance Initiative and Governance Incentive Tranche (3.50, <i>n</i> =6)	The World Bank's Worldwide Governance Indicators (3.36, <i>n</i> =8)	The World Bank's Development Policy Review (3.25, <i>n</i> =10)
Customs	The World Bank and IFC's Doing Business Report (3.68, <i>n</i> =22)	The World Bank's Diagnostic Trade Integration Studies (3.50, <i>n</i> =5)	The WTO's Accession Working Party Reports and Accession Protocols (3.18, <i>n</i> =10)
Public expenditure management	The IADB's Policy-Based Loans (4.24, <i>n</i> =9)	The EU's Partnership and Cooperation Agreements (4.22, <i>n</i> =6)	The EU's Stabilization and Association Agreements (4.10, <i>n</i> =6)

3.4.6. Agenda-Setting Influence by Policy Domain

External assessments of government performance appear to have exerted particularly strong agenda-setting influence in the health and family and gender policy domains, as reported by survey participants. The strong influence of external assessments over health sector reform efforts is consistent with the claim that the health sector is unique because it is easy for external actors to measure and monitor government performance on health related issues (OECD and DIE 2014; Honig 2014).⁷⁷ The agenda-setting influence of external assessments in the family and gender policy domain seems to reflect the focus of assessments like the UN MDGs and the MCC eligibility criteria (see Tables 1 and 2), which have not only cast a bright, public spotlight on family and gender issues but also created tangible rewards for governments to address them.⁷⁸ The USG, for example, links a country's eligibility for hundreds of millions of dollars in MCC assistance to performance on ten measures of gender equality in economic life and girls' primary education completion rates (Butts 2011; MCC 2013).⁷⁹

The policy domains that appear to be least susceptible to external influence at the agenda-setting stage are democracy and decentralization. This finding supports the notion that assessments containing (implicit or explicit) policy advice that threatens existing power structures will induce weaker policy responses (Parks and Rice 2013; Smets 2014). It is also consistent with previous research that suggests

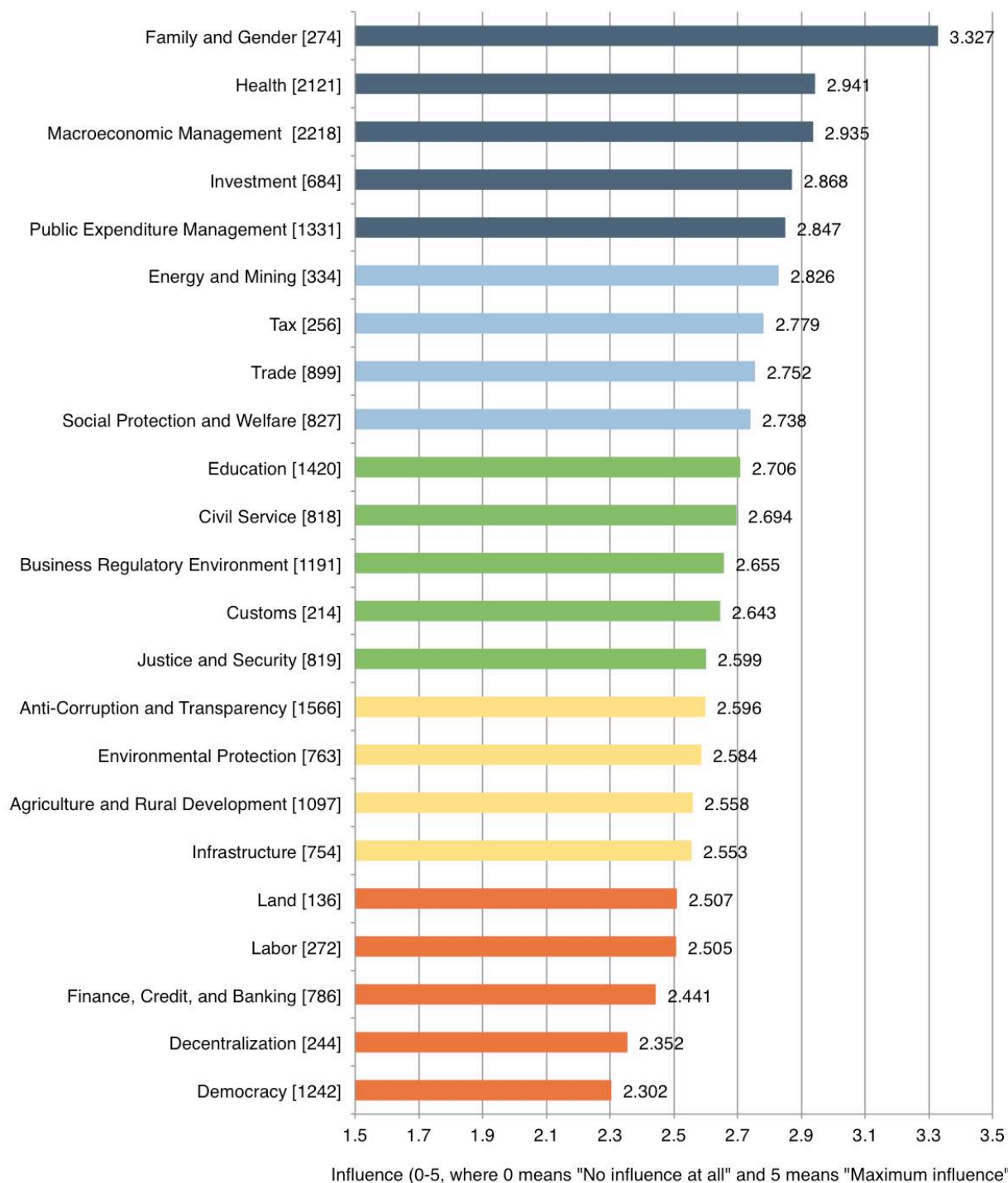
⁷⁷ Participants in a 2014 workshop co-sponsored by the OECD and the DIE coalesced around the idea that "[h]ealth is the most advanced sector with regards to results-based funding because expected results in this sector are easier to measure and verify. Donors know what to buy because outputs can be a good proxy for outcomes" (OECD and DIE 2014).

⁷⁸ See Phillips 2007; and MCC 2013a, 2013b. While the MDGs do not constitute a performance-based aid scheme, many donors tie budget support and other forms of assistance to improved performance on MDG indicators (EC 2008; Savedoff 2011).

⁷⁹ The MCC's "Gender in the Economy" index measures "the legal capacity of married and unmarried women to execute 10 economic activities: get a job, register a business, sign a contract, open a bank account, choose where to live, get passports, travel domestically and abroad, pass citizenship on to their children, and become heads of households" (Butts 2011).

decentralization reforms are likely to encounter significant domestic opposition from elected politicians and appointed bureaucrats who are motivated to preserve and consolidate their own power (Eaton et al. 2010: xv; Swift-Morgan 2014). Other policy domains that were less strongly influenced by external assessments of government performance include labor, land, and finance, credit, and banking. The difficulty of exerting outside influence in these policy domains may reflect some level of “pass through” where the insoluble nature of the domestic policy problem limits the influence that any external actor can wield (Kikeri et al. 2006). We will consider this issue at greater length later in this report.

Figure 10. The Agenda-Setting Influence of External Assessments, by Policy Domain



Note: The number of assessment-specific observations per policy domain is indicated in brackets. All else equal, there is an inverse relationship between the number of observations and the amount of error around a particular average influence score. Shading is by quintile and reflects the relative influence of external assessments in one policy domain versus another.

In Table 2, we listed the three most influential external assessments of government performance by policy domain and stage of the policymaking process. Several EU-produced assessments (e.g., Stabilization and Association Agreements, Partnership and Cooperation Agreements, Association Agenda, and European Neighborhood Policy Action Plans and Country Reports) enjoy far-reaching influence across a wide variety of policy domains.

Additionally, the UN MDGs seem to exert strong agenda-setting influence on how governments identify which reforms they will prioritize in the areas of health, education, social protection and welfare, and family and gender, while the WB's and the IFC's Doing Business Report appears to have significant agenda-setting influence in the areas of trade, customs, investment, and the business regulatory environment.

A number of assessment suppliers and sponsors appear to have benefitted from more specialized zones of policy influence. For example, the OECD's PISA has more agenda-setting influence than any other external assessment of government performance that focuses on education policy, and the World Economic Forum's Global Competitiveness Report exerts the greatest agenda-setting policy influence in the labor policy domain. The "Decision Point" and "Completion Point" assessments made under the auspices of the HIPC appear to influence the macroeconomic reform priorities of developing countries more than any other assessment.⁸⁰

A disproportionate number of the assessments identified by survey participants as particularly influential at the agenda-setting stage are tied to the potential acquisition of a tangible, material reward or set of rewards. Twenty-two, or 58%, of the thirty-eight assessments listed in Table 2 involve some "economic or financial reward for reform."⁸¹ Notwithstanding some important exceptions (e.g., the WB's Decentralization Indicators), the vast majority of external assessments that wield significant agenda-setting policy influence are performance-based aid and debt relief programs (e.g., AfDB and IADB policy lending, the EU's Governance Initiative and Governance Incentive) or other conditionality programs (e.g., the European Neighborhood Policy Action Plans and Country Reports).⁸²

3.4.7. Reform Design Influence by Policy Domain

As we indicated earlier in this report, external assessments of government performance do tend to exert greater influence during agenda setting than during reform design; however, there is important variation across policy domains. Figure 11 demonstrates that the greatest decreases in external assessment influence between the agenda-setting stage and the reform design stage are observed in the energy and mining, customs, investment, land, and education policy domains. These sharp reductions in influence may reflect the politically contested and transaction-intensive nature of energy, customs, land, and education reforms. We know from previous research that reform of the land and energy sectors, in particular, is difficult because it usually involves a substantial redistribution or restructuring of the wealth, authority, and property of vocal interest groups with high levels of pre-existing investment in illiquid assets (Shafer 1994; Kikeri et al. 2006).⁸³

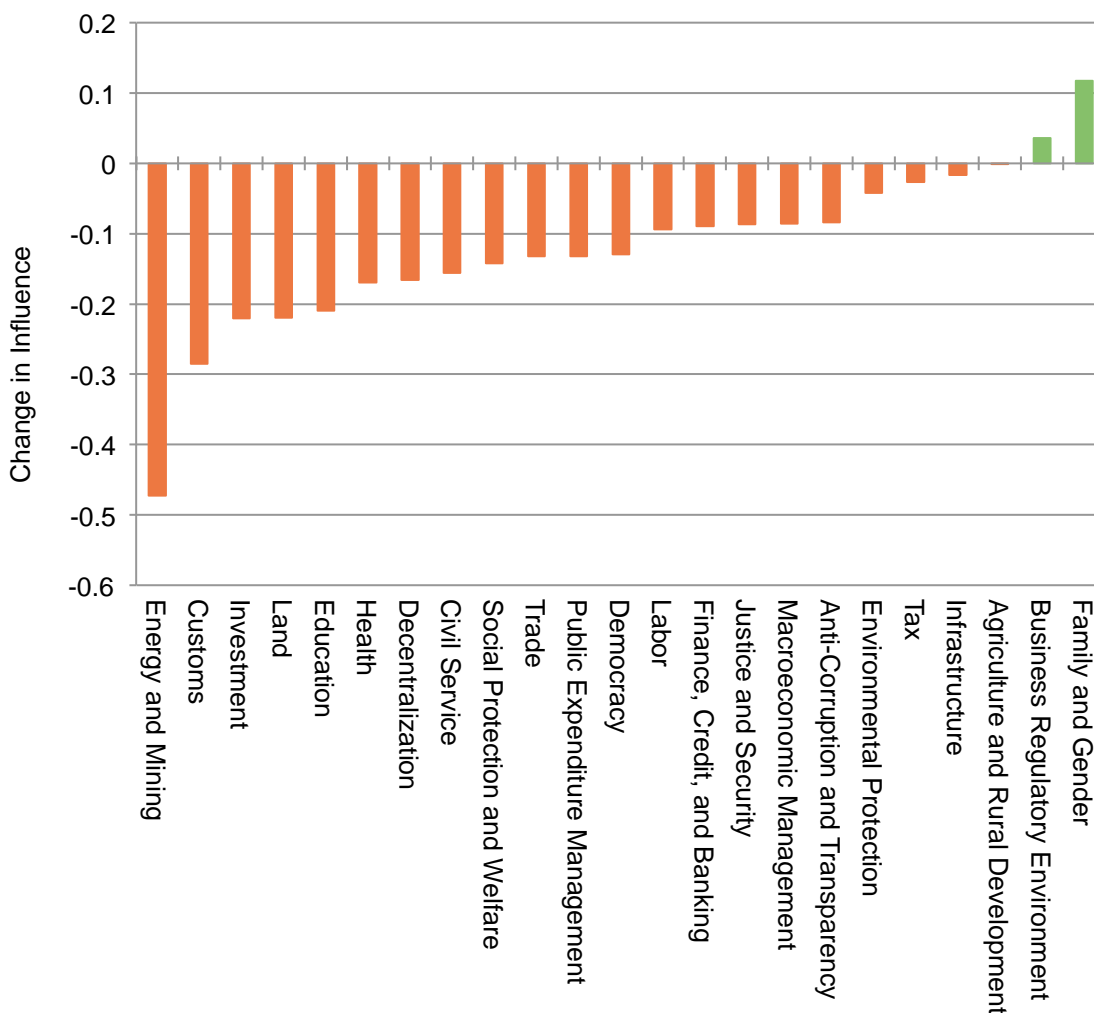
⁸⁰ These findings of narrowly targeted HIPC influence at the agenda-setting phase are consistent with previous research undertaken by Andrews (2010), Tavakoli et al. (2014), and Grek (2009).

⁸¹ By contrast, only 40.8% of the 103 assessments in our sample are attached to a material reward. See [Appendix G](#) for the methodology used to determine whether an assessment was tied to a material reward or set of rewards.

⁸² This finding at "one tail of the distribution" (i.e., the three most influential assessments by policy domain) adds important nuance to a broader empirical pattern: across the full distribution of 103 assessments, there is no statistically significant relationship between external assessment influence and the provision of extrinsic material incentives. While there is no significant trend overall, the most influential assessments are tied to material incentives.

⁸³ By contrast, the agricultural and rural development policy domain, which experiences a significantly smaller decline in assessment influence, is characterized by flexible production methods that require low sunk costs (Shafer 1994).

Figure 11. The Changing Influence of External Assessments by Reform Phase and Policy Domain



Note: The change in influence is equal to the influence of external assessments in a policy domain during reform design (on a scale of 0-5) minus their influence during agenda setting (also on a scale of 0-5). A positive value indicates an increase in influence between the agenda-setting stage and reform design stage, while a negative value indicates a decrease.

Family and gender is one of only two policy domains in which external assessments of government performance actually registered higher levels of influence at the reform design stage of the policymaking process.⁸⁴ This finding implies that while it may be relatively harder for reforms focused on family and gender issues to make it onto the policy agenda established by (mostly male) government decision-makers in developing countries (see Figure 1),⁸⁵ it is relatively easier for governments to design new policies and programs to address these issues once the need for reform is acknowledged. One reason why we may observe this pattern is that family and gender-related reforms (e.g., allowing women to own bank accounts or own property) are often not overly complex, but they usually require substantial political will (WB and IFC 2013).

⁸⁴ The second policy domain in which external assessments of government performance registered higher levels of influence at the reform design stage of the policymaking process is business regulatory environment.

⁸⁵ Also see WB 2011 and 2014.

In the infrastructure policy domain (see Table H.6 in Appendix H), we find that the AICD is one of the most influential assessments at the reform design stage. Similarly, we find that the WB's Poverty Assessment and the WB's Country Gender Assessment registered higher levels of reform design influence than agenda-setting influence in the labor, family, and gender policy domains. These assessments share a common feature: they are all *country-specific* studies or diagnostics that provide detailed policy analysis and guidance that is tailored to the particular country being assessed.

Indeed, one potential explanation for the influence of these assessments in particular is that they contain policy advice perceived to be useful in the design of specific reform features. It is less clear that these assessments exerted significant policy influence because developing country governments were seeking to reap a financial or economic reward offered by foreign governments and international organizations. In contrast to what we found at the agenda-setting stage, only 18 (or 47%) of the 38 assessments listed are linked to the potential acquisition of a material reward from an external actor.

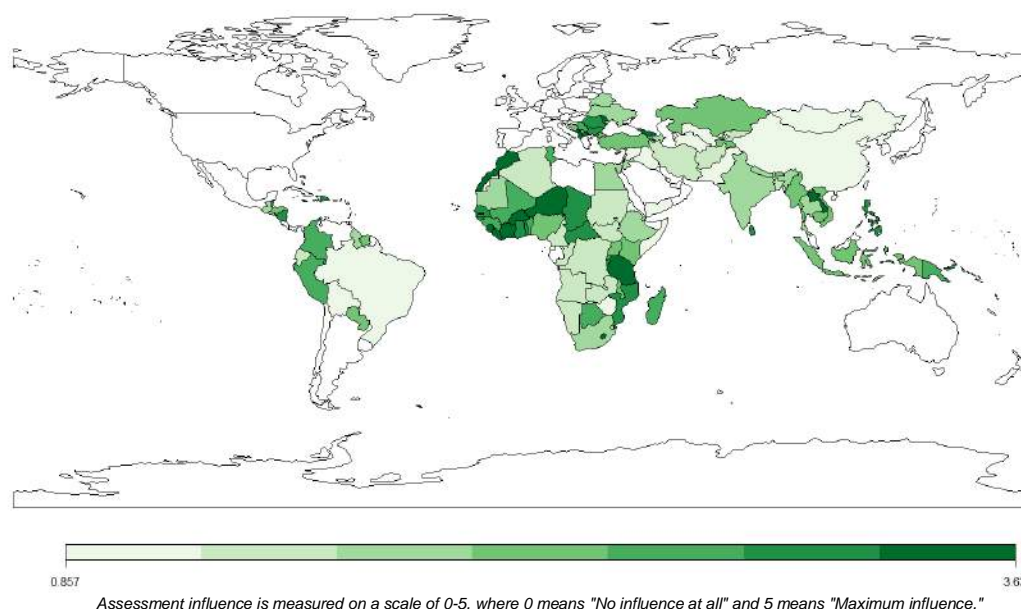
3.5. Geography matters: assessments achieve widely varying levels of uptake across regions and countries.

Our analysis of the data from the *2014 Reform Efforts Survey* indicates that external assessment influence varies significantly by country, and there is a strong tendency for assessment influence to cluster by geographic region.⁸⁶ Figure 12 provides a color-coded map of all countries included in the survey and reports the average agenda-setting influence of external assessments on policymakers in the developing world.⁸⁷ The lighter shades of green indicate lower levels of agenda-setting influence and the darker shades of green indicate higher levels of influence.

⁸⁶ Previous research suggests that there is tremendous variation in external assessment influence across countries and regions (Parks 2014; Kelley and Simmons 2014a, 2014b). Some governments are far more likely to consider external analysis and advice during the policy formulation and execution process than others. However, to date, neither the policy community nor the research community has had a strong grasp on the factors that best account for this geographical variation.

⁸⁷ Country-level scores were calculated by weighting according to the count of participants unique to each broad policy area within each country.

Figure 12. The Influence of External Assessments on the Decision to Pursue Reforms, by Country



Europe has the highest average agenda-setting influence scores, followed by Africa, Oceania, Latin America and the Caribbean, and Asia. In Europe, external assessments of government performance seem to exert particularly strong policy influence in Southern Europe (e.g., Macedonia and Albania), and slightly weaker influence in Eastern Europe. In Africa, assessment influence is especially high in West Africa, specifically, Côte d'Ivoire, Cape Verde, Sierra Leone, the Gambia, Liberia, Burkina Faso, and Niger.

While external assessment influence is still relatively high in Southern Africa, as one moves eastward, northward, and towards the center of the continent, influence begins to wane. In Asia, the data reveal higher levels of external assessment influence in Southeast Asia and lower levels of influence in Central Asia. In Oceania, assessment influence is highest in Polynesia and lowest in Micronesia.⁸⁸ Finally, assessment influence in Latin America is highest in Central America and lowest in South America.

Figure 13 similarly visualizes the average influence of external assessments on the reform design efforts of policymakers in the developing world. The lighter shades of orange indicate lower levels of assessment influence and the darker shades of orange indicate higher levels of influence. Overall, we see few differences between the countries that are influenced by external assessments at the agenda-setting stage of the policymaking process and countries that are influenced by external assessments at the reform design stage of the policymaking process.⁸⁹

This strongly positive and statistically significant correlation suggests that the same country-level attributes that make a government more inclined to recalibrate its reform priorities in response to an external assessment also influence the extent to which it uses assessments to select the specific features

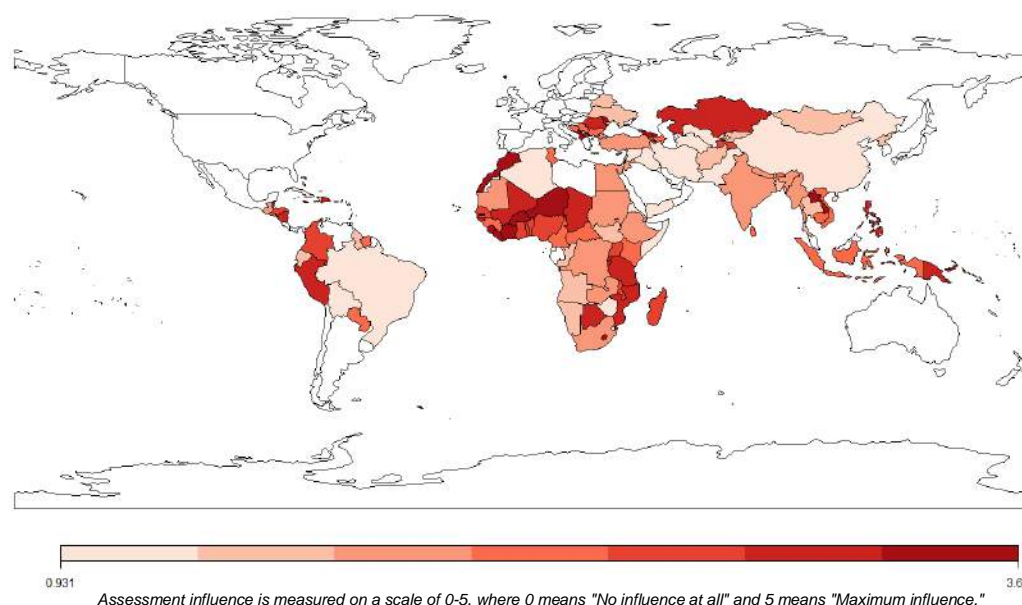
⁸⁸ We refer to Polynesia (Tuvalu, Western Samoa, and Tonga) and Micronesia (Kiribati, the Federated States of Micronesia, and the Marshall Islands) as sub-regions within Oceania, as defined by the UN Statistics Division.

⁸⁹ The bivariate correlation between these different country-level indicators of external assessment influence is 0.935. A Spearman's correlation between the two variables yields a similar value of 0.930.

of its reform interventions.⁹⁰ If external assessments influence a government's decision to pursue reform in the first place, they also seem to wield significant influence in shaping downstream reform design decisions. This suggests that influence at an early stage of the policymaking process can translate into more influence at subsequent stage.

The data reveal some other noteworthy differences in external assessment influence across different stages of the policymaking process. In many small Pacific island states (e.g., Papua New Guinea, Micronesia, Tuvalu, the Marshall Islands, and Samoa), external assessments of government performance seem to have exerted more influence on specific reform design decisions than on the reform priorities placed on the policy agenda.⁹¹ In these cases, assessments may have served more as sources of technical assistance than as agenda-setting inputs.⁹²

Figure 13. The Influence of External Assessments on the Design of Reforms, by Country



In addition to these general trends, we observe significant variation in the influence of *specific* assessments across regions. Figure 14 indicates that, in East Asia and the Pacific, the single most influential external assessment of government performance at the agenda-setting stage was TI's Corruption Perceptions Index. This finding is broadly consistent with the results reported in Parks and Rice (2013), as well as a recent study about corruption indicators in East and Southeast Asia, which concludes that TI's Corruption Perceptions Index has "placed the issues of combating corruption and improving governance squarely on the international policy agenda of governments and private enterprises" (Lin and Yu 2004: 143).⁹³

⁹⁰ This is consistent with the findings reported in Parks (2014). It should also be noted that this correlation most likely also reflects a lack of differentiation among survey participants between the stages of the policy process.

⁹¹ Similar increases in assessment influence were found in Central Asia (Mongolia, Kazakhstan, and Tajikistan) and the Caribbean (Jamaica and the Dominican Republic). Other, more isolated, increases in assessment influence were found in several countries, including Cameroon, Equatorial Guinea, Azerbaijan, Bolivia, and Cape Verde, among others.

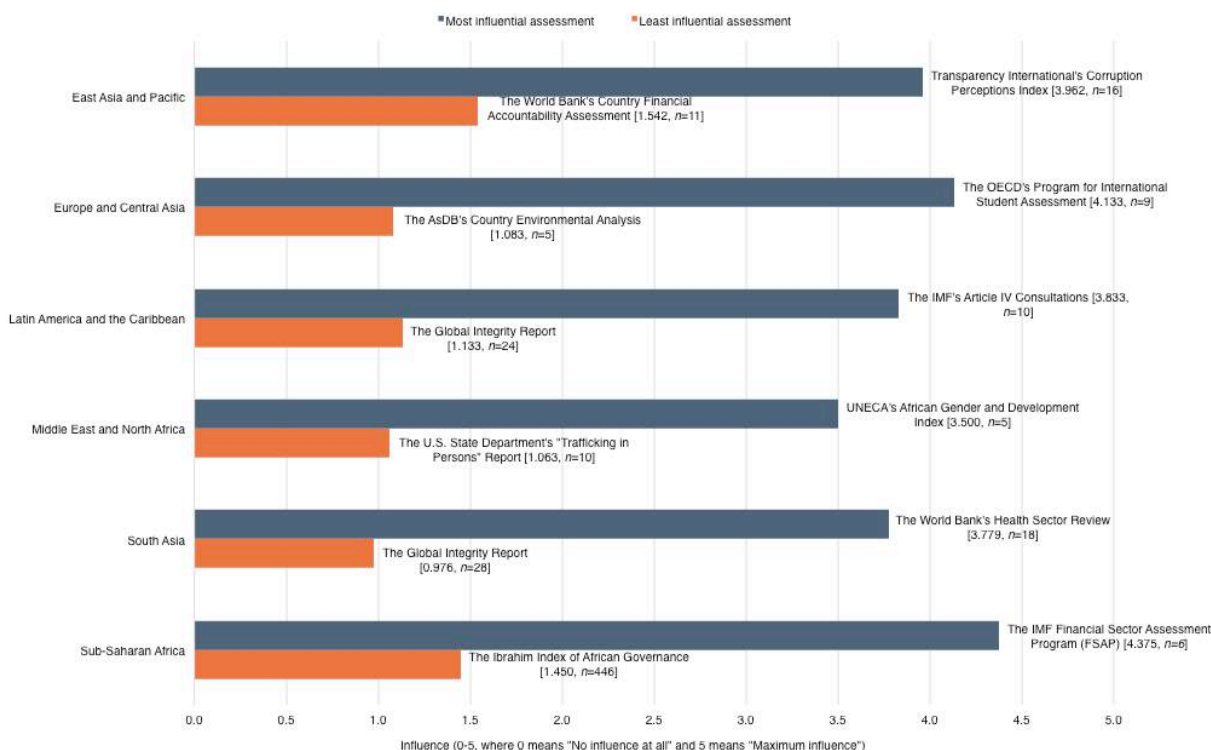
⁹² On the issue of whether and when technical assistance is used to implement policy recommendations contained in external assessments, see de Renzio and Masud 2011; IEG 2008; Geertson 2010; and Parks 2014.

⁹³ TI claims that its Corruption Perceptions Index "has been widely credited with putting the issue of corruption on the international policy agenda" (TI 2015). See Andersson and Heywood 2009.

In Europe and Central Asia, the OECD's PISA was the most influential assessment at the agenda-setting phase, which supports Sotiria Grek's claim that, in Europe, PISA is assigned "particularly significant weight as an indicator of the success or failure of education policy" (Grek 2009: 4).

Our findings for Latin America and the Caribbean are more surprising. The IMF's Article IV Consultations is the single most influential external assessment of government performance at both the agenda-setting and reform design stages, which does not comport with previous research suggesting that the IMF's policy analysis and advice is less influential in this region of the world (Chwieroth 2008; Helleiner and Momani 2008; and IMF 2013).⁹⁴ One possible explanation for the apparent disconnect between the conventional wisdom and our survey results (reported in Figures 14 and 15) is that government leaders in the Western Hemisphere value and quietly rely upon IMF policy analysis and advice while being careful not to publicly acknowledge the IMF's influence for domestic political reasons (Remmer 1986; Vreeland 2003).

Figure 14. The Most and Least Influential External Assessments at the Agenda-Setting Stage, by Region



Note: Regions follow World Bank classifications. Brackets indicate first the agenda-setting influence score and then the number (n) of observations for that assessment within the indicated region.

In the MENA region, our results suggest that the State Department's Trafficking in Persons Report is the single least influential external assessment of government performance at both the agenda-setting and reform design stages of the policymaking process. Previous research has shown the TIP Report to be generally effective as a reform catalyst (Kelley and Simmons 2014a). However, Figures 14 and 15 demonstrate that the policy influence of this assessment varies significantly across regions: while the TIP

⁹⁴ For example, in a 2011-2012 survey of central bank and finance ministry officials, the IMF concluded that "strong negative sentiment against the Fund ... would make it almost impossible for policymakers to follow Fund advice, because being seen to do so would be the political 'kiss of death.' ... Interviews confirmed the survey results, with the experiences from the Asian and Latin American crises still firmly entrenched in the thinking of senior policymakers today" (IMF 2013: 27).

Report ranks 60th out of 101 assessments on our measure of agenda-setting influence (see Figure 5), it ranks last in the MENA region. Our findings are consistent with those of Cho et al. (2014), who report that government commitment to anti-trafficking measures is weak and declining in the Middle East. The fact that governments in the MENA region have not mounted consistent efforts to confront this problem suggests that an external assessment's influence will be limited when it addresses an issue that lacks local resonance with the domestic authorities.

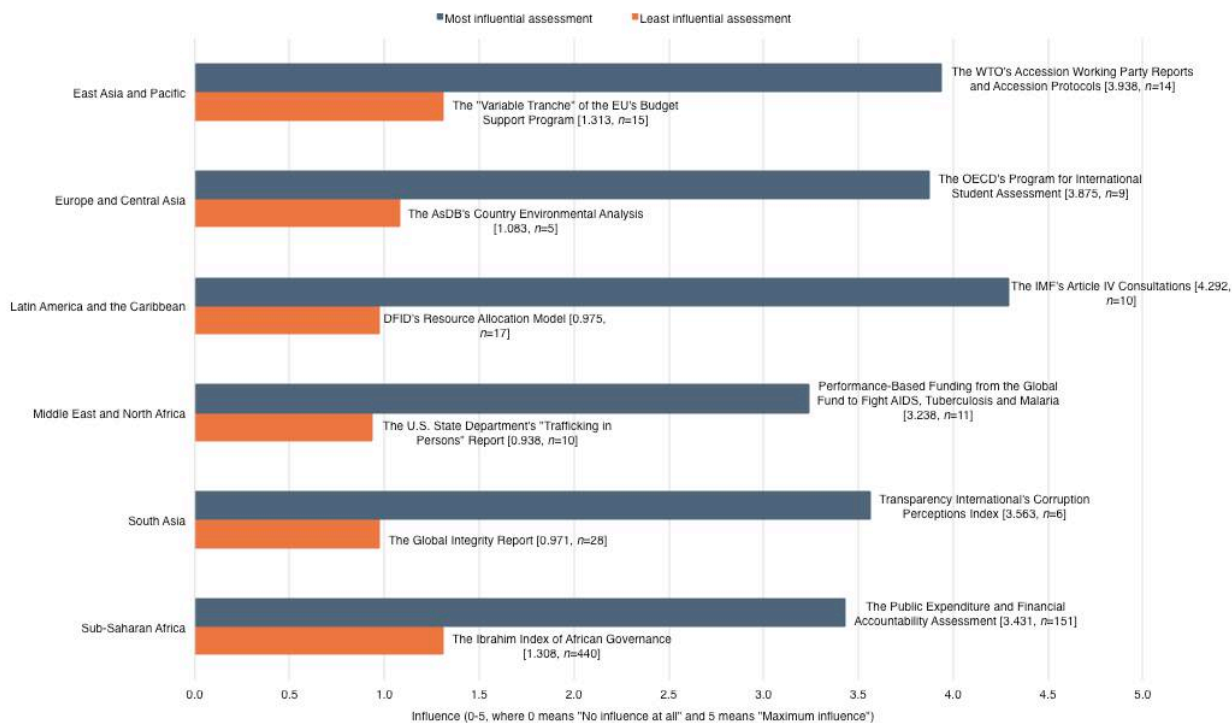
In Sub-Saharan Africa, South Asia, and Latin America and the Caribbean, two external assessments of governance—the Global Integrity Report and the Ibrahim Index of African Governance—registered exceptionally low levels of policy influence. On one hand, these assessments' lack of influence is somewhat unexpected and disappointing since they both exemplify the recent call from development policymakers, practitioners, and scholars for more “actionable governance indicators” that can be used to guide policy (WB 2006a; Knack 2007; UNDP and Global Integrity 2008; and Trapnell 2011). On the other hand, these assessments call attention to a set of vexing challenges that, in many cases, can only be resolved by taking measures that directly challenge vested interests or disrupt existing power structures. Previous research suggests that external pressures for governance reform may induce a weaker policy response for precisely this reason (Parks and Rice 2013; Smets 2014).⁹⁵

Moving from agenda-setting to reform design (see Figure 15), an overarching trend emerges: assessments with high levels of regional influence at the agenda-setting stage remain influential, while assessments with low levels of agenda-setting influence stay relatively un-influential. This adds further support to our consistent, repeated finding that perceptions of agenda-setting influence are strongly related to perceptions of reform design influence, and vice versa.⁹⁶

⁹⁵ It is also bears mentioning that among the various indicators that comprise the Ibrahim Index of African Governance are a significant number of “Africa Integrity indicators” derived from the organization—Global Integrity—that produces the Global Integrity Report.

⁹⁶ We observe this pattern not only in global influence scores, but also at the regional level. Indeed, the region with the strongest assessment-level correlation (0.927) between assessment influence at agenda-setting and reform design is Latin America and the Caribbean, and this correlation remains similarly strong across all regions, at most falling to 0.847 in South Asia.

Figure 15. The Most and Least Influential External Assessments at the Reform Design Stage, by Region



Note: Regions follow World Bank classifications. Brackets indicate first the reform design influence score and then the number (n) of observations for that assessment within the indicated region.

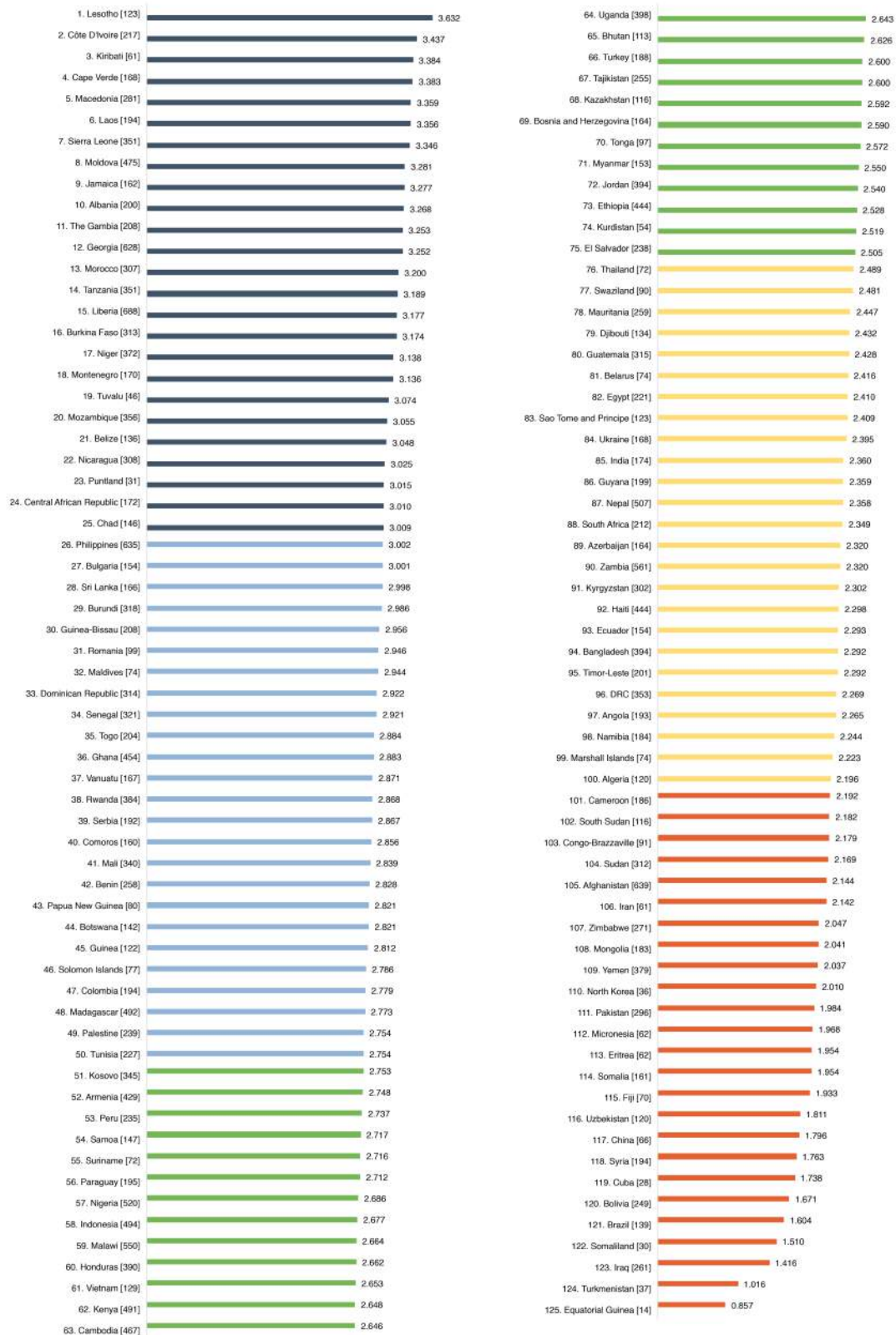
3.5.1. The Countries Most and Least Influenced by External Assessments

Within and across regions, there is significant country-level variation as to the extent to which the authorities drew on external sources of analysis and advice. As seen in Figure 16, “small states” appear to have been most influenced by external assessments of government performance.⁹⁷ The ten countries most influenced by external assessments have a median population combined of 2.8 million, compared to 9.4 million median population of countries included in our sample.⁹⁸ The correlation between population size and external assessment influence is also observable across the countries in the sample beyond the top ten.

⁹⁷ It should be noted that a threshold of $N \geq 5$ observations per country per survey question was used in this analysis. Thus, Zanzibar ($N=0$) was excluded. For more on how the country-level data were weighted to better enable apples-to-apples comparisons, see [Appendix C](#).

⁹⁸ These estimates were drawn from the WB's World Development Indicators and calculated based on averages of each country's population size over the 10 year period covered in our survey (2004-2013).

Figure 16. The Agenda-Setting Influence of External Assessments, by Country



Note: Participants in each country were asked to evaluate the agenda-setting influence of those assessments with which they were familiar, on a scale of 0 to 5, where 0 means "No influence at all" and 5 means "Maximum influence". The estimates presented in this chart are presented in rank order by quintile, and the location of a country within a specific quintile indicates the relative score of assessment influence in that country vis-a-vis other countries. The number of assessment-specific observations pertaining to each country is in brackets. Zanzibar was excluded from analysis because it did not meet a response sample size threshold of at least 10 observations. Due to respondent security concerns, the North Korean assessment influence score only reflects the experiences and perceptions of those development partner staff and officials and independent country experts who were physically located outside of North Korea on the survey launch date.

Unsurprisingly, the countries that were least susceptible to external assessment influence were closed economies, autocratic regimes, or countries that experienced long periods of political instability and war. Turkmenistan, the second least influenced country⁹⁹ on our measure of assessment influence, is governed by elites who have "zero tolerance ... for information that even implies a less-than-stellar performance by the government" (IREX 2014: 283).¹⁰⁰ Also on our list of "least influenced countries" are Equatorial Guinea, Iraq, Uzbekistan, China, Syria, Cuba, and Somaliland¹⁰¹, all of which Freedom House designates as "not free" countries.¹⁰² Two other notable countries are Brazil and Bolivia. The limited influence of external assessments in these two countries may reflect Brazil's longstanding sensitivity to perceived intrusions upon its national sovereignty (Santiso 2003; Laurance 2007; de Almeida 2013), and the fact that the Evo Morales administration (2006 to present) in Bolivia has become deeply suspicious of foreign powers seeking to influence its policy decisions and actions (Rieff 2005).

Moving from agenda-setting to reform design (see Figure 17), the most and least influenced countries remain quite similar. Lesotho (with a population of approximately 2 million) is still the single most responsive nation to external assessments at the reform design stage, followed by Cape Verde, Jamaica, Puntland, and Côte d'Ivoire.

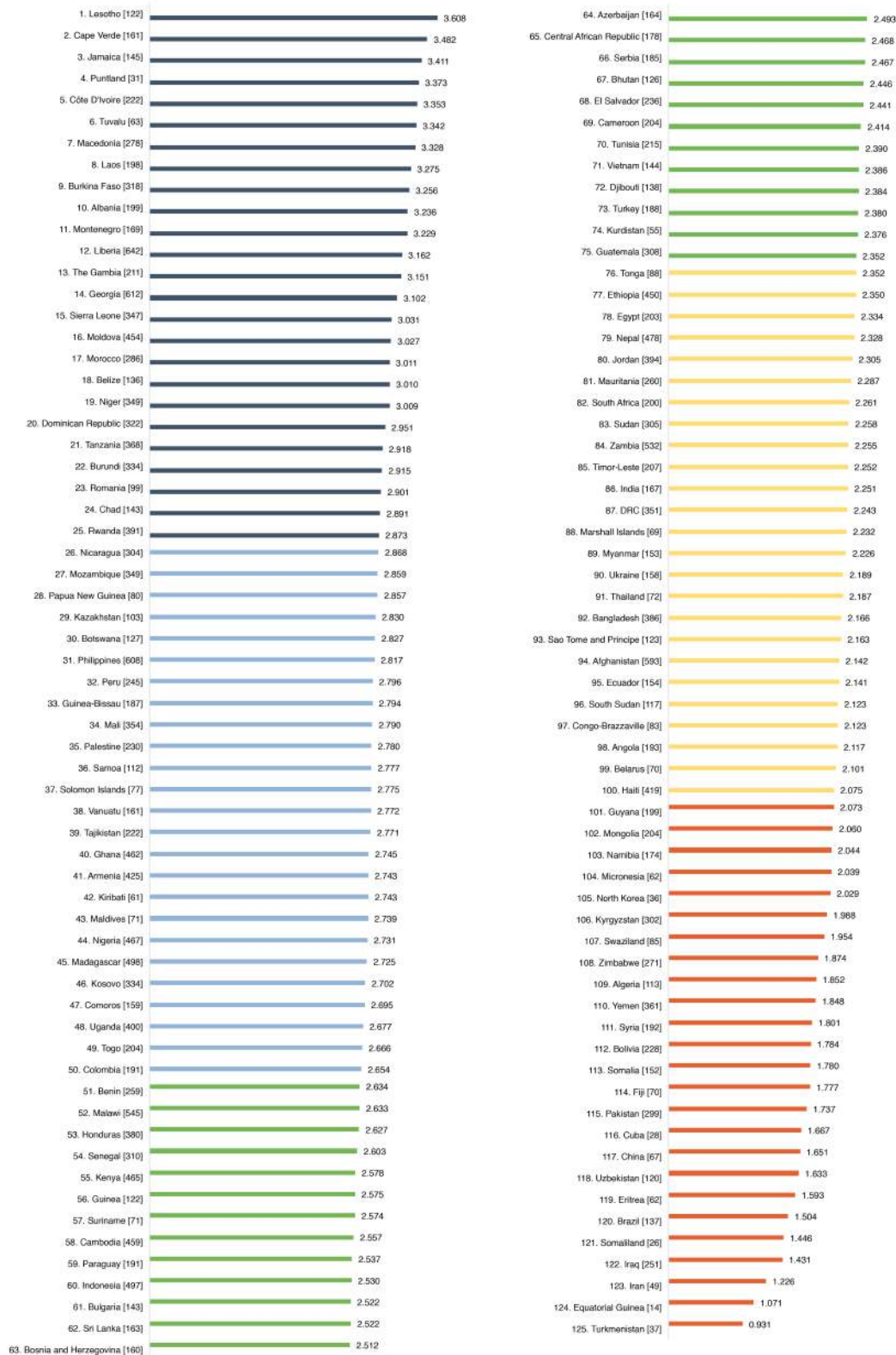
⁹⁹ Participants from Turkmenistan indicated the second lowest level of external assessment influence: 1.016 on a scale of 0 to 5, with 0 indicating no influence at all and 5 indicating maximum influence.

¹⁰⁰ IREX (2014: 283) also notes that, "Turkmenistan falls solidly among the worst of the worst in terms of the media's ability to sustainably inform the public with useful, relevant news and information."

¹⁰¹ Given the absence of Freedom House data for Somaliland, we imputed its score by relying Somalia's score.

¹⁰² These designations come from Freedom House's *Freedom in the World* Report.

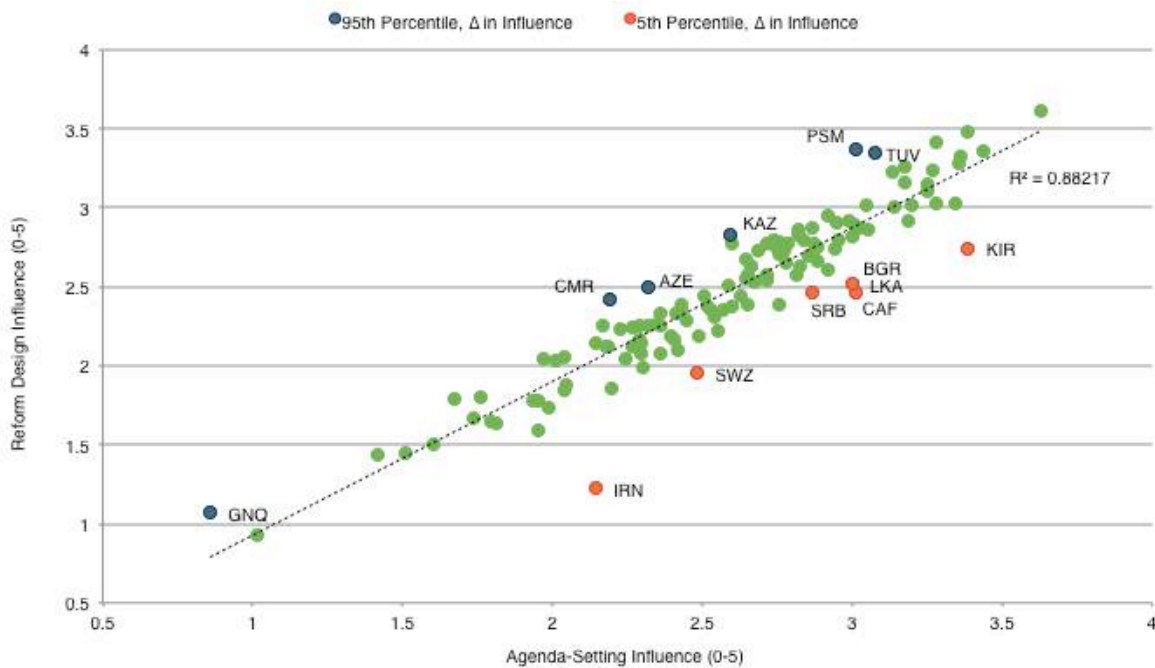
Figure 17. The Reform Design Influence of External Assessments, by Country



Note: Participants in each country were asked to evaluate the reform design influence of those assessments with which they were familiar, on a scale of 0 to 5, where 0 means "No influence at all" and 5 means "Maximum influence". The estimates presented in this chart are presented in rank order by quintile, and the location of a country within a specific quintile indicates the relative score of assessment influence in that country vis-a-vis other countries. The number of assessment-specific observations pertaining to each country is in brackets. Zanzibar was excluded from analysis because it did not meet a response sample size threshold of at least 10 observations. Due to respondent security concerns, the North Korean assessment influence score only reflects the experiences and perceptions of those development partner staff and officials and independent country experts who were physically located outside of North Korea on the survey launch date.

Figure 18 highlights the three countries with the most severe declines in assessment influence between the agenda-setting and reform design stages of the policymaking process: Iran (-0.916), Kiribati (-0.642), and the Central African Republic (-0.542). These results may point to the fact that very few assessment suppliers have an in-country presence in Tehran, South Tarawa, and Bangui. This lack of "ground game" most likely limits their ability to work with the authorities on the design features of reforms (Parks and Rice 2014).

Figure 18. The Strong Country-Level Correlation between Assessment Influence in Agenda Setting and Reform Design



Note: For both agenda setting and reform design, 0 means "No influence at all" and 5 means "Maximum influence". The R value of 0.935 indicates a very strong, positive relationship between the agenda-setting influence and reform design influence of assessments in an average sample country. Change (Δ) in influence is measured as assessment influence in reform design minus assessment influence in agenda setting. Countries in the 95th percentile are among the 5% of countries with the greatest, or most positive, changes in influence, while countries in the 5th percentile are among the 5% with the smallest, or most negative, changes in influence.

Box 6. Regional Differences in Assessment Familiarity

Examining assessment familiarity across six world regions yields results that are broadly similar to the global findings, but with a few notable differences. The UN, World Bank, EU, and Global Fund to Fight AIDS, Tuberculosis and Malaria have the most widely recognized assessments across regions. However, there is also some important variation between regions.

In Central and Eastern Europe and Central Asia, the two most widely recognized assessments are the EU's Stabilization and Association Agreements and the European Neighborhood Policy Action Plans and Country Reports. The EU not only has a large presence in this part of the world, but also the prospect of tariff-free access to European markets or eventual EU membership likely renders its assessments of government performance more salient and influential in the eyes of local stakeholders in those countries (Kelley 2006; Gavrilis 2014).

In Latin America, where education quality is an increasingly important issue, the OECD's Program for International Student Assessment appears to be widely known among policymakers and practitioners (Grindle 2004). By comparison, in South Asia, where problems related to child labor and worker rights are particularly acute (Cho et al. 2014), there are high levels of familiarity with the ILO's Global Monitoring and Analysis of Conditions of Work and Employment. Finally, in Sub-Saharan Africa, the region most severely affected by HIV/AIDS and malaria (Bor 2007), participants are particularly familiar with Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria.

These patterns in the data support the notion that assessments and their suppliers have higher levels of visibility with government decision-makers in the places that they seek to influence.

4. External Assessment Influence: Assessment-Level Findings

Using responses to the *2014 Reform Efforts Survey*, we also examined the extent to which assessment-level characteristics contribute to their uptake by and influence with senior government decision-makers in developing countries. We specifically consider the following assessment-level characteristics: policy focus, problem-solving orientation, geographic coverage, source of performance data, the general policy area, nature of the assessment supplier, and longevity.

Analyzing the survey data, we identified **12 assessment-level findings** on external assessment influence, which we will preview here and discuss at length in this section:

1. Paying attention to “nuts and bolts of government” may result in greater assessment influence.
2. Getting the incentives right is easier for some policy problems than others.
3. There are limits to what external assessment suppliers can reasonably expect to influence.
4. Country-specific diagnostics generally exert greater influence than cross-country benchmarking exercises.
5. Assessments that rely on host government data are more influential.
6. Governance assessments face the greatest obstacles to in-country uptake and influence.
7. Intergovernmental organizations with a global reach produce more influential assessments.
8. The longer an assessment’s track record of publication, the more influential it becomes vis-à-vis others.
9. Neither incentives nor penalties seem to easily explain assessment influence.
10. Assessment influence might depend upon the flexibility and specificity of policy advice.
11. Alignment with local priorities is a key determinant of external assessment influence.
12. Senior government leaders and their deputies engage with external assessments in very different ways.

4.1. Do the types of policy problems that external assessments seek to address affect their influence?

Do external assessments more successfully encourage “target” governments to focus their reform efforts on some types of policy problems more than on others? Using data from the *2014 Reform Efforts Survey*, we systematically categorized and coded the open-ended responses of survey participants to identify 16 policy *problem types* typically addressed by developing country reform efforts, including:¹⁰³

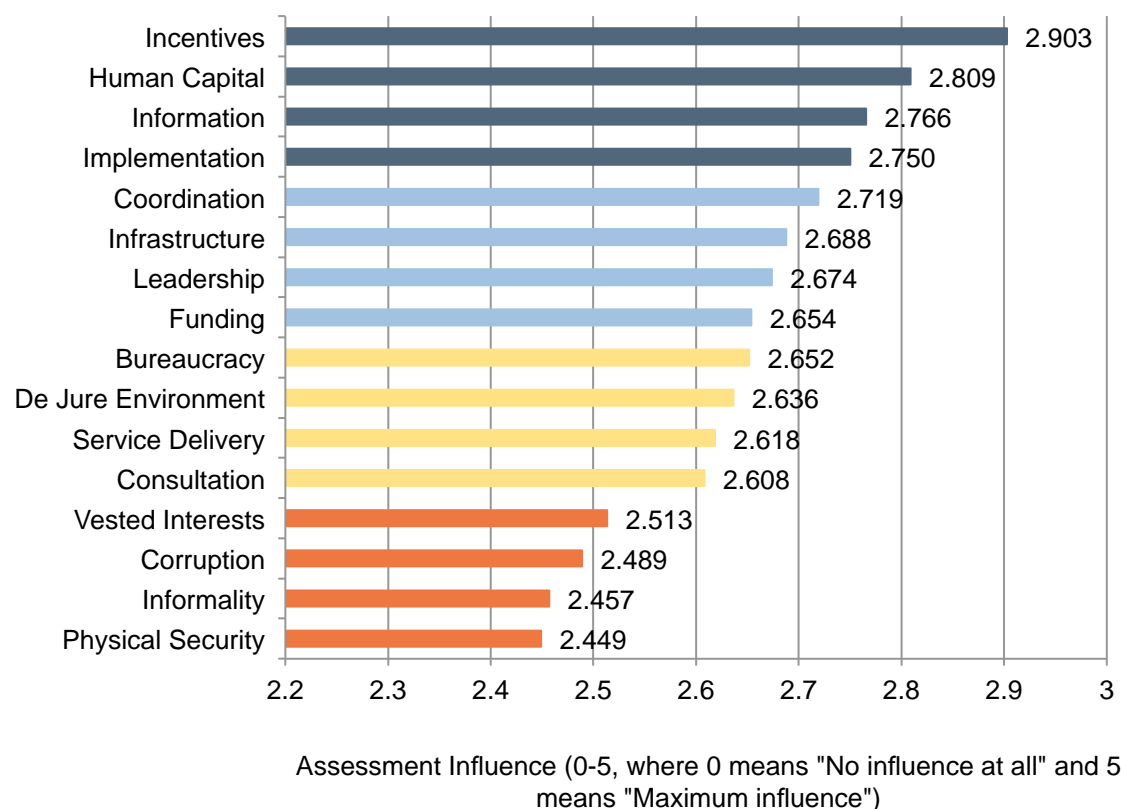
- **Bureaucracy:** Overly burdensome or inefficient bureaucratic procedures, practices, or regulations
- **Consultation:** Inadequate involvement of local stakeholders in decision-making
- **Coordination:** Poor inter- or intra-institutional communication or coordination
- **Corruption:** Fraud, corruption, opacity, cronyism, or political interference
- **De Jure Environment:** Inadequate, inconsistent, or unstable regulations, laws or policies
- **Funding:** Insufficient funding
- **Human Capital:** Insufficient human capital
- **Implementation:** Inadequate/inconsistent implementation or enforcement of regulations, laws or policies

¹⁰³ We thank Matt Andrews for providing input into the “problem types” and NORC at the University of Chicago for helping to code open-ended survey responses into these sixteen “problem type” categories. Problem types were identified using data from question 20. We utilized data from question 31 to calculate the weighted agenda-setting influence of external assessments on reforms focused on each problem type.

- **Incentives:** Behavior or incentives of private individuals or firms needs to change
- **Informality:** Disconnect between formal policies or institutions and informal administrative, cultural, or economic practices or norms
- **Information:** Inadequate data collection or information systems
- **Infrastructure:** Insufficient equipment, technology, or infrastructure
- **Leadership:** Inadequate management, strategic direction, focus, or political will
- **Physical Security:** Insufficient safety and security due to conflict, terrorism, or instability
- **Service Delivery:** Poor delivery of public services
- **Vested Interests:** Overly centralized decision-making dominated by the interests of a single group

In Figure 19, we depict the relative agenda-setting influence of external assessments by the problem focus of their specific reform efforts.

Figure 19. Substantive Focus of Reforms Influenced by External Assessments



Note: *Shading is by quartile. The position of a given problem type within a given quartile reflects the relative influence of external assessments on the decision of an average government (in our sample) to undertake at least one reform effort focused on solving that particular type of problem. In question 31 of the survey questionnaire, participants were asked to evaluate -- on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence" -- the amount of influence that specific external assessments had on a government's decision to pursue reforms focused on three specific problems. These three problems were identified by the survey participants themselves in a prior, open-ended question (question 20). After the survey field period, we coded each problem by problem type. This graph shows the weighted average influence score of all external assessments of government performance on reforms that tried to solve each type of problem.*

Additionally, the external assessments of government performance that exert the greatest agenda-setting influence on each problem type (see Table 3) shows that there is significant variation in and specialization

among the external assessments that help put specific types of problems on the policy agendas of developing country governments.¹⁰⁴

Table 3. The 3 Most Influential Assessments in a Government's Decision to Pursue Reforms, by Specific Problem Type

Incentives (Influence of external assessments = 2.903)	The EU's Poverty Reduction Budget Support Program (3.5, <i>n</i> =14)	The "Variable Tranche" of the EU's Budget Support Program (3.4, <i>n</i> =5)	UNDP's Human Development Index (3.4, <i>n</i> =18)
Human Capital (2.809)	The EU's Association Agenda (4.4, <i>n</i> =6)	The EU's Stabilization and Association Agreements (3.8, <i>n</i> =14)	The EU's Association Agreements (3.7, <i>n</i> =17)
Information (2.766)	The GAVI Alliance's Health Systems Strengthening Window (3.8, <i>n</i> =7)	The Public Expenditure and Financial Accountability Assessment (PEFA) (3.75, <i>n</i> =22)	The World Bank's Diagnostic Trade Integration Studies (3.7, <i>n</i> =5)
Implementation (2.750)	The EU's Stabilization and Association Agreements, (4.1, <i>n</i> =12)	NATO's Membership Action Plan and Annual Progress Report (4.0, <i>n</i> =8)	The World Bank's Country Financial Accountability Assessment (4.0, <i>n</i> =5)
Coordination (2.719)	The EU's Stabilization and Association Agreements (4.0, <i>n</i> =11)	The World Bank's Education Sector Review (3.6, <i>n</i> =5)	The European Neighborhood Policy Action Plans and Country Reports (3.6, <i>n</i> =5)
Infrastructure (2.688)	The World Bank's Health Sector Review (3.8, <i>n</i> =5)	Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria (3.8, <i>n</i> =9)	The WTO's Accession Working Party Reports and Accession Protocols (3.4, <i>n</i> =7)
Leadership (2.674)	The GAVI Alliance's Immunization Services Support (ISS) Window (3.8, <i>n</i> =9)	Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria (3.5, <i>n</i> =17)	The Public Expenditure and Financial Accountability Assessment (PEFA) (3.5, <i>n</i> =72)
Funding (2.654)	The EU's Stabilization and Association Agreements (4.1, <i>n</i> =18)	The World Bank's Education Sector Review (3.8, <i>n</i> =8)	The IADB's Country Institutional and Policy Evaluation (CIPE) and Performance-Based Allocation System (3.7, <i>n</i> =7)

¹⁰⁴ After restricting our participant sets to only those who mentioned a particular problem type, we were able to replicate the assessment-specific weighting system used to generate the aggregate statistics presented throughout this report—and thereby identify the three assessments that exerted the most agenda-setting influence over reforms focused on each problem type.

Bureaucracy (2.652)	Transparency International's Corruption Perceptions Index (4.5, <i>n</i> =6)	The HIPC Initiative's "Decision Point" and "Completion Point" (3.7, <i>n</i> =9)	The World Bank and IFC's Doing Business Report (3.7, <i>n</i> =44)
De Jure Environment (2.636)	The EU's Association Agenda (4.0, <i>n</i> =5)	The World Bank's Education Sector (3.9, <i>n</i> =7)	The World Bank's Health Sector Review (3.8, <i>n</i> =11)
Service Delivery (2.618)	The EU's Association Agenda (4.5, <i>n</i> =8)	The World Bank's Health Sector Review (3.7, <i>n</i> =39)	Transparency International's Corruption Perceptions Index (3.4, <i>n</i> =11)
Consultation (2.608)	The World Bank's Development Policy Review (3.4, <i>n</i> =13)	The EU's Stabilization and Association Agreements (3.3, <i>n</i> =5)	The World Bank's Country Economic Memorandum (3.2, <i>n</i> =11)
Vested Interests (2.513)	The World Bank's Education Sector Review (4.0, <i>n</i> =6)	The EU's Stabilization and Association Agreements (4.0, <i>n</i> =7)	The World Bank's Growth Diagnostic Studies (3.3, <i>n</i> =14)
Corruption (2.489)	The Global Environment Facility's Performance Index and Resource Allocation Framework (4.2, <i>n</i> =5)	The AsDB's Country Poverty Analysis (3.70, <i>n</i> =5)	The EU's Stabilization and Association Agreements (3.6, <i>n</i> =42)
Informality (2.457)	The World Bank and IFC's Doing Business Report (3.1, <i>n</i> =7)	The UN's Millennium Development Goals (3.1, <i>n</i> =13)	The Millennium Challenge Corporation's Eligibility Criteria and Country Scorecards (2.6, <i>n</i> =7)
Physical Security (2.449)	The IADB's Policy-Based Loans (4.3, <i>n</i> =6)	The IADB's Debt Relief Initiative (3.6, <i>n</i> =7)	The World Bank Poverty Assessment (3.3, <i>n</i> =8)

Putting these data points together, we identified three important findings:

1. Paying attention to “nuts and bolts of government” may result in greater assessment influence.
2. Getting the incentives right is easier for some policy problems than others.
3. There are limits to what external assessment suppliers can reasonably expect to influence.

4.1.1. Paying attention to the “nuts and bolts of government” may result in greater assessment influence.

External assessments of government performance seem to have relatively more agenda-setting influence when they seek to place issues on the policy agenda that relate to the “nuts and bolts of government”

(e.g., providing sufficient personnel or training so that they can discharge basic functions, the creation of modern data collection and information management systems, and the provision of modern equipment and technology infrastructure). Notwithstanding the fact that assessment influence scores fall within a narrow band, this pattern is clearly evident in the data (see Figure 19).

To the extent that the results reported in Figure 19 capture the revealed preferences of governments in the developing world, they are consistent with a growing body of literature on “task complexity,” which suggests that donors, international organizations, and other external actors often ask public sector institutions in developing countries to “run before they walk” (Pritchett et al. 2013; Krasner and Risse 2014; and Hoey 2015).¹⁰⁵ These findings are also consistent with new research that suggests the limited organizational capacity and technical expertise of developing country governments can render some of the most “demanding” content in external assessments of government performance less influential (Hille and Knill. 2006; Parks 2014; IMF 2014). A less sanguine interpretation of our results in Figure 19 is that the authorities in developing countries would prefer to get lots of “goodies” (funding, staff, and equipment) rather than making difficult policy and institutional changes that will improve the de facto quality of governance.

4.1.2. Getting the incentives right is easier for some policy problems than others.

External assessments appear more likely to influence a “target” government’s policy agenda when their diagnostic or advisory content focuses on the behaviors and incentives of private firms and individuals.¹⁰⁶ Information provision and other interventions geared towards influencing the behavior of individuals and groups are not necessarily too costly, complex, or time-consuming for developing country governments to pursue.¹⁰⁷

Conversely, external assessments of government performance appear to have relatively less influence when they seek to address deeply-rooted, complex problems, and substantial domestic political economy challenges (e.g., misalignment between de jure policies and de facto practices, instability and low levels of adherence to the rule of law, or public service delivery). Similarly, reform efforts that seek to alter existing political power structures by decentralizing decision-making processes, combatting corruption, or increasing the involvement of local stakeholders also have lower levels of influence. These results are consistent with research on the difficulty of providing external support to facilitate reforms that effectively confront issues of power, rent-seeking, corruption, and institutional dysfunction.¹⁰⁸

¹⁰⁵ Khwaja (2008) and Schäferhoff (2014) demonstrate that task—or project—complexity is negatively correlated with implementation success.

¹⁰⁶ Consider the recently installed Prime Minister of India, Narendra Modi, who decided shortly after assuming power to take on the taboo topic of open defecation in India and establish an explicit policy goal of ending the practice by 2019 (The Economist 2014). Modi’s use of the “bully pulpit” to focus attention on this challenge followed the publication of studies and assessments that documented the social, economic, and environmental costs of poor sanitation in India.

¹⁰⁷ It is important to keep in mind, though, that, in spite of the relative ease with which governments may prioritize reforms of this nature, the challenge of actually changing the behavior of individuals and firms is non-trivial (Rosenzweig and Wolpin 1982; Zwane and Kremer 2007; and Pritchett 2013).

¹⁰⁸ For an introduction to this literature, see Pritchett and Woolcock 2004; WB 2006b; Hout 2012; Innes 2013; Persson et al. 2013; Andrews 2013; Pritchett 2013; Buch et al. 2014; Krasner and Risse 2014; and Lake and Farris 2014.

Box 7. Aligning External and Domestic Interests to Address Specific Problem types

The data from the *2014 Reform Efforts Survey* reveal significant differentiation among external assessments of government performance in terms of the degree to which they focus developing country reform efforts on different problem types.

The three external assessments that are most associated with changing the behavior and incentives of private individuals and firms (the “most influenced” problem type) are the EU’s Poverty Reduction Budget Support Program and “Variable Tranche” Initiative and the UNDP’s Human Development Index. These assessments use various means to encourage governments to meet health and education benchmarks, which involve incentivizing behavioral changes such as convincing individuals to wash their hands more regularly or paying parents to send [female] children to school. The GAVI Alliance and the Global Fund to Fight AIDS, Tuberculosis and Malaria, by contrast, have particularly high levels of agenda-setting assessment influence on issues of “inadequate management, leadership, strategic direction, focus or political will”. This result is not surprising, given that these “vertical” funds target very specific diseases that require buy-in at the highest levels of government (Bor 2007; Piot et al. 2007; Sherry et al. 2009).¹⁰⁹

In terms of orienting government reform priorities towards issues of implementation and the rule of law, we find that the WB and IFC’s Doing Business Report has succeeded in provoking governments in the developing world to pay more attention to cutting red tape, while the IADB has successfully used its assessments (and lending instruments) to place issues of safety and security on the policy agendas of governments in Latin America and the Caribbean. We also see in Figure 3.1 that the EU’s Association Agenda has substantially influenced the attention that developing country governments pay to issues of legislative reform and legal compliance, which is not surprising in light of the emphasis that the EU places on compliance with the *Acquis Communautaire* (Grabbe 2002; Kelley 2004; and Böhmelt and Freyburg 2013).¹¹⁰

4.1.3. There are limits to what external assessment suppliers can reasonably expect to influence.

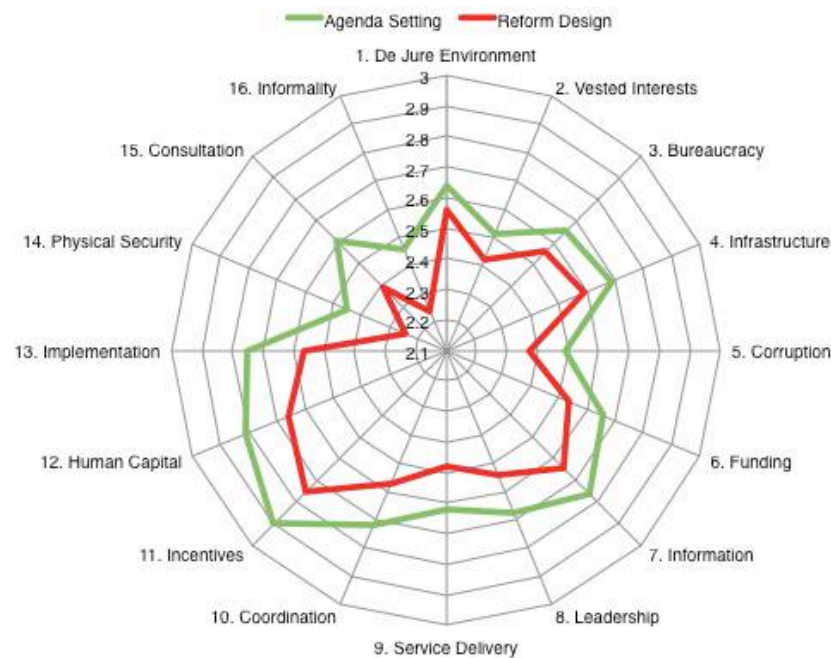
Some “problem types” are characterized by larger declines in external assessment influence between the agenda-setting and reform design stages than others. This is particularly true of the following four problem types in Figure 20: (1) disconnect between formal policies and informal practices; (2) inadequate involvement of local stakeholders in decision-making processes; (3) insufficient safety and security; and (4) inadequate or inconsistent implementation or enforcement of laws and policies. A common thread among these problem types is that they all involve issues of implementation or the rule of law. This pattern, in combination with the overall decline in assessment influence between agenda-setting and reform design, underscores the fact that there are limits to what external actors can reasonably expect to influence and help developing country government accomplish.¹¹¹

¹⁰⁹ One of the principal critiques of these vertical programs is that do not necessarily work within and help strengthen primary health care systems, but instead prioritize the fastest and most cost-effective means to combat diseases such as malaria and HIV/AIDS (Dervis et al. 2010). By comparison, “horizontal” funds and programs seek to improve primary care services in order to address a wide range of illnesses and improve overall health outcomes.

¹¹⁰ *Acquis Communautaire* is the accumulated body of EU law.

¹¹¹ See Devarajan and Holmgren 2001; Andrew 2009, Andrew 2013; Booth 2011; Buch et al. 2014; and Parks 2014.

Figure 20. Assessment Influence by Policymaking Phase and Problem Type



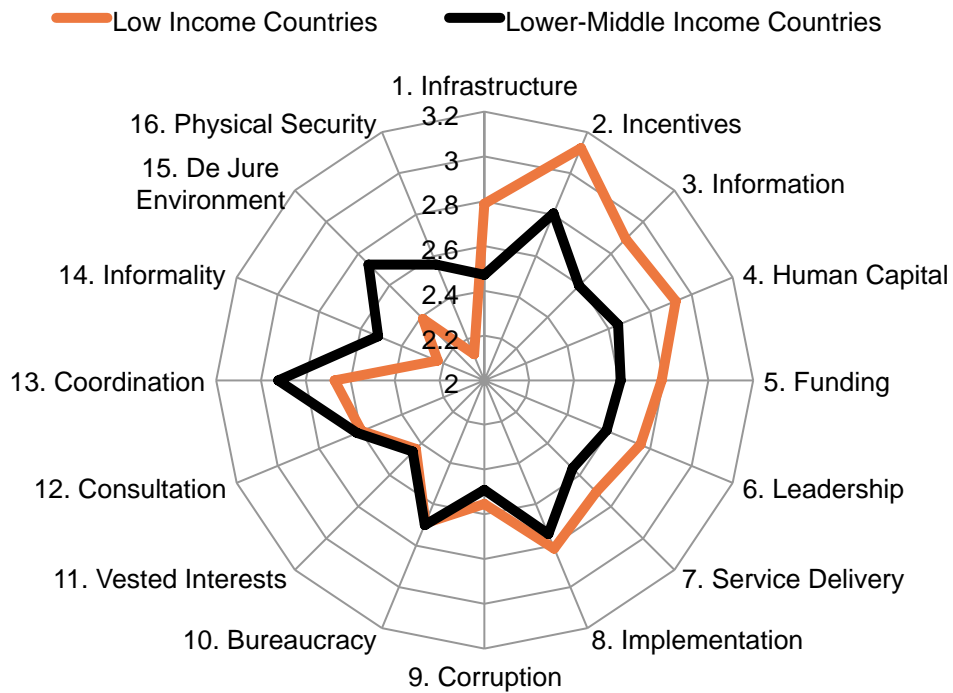
Note: Assessment influence is on a scale from 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". The spider graph displays both the agenda-setting and reform design influence of external assessments on each problem type. Each problem type has a dedicated radial line emanating from the center of the graph. Higher values of influence are located further from the center of the graph on each line. Problem types are ordered clockwise, beginning with 1, by the magnitude of the decrease in assessment influence from the agenda-setting to reform design stage (from smallest to largest).

By contrast, in four other cases—inconsistent and unclear drafting of laws and policies, overly centralized decision-making, overly burdensome or inefficient bureaucratic procedures, and insufficient equipment, technology or infrastructure—external assessments seem to have nearly equal levels of reform design and agenda-setting influence. This may be a reflection of the fact that these problems are largely technical in nature, and it relatively easier for development partners and other suppliers of external assessments to help design reforms that address these problems.

This speaks to an important debate about whether external actors are encouraging *form-based* reforms (e.g., the creation of new laws or agencies) or *function-based* reforms (e.g., the timely and impartial implementation of existing laws and regulations). A growing literature on isomorphic mimicry is concerned that development partners encourage developing countries to pursue cosmetic changes that “mimic” the formal structures of developed countries rather than improving the de facto quality of governance (WB 2006b; Andrews et al. 2013; Andrews 2013; Sjöstedt 2013; Pritchett et al. 2013; Samuel 2013; and Buch et al. 2014).

We also find (see Figure 21) that external assessment suppliers tend to steer low-income countries towards form-based changes and capacity building measures (e.g., staffing, funding, equipment, and information and data collection systems). By contrast, external actors are more likely to persuade policymakers in lower-middle and upper-middle income countries to address issues related to the effective functioning of public sector institutions (e.g., corruption, political interference, and the decoupling of informal practices and norms from formal policies and institutions). These findings corroborate the evidence presented in Andrews (2013), Pritchett et al. (2013), and Buch et al. (2014).

Figure 21. Assessment Influence by Income and Problem Type



Note: Assessment influence is on a scale from 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Problem types are ordered clockwise, beginning with 1, with the problem type with the largest difference in assessment influence between low-income countries and lower-middle income countries first and the problem type with the smallest difference in assessment influence between low- and middle-income countries sixteenth. We calculated this difference as assessment influence in low-income countries minus assessment influence in lower-middle income countries.

Box 8. Cross-Country Benchmarking Initiatives and Tackling De Facto Policy Problems

Notably, the three external assessments that have most effectively encouraged governments to tackle de facto policy problems (e.g., the decoupling of informal practices and norms from formal policies and institutions) are all *high-profile, cross-country benchmarking initiatives*. These initiatives include: the WB and IFC's Doing Business Report, the UN's MDGs, and the MCC's Eligibility Criteria and Country Scorecards. This result may be surprising to some observers in light of recent critiques that take aim at precisely these assessments (Arruñada 2007; Goldsmith 2011; and Krever 2013).¹¹²

One potential explanation for this pattern in the data is that cross-country benchmarking initiatives often lack detailed, country-specific diagnostic and advisory content, which may preserve the policy autonomy and maneuverability of decision-makers in developing countries and thus enable them to “crawl the design space” (Pritchett et al. 2012) in search of difficult-to-identify solutions that fit the local context and address de facto problems (Natsios 2010; Sjöstedt 2013; and Perakis and Savedoff 2015). Elsewhere in this report we find that developing country policymakers value country-specific diagnostics that provide actionable policy advice and technical assistance as a *general matter*; however, when it comes to confronting the most deeply rooted and difficult-to-resolve problems in their societies, they seem to prefer cross-country benchmarking initiatives that provide broad and flexible policy guidance.

4.2. External Assessment Characteristics and Uptake Among Developing Countries

There is a large theoretical literature that has generated dozens of hypotheses about the determinants of assessment influence. Combining the data from the *2014 Reform Efforts Survey* with information on the observable attributes of external assessments (e.g., the identity of the assessment sponsor, the geographic and policy domain focus of the assessment, the provision of specific policy recommendations, and the inclusion of financial or reputational rewards and/or penalties), we were able to analyze and test some of these hypotheses.¹¹³

We find that only a small number of factors appear to have an impact on an assessment's policy influence. Figure 22 calls attention to five statistically significant findings:¹¹⁴

1. Country-specific diagnostics generally exert greater influence than cross-country benchmarking exercises.
2. Assessments that rely on host government data are more influential.
3. Governance assessments face the greatest obstacles to in-country uptake and influence.
4. Intergovernmental organizations with a global reach produce more influential assessments.

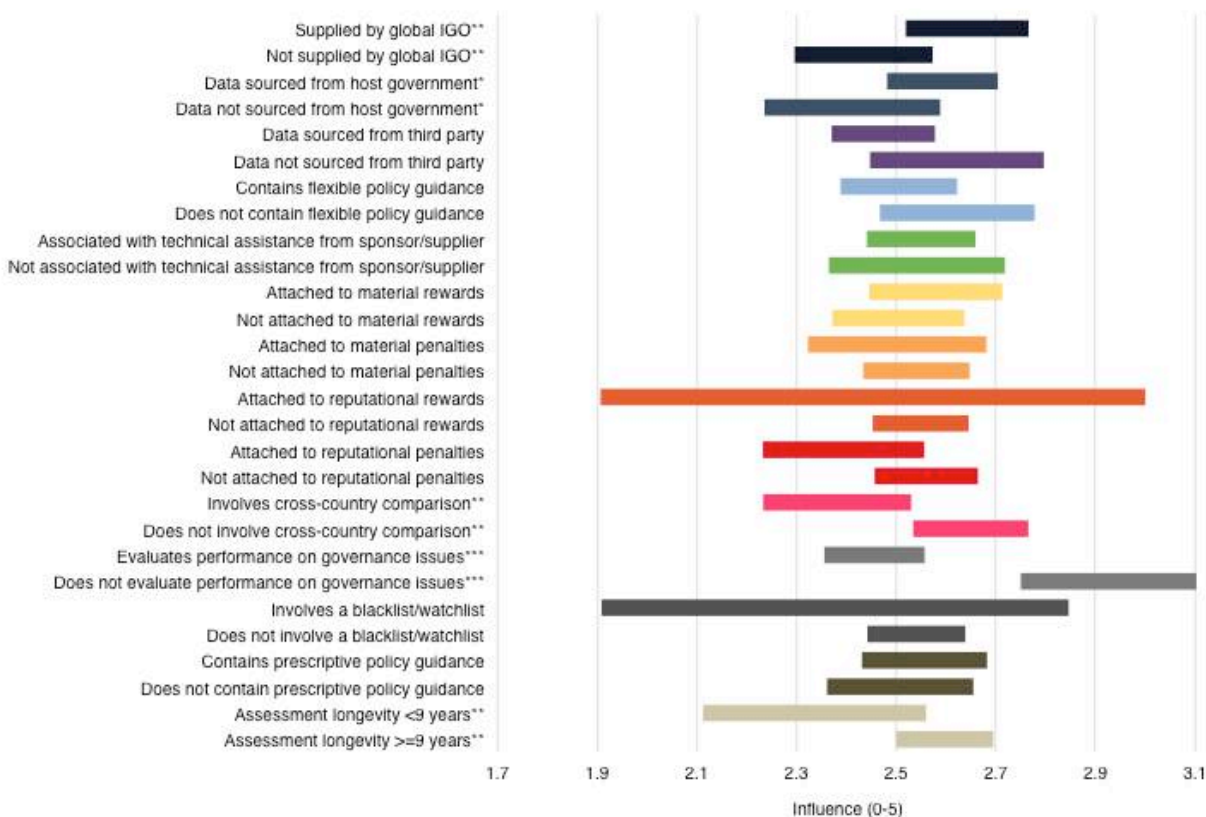
¹¹² These findings contradict recent criticisms that the Doing Business Report incentivizes governments to focus on solving de jure rather than de facto business regulatory problems (Kaufmann et al. 2005; Hallward-Driemeier et al. 2010; and Hallward-Driemeier and Pritchett 2011). Our findings are more consistent with a recent independent evaluation of the WB's Doing Business indicators, which concluded: “the [Doing Business] indicators have not distorted policy priorities or encouraged policymakers to make superficial changes to improve rankings” (IEG 2008: xix).

¹¹³ Independent of the data collected through the *2014 Reform Efforts Survey*, we assembled a database on the observable attributes of the external assessments of government performance that participants were asked to evaluate. The assessment-level variables in this database include: the assessment sponsor; country-specific or cross-country focus; inclusion of specific policy recommendations; whether the assessment is attached to any financial and/or reputational rewards; whether the assessment is attached to any financial and/or reputational penalties; and the policy domain scope of the assessment. More information is available in [Appendix G](#).

¹¹⁴ Difference-in-means tests were conducted for most of the variables listed in [Appendix G](#), with few exceptions.

- The longer an assessment's track record of publication, the more influential it becomes vis-à-vis others.

Figure 22. External Assessment Characteristics and Influence at the Agenda-Setting Stage



Note: * Indicates a significant difference-in-means at $p < 0.10$, ** indicates a significant difference-in-means at $p < 0.05$, and *** indicates a significant difference-in-means at $p < 0.01$. The width of each bar corresponds to the 95% confidence interval surrounding the average assessment influence score for assessments with the indicated characteristic. Assessment influence is on a scale from 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Only assessments with a sample size of at least 10 observations were considered for analysis.

4.2.1. Authorities generally prefer country-specific, rather than cross-country assessments.

Scholars have long debated whether cross-country benchmarking initiatives or country-specific assessments exert greater influence over the policy decisions and actions of governing leaders in developing countries. Proponents of cross-country performance comparisons argue that they are particularly influential because governments respond to social pressures that signal "appropriate" and "inappropriate" forms of behavior (Schueth 2011; Sharman 2011),¹¹⁵ and encourage countries to compete with their peers by highlighting a country's relative strengths and weaknesses (Kelley and Simmons 2014b). Conversely, other scholars argue that assessments more tailored to the local circumstances of

¹¹⁵ Various critiques of cross-country benchmarking include Arruñada 2007; Arndt 2008; Pham 2009; Goldsmith 2011; Wolff et al. 2011; Høyland et al. 2012; Davis et al. 2012; Hood 2012; and Krever 2013.

particular countries enjoy higher credibility and local resonance with the domestic authorities (Momani 2007; Lombardi and Woods 2008; and Edwards 2011).¹¹⁶

Comparing the agenda-setting and reform design influence scores for the cross-country benchmarking initiatives and country-specific assessments in our sample, we find that the average country-specific assessment is significantly more influential than the average cross-country benchmarking exercise (see Figure 22).¹¹⁷ That is to say, country-specific diagnostic and advisory tools generally have more influence than cross-country benchmarking exercises in shaping the reform decisions of governing policymakers. This finding poses a significant challenge to the conventional wisdom, as much of the existing literature on this topic emphasizes the competition-inducing effects of cross-country benchmarking. The local resonance and uptake of country-specific diagnostics by developing country policymakers arguably constitutes a major blind spot among those who study the phenomenon of external assessments of government performance.¹¹⁸

In Figure 23, we go a step further and look at how perceptions of cross-country benchmarking exercises and country-specific assessments vary by the five stakeholder groups included in our sample.¹¹⁹ Nearly all stakeholder groups appear to view country-specific assessments as more influential than cross-country benchmarking exercises, with private sector representatives serving as the only exception.¹²⁰ It is also important to note that compared to development partner staff, host government officials assign relatively high-assessment influence scores to both country-specific assessments and cross-country benchmarking initiatives at the agenda-setting stage.

¹¹⁶ A 2008 UNDP and Global Integrity study on the uses of governance data found that many practitioners view the currently available metrics as, “only loosely relevant to the daily work of putting together specific governance and anti-corruption reforms”. The report cites an example of a Latin American Government working group seeking to improve overall governance and anti-corruption performance. The group sought to learn from international corruption datasets, but quickly abandoned the effort because “nothing in the data identified actual points of intervention” (UNDP and Global Integrity 2008: 35-36). A 2008 evaluation of the WB and IFC’s Doing Business Report also calls attention to this issue of local resonance by providing the example of how the Chinese government has responded: “the Chinese authorities pay less attention to [Doing Business] than to other cross-country benchmarks on rule of law, the investment environment, global competitiveness, trade and logistics, and corruption, because they consider China’s poor [Doing Business] rankings inconsistent with its strong private sector growth” (IEG 2008: 44).

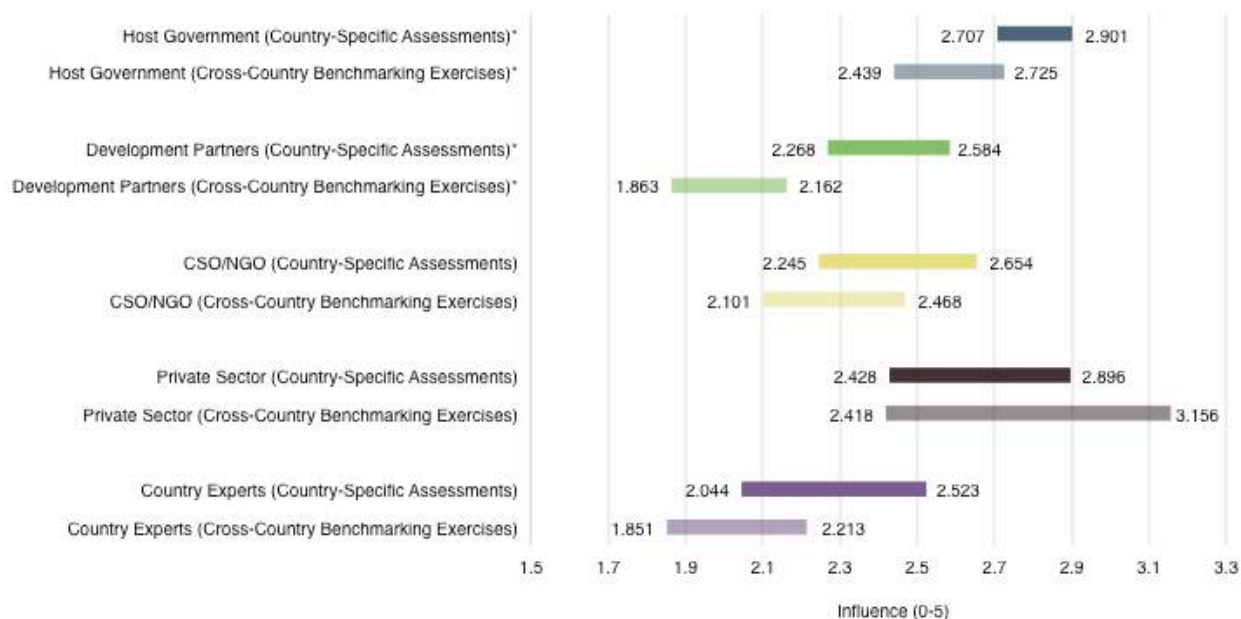
¹¹⁷ Difference of means tests between the average influence of cross-country benchmarking initiatives and country-specific studies reveals a statistically significant difference at the agenda-setting stage ($p=0.011$) and reform design stage ($p=0.014$). While there is some overlap between the 95% confidence intervals at both the agenda-setting (0.02) and reform design (0.03) stages, this overlap is not sufficiently large to render the tests statistically insignificant.

¹¹⁸ Many country-specific diagnostic and advisory tools are developed to identify country-specific policy recommendations regarding possible policy, legal, regulatory, institutional, and administrative reforms.

¹¹⁹ The bars indicate 95% confidence intervals around the point estimates.

¹²⁰ Moreover, the perception of the influence of country-specific assessments as being more influential than cross-country benchmarking exercises is significant at $p<0.05$ among the two stakeholder groups most central to the policymaking process: host government officials and development partner staff.

Figure 23. The Agenda-Setting Influence of Country-Specific Assessments vs. Cross-Country Benchmarking Exercises, by Stakeholder Group



Note: * Indicates a significant difference-in-means at $p < 0.05$. The width of each bar corresponds to the 95% confidence interval surrounding the average assessment influence score, as rated by members of the indicated stakeholder group. Assessment influence is on a scale from 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Only assessments with sample size of at least five observations were considered for analysis within a given stakeholder group.

One way of interpreting this finding is that governing decision-makers in developing countries value the nuance and context specificity contained in tailored, country-specific analysis, but they are also more mindful of and motivated by the forces of inter-jurisdictional competition. This is perhaps one way of knitting together insights from the literature on the “competitive state” (Cerny 1997; Sharman 2011; and Schueth 2011) and the body of research on why governance assessments need to better account for the local actors, institutions, and incentives that instigate and impede change (Dahl-Østergaard et al. 2005; Hyden 2008; and Unsworth 2009).

4.2.2. Assessments that rely on host government data are more influential.

How external assessments source their data also appears to matter, but not in the way suggested by the conventional wisdom. While it is often argued that independent third parties should be in charge of data collection and policy surveillance to mitigate the risk of biased advice and ensure the credibility of incentives and disincentives,¹²¹ we find that the most influential assessments are those that rely on host government data.¹²² This pattern in the data suggests that performance measurement by a developing country government increases the local resonance of the assessment and the resulting policy recommendations (Swedlund 2013; Busia 2014).

¹²¹ There are several reasons why sovereign states might delegate responsibility for surveillance and evaluation of other countries to third parties (Rodrik 1996; Parks and Tierney 2004; Hagen 2009; and Fang and Stone 2012). States may delegate surveillance and evaluation authorities to a third party in order to capture scale economies, benefit from higher levels of perceived credibility and neutrality (Hawkins et al. 2006), and prevent “the monitored” from behaving opportunistically by directly influencing the measurement process (Parks 2014; Momani 2007).

¹²² The point-biserial correlation coefficient between *AgendaSettingInfluence_Q31* and *PartyResponsible_2* is 0.120, a value that is significant at $p < 0.05$. This indicates a significantly positive relationship between the use of host government data and the influence of external assessments.

4.2.3. Governance assessments face the greatest obstacles to in-country uptake and influence.

External assessments focused on governance issues wield less policy influence than their economic and social counterparts, according to our survey participants (see Figure 22). While disappointing, this may be unsurprising as theory and common sense suggest that any assessment encouraging reforms which threaten to remove an official from office or threaten existing power structures will likely induce a weak response (Parks and Rice 2013; Smets 2014).¹²³

4.2.4. Intergovernmental organizations with a global reach produce more influential assessments.

Interestingly, assessments supplied by global intergovernmental organizations appear to be more influential than assessments provided by other organizations. This distinction is likely a reflection of the significant in-house technical expertise that these intergovernmental organizations possess, their “ground game” in developing countries, and the “influence dividend” associated with higher levels of assessment familiarity.¹²⁴

4.2.5. The longer an assessment’s track record of publication, the more influential it becomes vis-à-vis others.

According to expectations theory, people make today’s decision based on their expectations of the future (Muth 1961; Grossman 1980).¹²⁵ We find preliminary support for this hypothesis that the longer an assessment has been active—and is expected to remain active—the more influence it will gain with developing country policymakers (Schimmelfenning and Sedelmeier 2004; Parks 2014).¹²⁶

4.3. Which Assessment Attributes Do Not Appear to Influence Behavior Change?

There are a number of other assessment attributes that have received a great deal of attention in the scholarly literature, but the following two in particular do *not* appear to influence on the policy behavior of developing country policymakers:

1. Neither incentives nor penalties seem to easily explain assessment influence.
2. Assessment influence might depend upon the flexibility and specificity of policy advice.

¹²³ The point-biserial correlation coefficient between *AgendaSettingInfluence_Q31* and *GovernancePolicyBroad* is -0.395, a value that is significant at $p < 0.01$. This indicates a significantly negative relationship between the focus of an assessment on governance issues and its influence at the agenda-setting stage.

¹²⁴ Assessment familiarity is a function of the effort that these organizations devote to branding and disseminating their assessments and findings (IEG 2008, 2012; WB 2010, 2014; Doemeland and Trevino 2014; and Parks et al. 2014). See Figure 9.

¹²⁵ This insight is also related to a broader literature on the clarity and the predictability of the benefits that policymakers can expect to achieve by seeking to improve their governments’ performance on an external assessment (Perrin 2002; Lawson et al. 2005; Nsouli 2006; Molina-Gallart 2008; and Savedoff 2011).

¹²⁶ We operationalize “assessment longevity” as the number of years between 2004-2013 when a given assessment was produced and published. We split our sample according to whether or not a given assessment had been in operation for at least nine years (the approximate average assessment longevity within the period of interest). The Pearson correlation coefficient between *ActiveYears* and *AgendaSettingInfluence_Q31* is 0.251, a value significant at $p < 0.05$. The point-biserial correlation coefficient between our dichotomized measure of assessment longevity and *AgendaSettingInfluence_Q31* is 0.244, a value also significant at $p < 0.05$.

4.3.1. Neither incentives nor penalties seem to easily explain assessment influence.

Few hypotheses have received more attention than the assertion that external assessments of government performance will exert greater influence when they are tied to material and reputational benefits and penalties. Economists and political scientists frequently argue that the strength of the benefits and the penalties that are attached to a country's "compliance" with the requirements of an assessment is a key predictor of policy influence with the domestic authorities in developing countries.¹²⁷

The intuition underpinning this argument is that external actors can strategically alter a government's policy behavior by increasing the expected return on reform and increasing the expected costs of inaction and reform reversals.¹²⁸ A related literature suggests that governments are also responsive to *reputational* rewards and penalties (Sharman 2008, 2009; Pitlik et al. 2010; and Kelly and Simmons 2014a, 2014b). For example, governments that need to signal their credibility to donors, creditors, or investors may be more likely to make thoroughgoing efforts to meet the reform requirements of external assessment sponsors and suppliers (Parks 2014; Buch et al. 2014).

However, we do not find that external assessment influence systematically depends upon the provision of material or reputational rewards.¹²⁹ Nor do assessments appear overall to wield more influence when non-performance by "target" states can result in material or reputational penalties. We also find no evidence that watch lists and blacklists are more or less influential than other assessments.

4.3.2. Assessment influence might depend upon the flexibility and specificity of policy advice

Another question that we set out to answer is whether descriptive or prescriptive assessments exert greater influence on the decisions and actions of policymakers in developing countries. Some argue that assessments providing policy advice are more likely to induce a policy response from public sector leaders in developing countries (UNDP and Global Integrity 2008; Schimmelfennig and Sedelmeier 2004; and Parks 2014). Others propose that developing country leaders value, benefit from, and seek out policy flexibility and maneuverability, which descriptive assessments preserve (Grindle 2004; Rodrik 2007; and Andrews 2013).

To this end, we tested whether assessments that provide the government with the flexibility needed to successfully adapt to changing circumstances during the design and implementation of reform efforts are more influential. We also examined whether prescriptive assessments that provide concrete policy guidance exert greater influence than descriptive assessments.

We find that prescriptive assessments are slightly more influential than descriptive assessments, and that decision-making elites in the developing world want more specific rather than less specific forms of policy guidance.¹³⁰ While neither of these results reach conventional levels of statistical significance, they do

¹²⁷ For an introduction to this literature, see Noland 1997; Kelley 2004; Lawson et al. 2005; Hovi et al. 2005; Schimmelfennig and Sedelmeier 2004; Krasner 2011; Molenaers 2012; David-Barret and Okamura 2013; and Parks 2014.

¹²⁸ On the other hand, De Renzio and Woods (2008) and Barder (2010) have expressed skepticism that extrinsic incentives can substantially influence the policy behavior of governing decision-makers in developing countries.

¹²⁹ Assessments with high stakes rewards have both exceptionally high—and exceptionally low—levels of influence. For example, while the EU's Association Agenda, the EU's Stabilization and Association Agreements and Performance-Based Funding from the Global Fund to Fight AIDS, Tuberculosis, and Malaria are ranked among the top five most influential assessments, the MCC's Eligibility Criteria, the AsDB's Results-Based Lending, and DFID's Resource Allocation Model are ranked 89th, 92nd, and 94th, respectively.

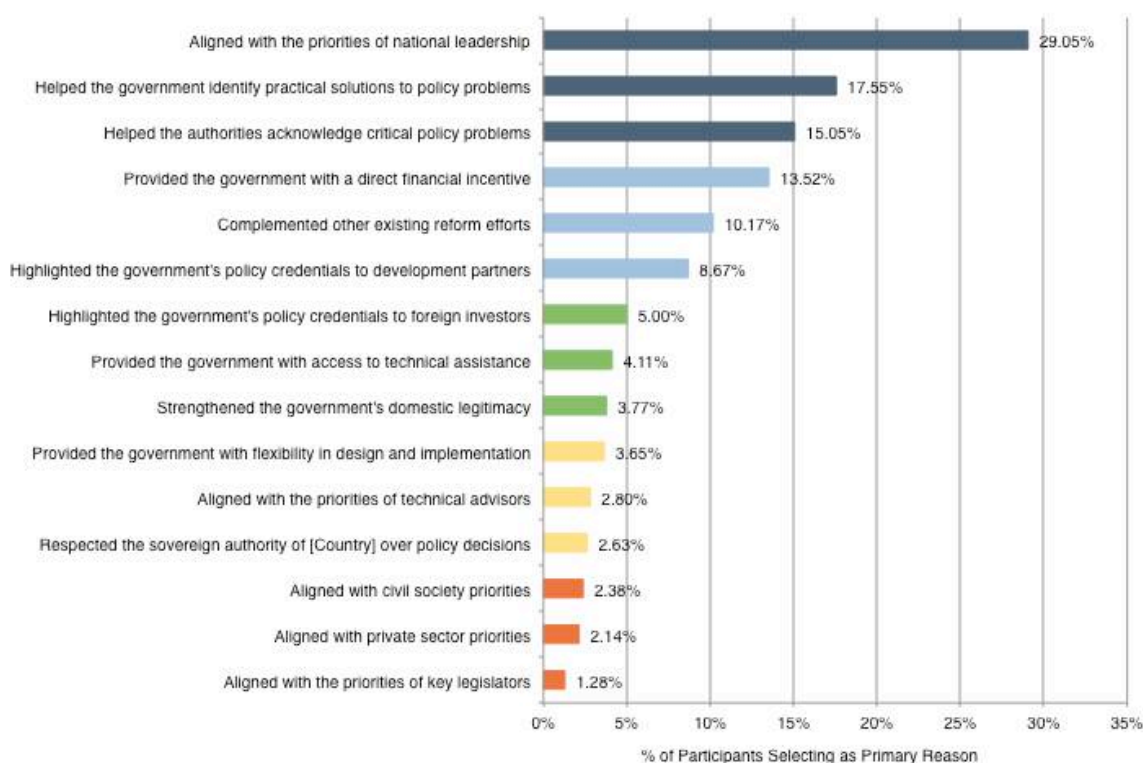
¹³⁰ Some external assessments are produced and delivered by a small group of experts who do not devote much effort to interacting with domestic authorities in target countries, while other assessments have expert advisers who are willing and able to assist the authorities. We tested out whether assessments that provide the government with access to the technical assistance of development partner staff are any more influential. The difference of means test in Figure 22 suggests that there is no discernible difference in the influence levels of assessments that involve the provision of technical assistance and those that do not involve the provision of technical assistance.

preliminarily suggest that development scholars may be more concerned about the provision of specific policy advice diminishing the policy autonomy and flexibility of governments than development policymakers and practitioners.¹³¹

4.4. Alignment with local priorities is a key determinant of external assessment influence.

In the interest of obtaining a deeper understanding of the drivers of assessment influence, we asked participants to identify the *primary* reason for the outsized influence of an external assessment. The question posed to each participant probed the reasons why a specific assessment (that was familiar to the participant) exerted significant influence over a defined set of reform efforts in a particular country and policy domain.¹³² The results are reported in Figure 24.

Figure 24. The Primary Reason for External Assessment Influence



Note: Shading is by quintile and reflects the relative percentage of survey participants selecting one explanation of assessment influence versus another.

The most frequently selected explanation for assessment influence—accounting for 29% of participants—was that an **assessment was influential because it promoted reforms that aligned with the priorities of national leadership.**

¹³¹ On this point, also see Parks and Rice (2013). The difference of means tests are reported in Figure 22.

¹³² Each participant was assigned and asked about a single assessment among those that they perceived as having the highest level of influence in both the agenda-setting and reform design stages of the policy process. “Highest” was determined by first taking the sum of 0-5 scores provided in questions 31 and 32 and determining which assessment(s) had the maximum sum. If there was a tie, then the assessment with the greatest response value in question 31 was selected. If there was still a tie between multiple assessments, then a single assessment was randomly selected from among this final list.

The second most popular explanation, accounting for 18% of participants, was that **an assessment was influential because it helped the authorities identify practical approaches to solve critical policy problems**. This suggests that there is wisdom in recent calls for donors and international organizations to “go with the grain” of existing socio-economic and political forces in the countries where they work, rather than foisting unpopular and disruptive agendas upon their counterpart governments (Kelsall 2011, 2012; Srivastava and Larizza 2013; and Levy 2014).

However, the fact that 15% of participants believe external assessments were influential because they helped the host government more fully acknowledge and appreciate policy problems also suggests that assessment sponsors should not soft-pedal or self-censor. There is still scope for raising uncomfortable, and perhaps controversial, policy issues in ways that are not counterproductive, as long as they focus on solving problems that are not diametrically opposed to the interests of national leadership.

Interestingly, when required to identify the primary driver of external assessment influence, 13.5% of participants cited the lure of a financial incentive.¹³³ Domestic authorities not only care about external rewards, but actively undertake and design reforms to secure those rewards. This finding lends some support to the argument that the presence of a material reward can, in certain contexts, increase the policy influence of an external assessment of government performance (Schimmelfennig and Sedelmeier 2004; Parks 2014).

4.4.1. The Primary Reason for Assessment Influence, by Policy Domain

More often than not, participants believed that **the primary reason for assessment influence within policy domains was related to alignment with domestic priorities and needs** (see Table 3). This might include the extent to which assessments: promoted reforms that complemented existing efforts, helped the government understand the nature of critical policy problems, or provided practical approaches for addressing difficult policy problems. The unique attributes of policy domains also appear to yield some differentiation in the primary reason given for assessment influence (e.g., health, infrastructure, and education lend themselves to technical solutions and measurable outcomes).

Notably, there were only three policy domains in which participants identified financial incentives as the primary reason for influence: civil service, macroeconomic management, and democracy. One potential explanation for this pattern is that reform in these policy domains is particularly costly to the parties involved, thereby making financial incentives necessary to alter the cost-benefit calculus of reform. However, this is only one of several potential explanations and further research will be needed to better understand these results.

Table 4. The Primary Reason for Assessment Influence in Each Policy Domain

Primary Reason for Assessment Influence	Policy Domains (% of Participants Selecting Reason)
It helped the government clearly identify practical approaches for addressing critical policy problems.	Customs (33.33%) Health (26.96%) Infrastructure (26.89%) Education (26.03%) Trade (25.64%) Land (25.00%) Anti-corruption and transparency (23.56%) Investment (23.39%) Labor (20.97%) Decentralization (20.67%)

¹³³ The relative increase in the popularity of this explanatory factor—from the sixth most popular response (in question 33) to the fourth most popular response (in question 34)—suggests that, when a financial incentive contributes to the influence of an external assessment, it is also likely to be the primary driver of its influence.

	Justice and security (20.44%) Environmental protection (19.81%) Public expenditure management (18.46%)
It promoted reforms that complemented other existing reform efforts in [Country].	Family and gender (30.00%) Energy and mining (22.86%)
It provided the government with a direct financial incentive to undertake specific reforms.	Civil service (18.71%) Macroeconomic management (18.67%) Democracy (15.29%)
It helped the authorities fully acknowledge the critical nature of policy problems that were not otherwise completely understood or appreciated.	Business regulatory environment (25.19%) Finance, credit, and banking (22.10%) Tax (21.59%) Agriculture and rural development (21.12%) Social protection and welfare (16.16%)

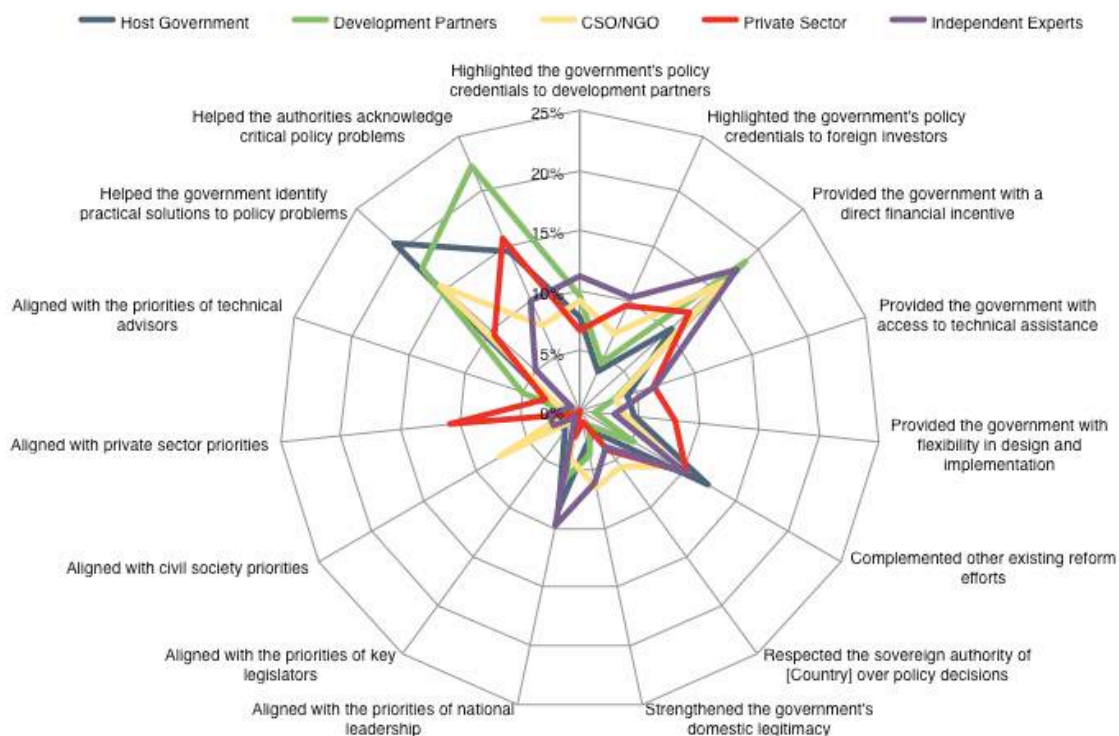
4.4.2. The Primary Reason for Assessment Influence, by Stakeholder Group

Different stakeholder groups perceive and interact with external assessments of government performance in unique ways. Civil society and private sector participants may provide unique insight into how external assessments are seen by those outside of government. Host government officials—often the targets or suppliers of these external assessments—are arguably best positioned to articulate how external assessments are received and used by the authorities themselves. Development partners who finance, design and promote assessments may suffer from some degree of confirmation bias;¹³⁴ however, they are also directly involved in discussions with government decision-makers about the content contained in these assessments and may be well-positioned to speak to their perceptions of local resonance and uptake of their policy advice.

Overall, survey participants across the stakeholder groups expressed a clear preference for three primary explanations of assessment influence: (1) assessments helped the government identify and acknowledge policy problems; (2) assessments helped the government identify practical solutions to problems; and (3) assessments provided financial incentives for reform (see Figure 25).

¹³⁴ Development partner perceptions of the reasons for assessment influence may not reflect how host government policymakers actually perceive and use the external assessments they produce.

Figure 25. The Primary Reason for External Assessment Influence, by Stakeholder Group



Note: The scale indicates the percentage of survey participants from each stakeholder group who identified each explanation as the primary reason for assessment influence. Reasons for influence are not presented in any particular order, but are clustered by their similarity to one another. Each explanation of assessment influence has its own dedicated radial line emanating from the center of the graph. Higher percentages are located further from the center of the graph on each line.

Development partners were more likely than other participants to cite helping the government identify and acknowledge policy problems as the main reason for an assessment's influence. By contrast, civil society and independent expert participants were among those least likely to select this reason. This pattern suggests a potential source of confirmation bias in development partner perceptions, perhaps because they administer many external assessments of government performance with this goal in mind (Kelley and Simmons 2014b).¹³⁵

Host government participants were less likely than any other stakeholder group to choose the provision of a financial incentive as the primary reason for assessment influence. Only 10% of host government participants chose this reason, compared to 16-19% of development partners, independent experts, and civil society representatives. This suggests two things. First, host government officials are less likely than "outsiders" to perceive material rewards as the *primary* reason for assessment influence.¹³⁶ Second, while

¹³⁵ There is likely some degree of perception bias in the determinants of external assessment influence selected by all or most stakeholder groups, not only the development partner participants.

¹³⁶ In contrast to host government officials, independent country experts indicate that financial incentives are the single most important determinant of external assessment influence. This may reflect the arms-length relationship the independent experts have with the policymaking process in developing countries. The relative neutrality of independent experts may also suggest bias on the part of host government participants who, by comparison, report financial incentives to be less influential. On the other hand, the "distance" of independent experts from the policymaking process may render them less knowledgeable about the inner-workings of government.

assessment suppliers view financial leverage as important to assessment influence, the targets of those assessments may believe that the role of financial leverage is overstated.

Notably, civil society and private sector participants are significantly more likely to emphasize their own role in the policymaking process.¹³⁷ By contrast, we observe that less than 3% of participants from any other stakeholder group selected alignment with either civil society priorities or private sector interests as a primary reason for assessment influence.

4.5. Survey Participants Speak to Factors Contributing to External Assessment Influence

In order to better understand how the attributes of external assessments influence local uptake by developing country authorities, we also asked survey participants to identify, from a fixed list of options, all of the factors that contributed to the influence of a given external assessment.¹³⁸ The responses that we received largely reinforce the finding that **external assessment influence is largely a function of whether it reinforces and advances a “target” government’s pre-existing policy priorities.**

Overall, participants indicated that the most influential external assessments helped governments understand policy problems and identify practical solutions to those problems. Approximately 59% of participants said that influential assessments helped the government clearly identify practical approaches for addressing critical policy problems (see Figure 26). This was the single most frequently selected response option.¹³⁹ Similarly, 54% of participants said that influential assessments helped the authorities fully acknowledge the critical nature of policy problems that were not otherwise entirely understood or appreciated.¹⁴⁰

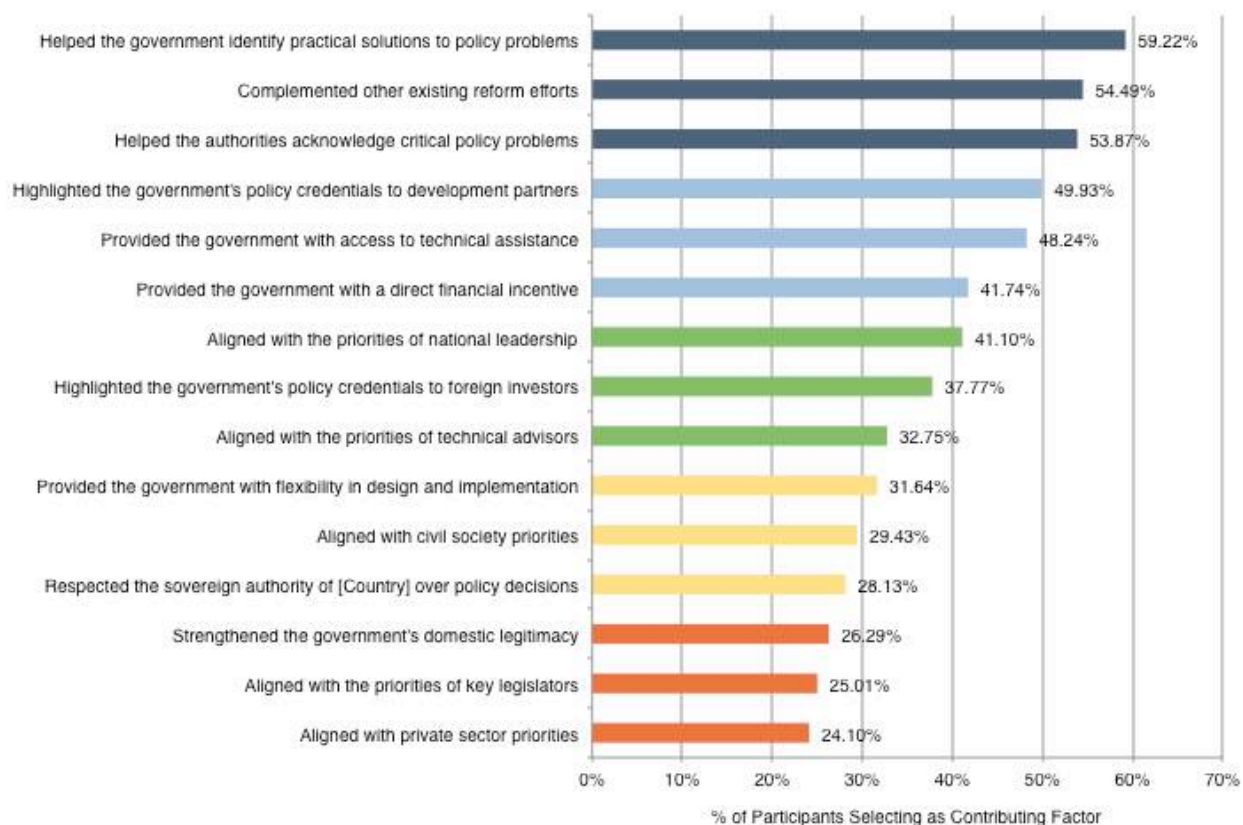
¹³⁷ It is conceivable that CSO/NGO and private sector participants are more knowledgeable about—and therefore more likely to answer questions pertaining to—reform efforts in which CSO/NGO and private sector groups played a more central role. This pattern in the data likely reflects the tailored design of the survey questionnaire and the “bounded” perceptions of actors who play a less central role in the policymaking process (Weyland 2007).

¹³⁸ Our fixed list of reasons for assessment influence was developed through a literature review (e.g., Schimmelfennig and Sedelmeier 2004; Girod and Walters 2012; Simmons and Kelley 2014a, 2014b; and Parks 2014), questionnaire pre-testing, and cognitive interviews with policymakers and practitioners at the OpenGov Hub, AidData, the College of William & Mary’s Institute for the Theory and Practice of International Relations, and Harvard University’s Kennedy School of Government. We sought to minimize participant burden and maximize the usability and salience of our resulting survey questions.

¹³⁹ This finding is consistent with a recent evaluation, which found that the WB’s “clients do not always see the value of [analytical and advisory activities] if [they are] not demand-driven and [do] not provide options and solutions to problems” (IEG 2012: 52).

¹⁴⁰ Kelley and Simmons (2014b) argue that an assessment’s ability to call attention to issues or problems that have not been fully understood or appreciated by the domestic authorities will co-vary positively with its policy influence (Kelley and Simmons 2014b). Conversely, if an assessment highlights concerns that are already widely understood by decision-makers in the target government, one would expect it to wield less influence.

Figure 26. Contributing Factors to External Assessment Influence



Note: Shading is by quintile and reflects the relative percentage of survey participants selecting one explanation of assessment influence versus another.

Participants also identified assessments as influential when they promoted reforms that complemented existing reform efforts, which suggests that public sector leaders in developing countries pay more attention to external assessments when they align with pre-existing government interests, policies, and programs.¹⁴¹ Governments appear to be selective and strategic about the assessments they use, picking and choosing those assessments perceived as helping move the country in a direction that aligns with domestic priorities. Moreover, governments do not appear to be easily cajoled or coerced into pursuing reforms that align with donor priorities but displace domestic priorities.¹⁴²

That is not to say that domestic authorities in developing countries are unconcerned with external rewards or credibility with donors, investors, and creditors. **Nearly 50% of participants indicated the government's desire to signal its policy credentials to the donor community contributed to the outsized influence of an external assessment.** Similarly, 42% of participants indicated that

¹⁴¹ This finding suggests that a key determinant of external assessment influence is alignment with domestic beliefs, interests, and priorities in the country that is being assessed. Previous research suggests that governing elites in developing countries are motivated by a desire to stay in power (Bueno de Mesquita 2011). As such, external assessments promoting changes that align with the priorities of national leadership may be more influential (Parks 2014). Conversely, if an assessment is perceived to undermine the sovereign control of the domestic authorities over their own policymaking process, it may exhibit a lower level of policy influence (Dharmasaputra and Wahyudi 2014).

¹⁴² This finding is consistent with research that suggests a key determinant of assessment influence is "country ownership" of the diagnostic and advisory content contained in the assessment (Busia 2014).

assessment influence was attributable, at least in part, to a direct financial incentive. Nonetheless, these factors appear to be secondary, rather than primary, considerations.¹⁴³

Notably, significantly fewer participants reported that external assessments of government performance were influential when they aligned with the priorities of various domestic actors outside of the executive branch of government. Only 29% of participants said that assessments were influential in part because they aligned with the priorities of civil society organizations.¹⁴⁴ Similarly, 25% of participants said that influential assessments promoted reforms that aligned with the priorities of key legislators, and 24% of participants said that influential assessments promoted reforms that aligned with the priorities of influential private sector groups.

The results may point to the generally low level of policymaking influence that actors outside of the executive branch of government have in developing countries. However, it is also possible that this pattern in the data reflects a bias “baked into” the very process of producing and delivering assessment results. There are good reasons to believe that assessment sponsors tend to focus their policy advice and outreach and consultation efforts on leading decision-makers in the executive branch whom can “speak on behalf of the government” (Parks 2014).¹⁴⁵

4.6. Position-Driven Perceptions of Contributing Factors to External Assessment Influence

We also wondered whether perceptions of external assessment influence might depend upon where one sits. Policy professionals holding positions with varying levels of authority are likely to interact with different stakeholders, engage at different stages of the policymaking process, have access to different types of information, and express different perspectives. In Figure 27, we highlight differences of opinion across participants from three different “position types”—organizational leadership, administrative management, and programmatic and technical staff—about why external assessments of government performance proved influential.¹⁴⁶

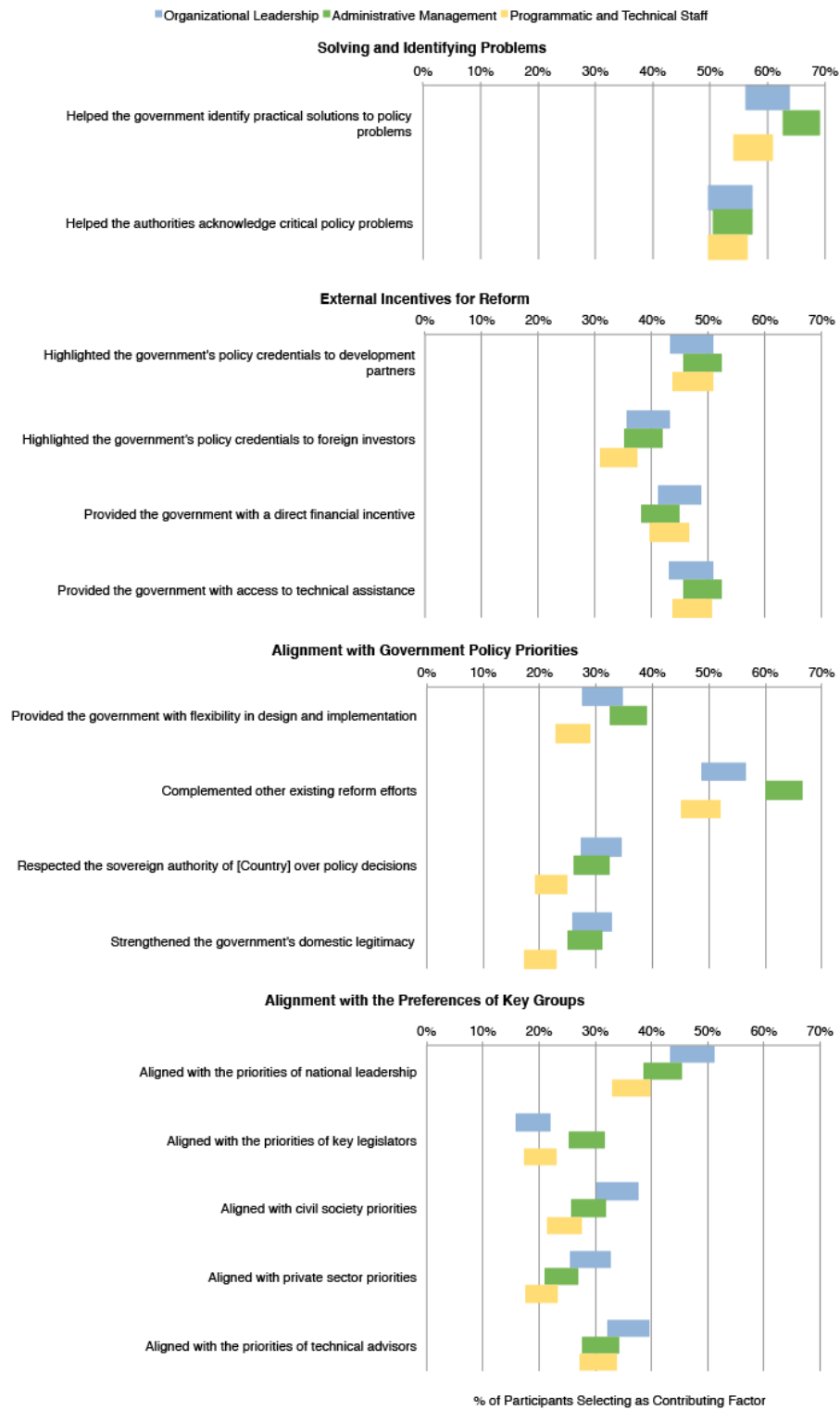
¹⁴³ Figure 26 also suggests that governments apparently use external assessments of their performance to obtain access to technical experts for help in implementing reform efforts (Easterly 2014).

¹⁴⁴ See Diamond 2008, Krasner 2011, McArthur 2013, Parks 2014, and Diofasi and Birdsall 2015.

¹⁴⁵ Previous research indicates that aid agencies and international organizations are often eager to ally themselves with “sympathetic interlocutors” in the executive branch with whom they share causal and principled development policy beliefs (Chwieroth 2007, 2009a, 2009b, 2010; Kahler 1992; Corrales 2006; Momani 2005; Andrews 2011; and Parks 2014).

¹⁴⁶ The vertical, y-axis lists 15 potential reasons for assessment influence participants could choose from. The horizontal, x-axis scale measures the weighted percentage of participants holding each of the three types of positions that selected a particular reason as contributing to assessment influence. Colored bars depict the 95% confidence intervals specific to each position type.

Figure 27. Contributing Factors to Assessment Influence, by Respondent Position Type



Note: The width of each bar corresponds to the 95% confidence interval surrounding the percent of each type of survey participant who selected a given reason for influence as a contributing factor.

Despite some variation, there is a relatively high level of agreement among participants from each of the three different position types.¹⁴⁷ Participants holding all three position types generally agreed that the following four factors explain the influence of external assessments: (1) ability to help the government better understand critical nature of specific policy problems, (2) highlighting the government's policy credentials to key development partners, (3) provision of development partner technical assistance, and (4) provision of a direct financial incentive to the government.

At the same time, some explanations for assessment influence provoked different reactions from those holding different position types. These differences of opinion seem to generally follow formal differences in the level of the survey participant's seniority, authority, and level of political-mindedness.¹⁴⁸ Additionally, the likelihood of any given reason for assessment influence being selected seems to increase with the participant's level of seniority and breadth of perception.

For example, less senior technical and programmatic staff are significantly less likely than more senior organizational leaders or administrative managers to select the following three reasons for assessment influence: (1) flexibility to successfully adapt to changing circumstances during the design and implementation of reform efforts, (2) strengthening the government's legitimacy among key domestic political constituencies, and (3) respecting the sovereign authority of the government over final policy decisions. These findings likely reflect the relatively apolitical nature of technical and programmatic staff work, but may also suggest a lack of satisfaction with the level of policy flexibility offered to those actually overseeing reform implementation efforts.

4.6.1. Senior government leaders and their deputies engage with external assessments in very different ways.

We also analyzed the divergence of opinion between host government participants holding different position types. Heads of state and heads of government felt more strongly than other host government participants that external assessments were influential when they respected their country's sovereign authority to make final policy decisions and promoted reforms that aligned with the priorities of national leadership. These results support the notion that developing country leaders are motivated by a desire to minimize the domestic audience costs of "bowing" to international pressure (Vreeland 2003).

Heads of state and heads of government are also less likely than any other position to accept the idea that external financial incentives encouraged the government to undertake specific reforms. Yet curiously, at the next level down of host government leadership (e.g., vice heads of state, vice heads of government, and their respective chiefs of staff) participants are *more* likely than any of the other positions within the host government stakeholder group to accept that external financial incentives encouraged the government to undertake specific reforms.

One way of interpreting this finding is that developing country governments behave strategically when they encounter external evaluators of their performance. Members of leadership, who are mindful of domestic audience costs, might project strength in the face of external pressures, while their deputies work behind the scenes to do what is necessary to reap external rewards from international organizations and foreign governments.

Notably, the highest-ranking civil servants in line ministries who often remain in office during political transitions—Secretary Generals, Permanent Secretaries Director Generals—responded in substantially

¹⁴⁷ Most of the variation falls within a narrow band (+/-10%) of deviation from overall average scores.

¹⁴⁸ Administrative managers were more likely to select any one reason as contributing to assessment influence than any other group.

different ways in comparison to other officials. They were less likely to indicate that external assessments were influential because of a desire on the part of the host government to highlight its policy credentials to foreign investors or development partners. These government technocrats (with longer time horizons) appear to be less interested in satisfying the performance requirements of external actors or reaping external benefits from donors and investors. Conversely, this cohort of senior civil servants was also more likely to attribute influence to assessments that either promoted reforms complementing existing efforts, or helped the host government clearly identify practical approaches to addressing critical policy problems. Less politically oriented government officials appear to focus more on problem solving and alignment with pre-existing domestic reform priorities.

5. External Assessment Influence: Country-Level Findings

Which *country-level* factors determine a government's responsiveness to external sources of analysis and advice? The existing literature suggests that external assessment influence could be related to various country characteristics, including: country size, regime type, the government's commitment to reform, and bargaining power vis-à-vis the sponsors and suppliers of external assessments.¹⁴⁹

Using responses to the *2014 Reform Efforts Survey*, we examined the extent to which a government's propensity to use external analysis and advice correlates with unique country-level characteristics in five domains: demographics, governance, reform history, bargaining power, and normative and causal beliefs. After analyzing the data, we identified **6 country-level findings**:

1. External assessments are more influential in smaller countries.
2. External assessments are more influential in countries with democratically elected leaders and a free press.
3. Countries with effective public sector institutions are more amenable to external analysis and advice.
4. Reform-minded governments are more likely to draw upon external sources of analysis and advice.
5. Aid-dependent countries are more influenced by external assessments.
6. The influence of external assessments may depend upon the training and experience of political leaders.

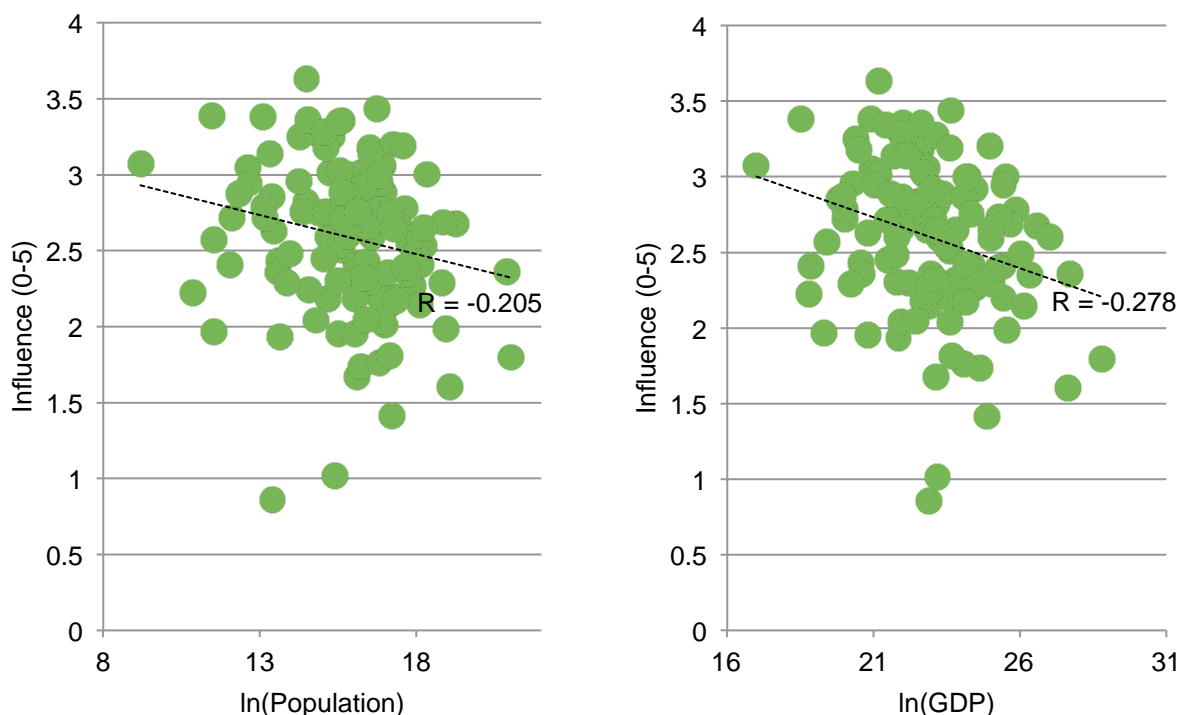
5.1. External assessments are more influential in smaller countries.

Figure 28 suggests that external assessments of government performance are, on average, more influential in small countries than in large countries.¹⁵⁰ There are several reasons why the size of a country's population might correlate negatively with the ability of external assessment suppliers to influence the decision-making of another government. One possibility is that with limited internal administrative and technical capabilities, smaller countries may rely more heavily on external diagnostics of performance (Edwards 2012). Alternatively, this pattern in the data could indicate that small states lack the credibility or bargaining power to evade external pressures for reform (Gruber 2000; Pop-Eleches 2009; and Odell and Tingley 2013).

¹⁴⁹ For an introduction to this literature, see Noland 1997; Vreeland 2003; Kelley 2004; Hille and Knill 2006; Chwieroth 2007, 2009; Sharman 2009; Pop-Eleches 2009; Montinola 2010; Pitlik et al. 2010; Girod and Walters 2012; Girod and Tobin 2011; Öhler et al. 2012; Parks 2014; Kelley and Simmons 2014a, 2014b; and IMF 2014.

¹⁵⁰ This is true regardless of whether country size is measured demographically (i.e., by population) or economically (i.e., by gross domestic product).

Figure 28. The Negative Relationship between Country Size and Assessment Influence at the Agenda-Setting Stage

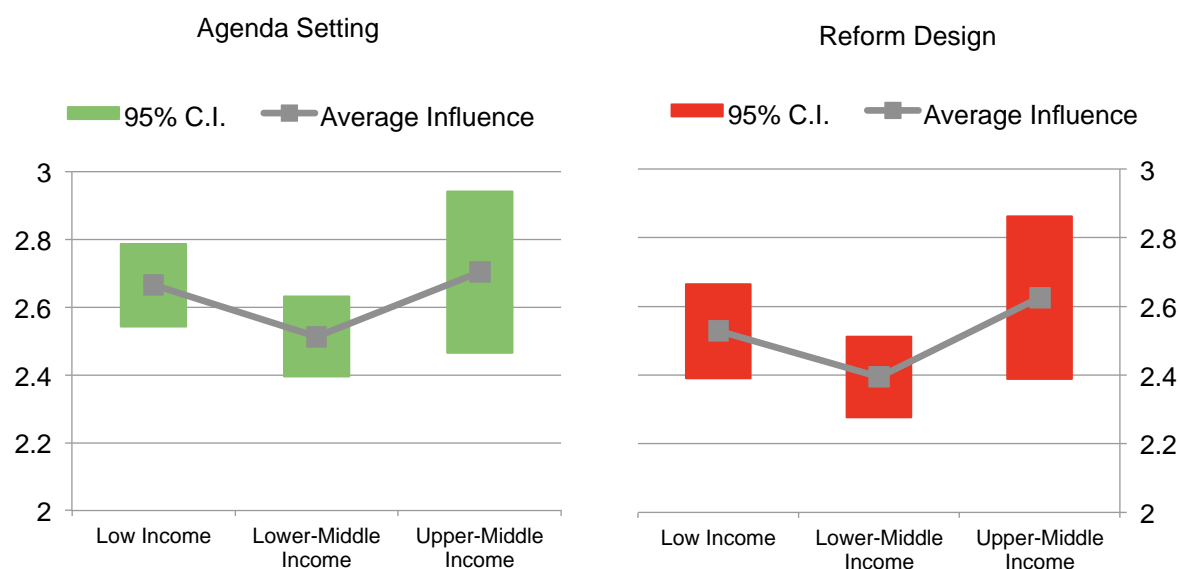


Note: Assessment influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". The natural logarithms of population and GDP better estimate the linear relationship between country size and assessment influence via Pearson's correlation coefficient. Both population and GDP represent the average available values for each sample country from 2004 to 2013. The R-values of -0.205 and -0.278 indicate negative relationships between an average country's demographic and economic size and the agenda-setting influence of external assessments in that country. On average, smaller countries experience higher levels of assessment influence, while larger countries experience lower levels of assessment influence.

There also several reasons why external assessment influence might correlate positively with a country's per capita income level, including lower demand for reform and capacity to execute reform (Andrews 2009; Loayza et al. 2012). The results that we report in Figure 29 suggest that upper-middle income countries are most influenced by external assessments, followed by low income and lower-middle income countries.¹⁵¹ Our findings are consistent with Loayza et al. (2012), who find that as the size of the middle class increases in a country, reform efforts intensify and the overall quality of governance improves. Upper-middle income countries are generally characterized by a growing middle class, which may explain why governments in these countries are more interested in using external performance diagnostics. Another possible explanation for higher levels of external assessment influence among upper-middle countries may lie in the high concentration of upper-middle income countries in and around Europe, which in turn may reflect (at least in part) the strong influence of EU-sponsored assessments.

¹⁵¹ Countries are grouped according to World Bank income group designations over the period 2004-2013. For each country, we rely on the *modal* income group designation over the ten-year period. This procedure for grouping countries into income brackets is not the same procedure we use to identify the cohort of countries for inclusion in our sample, which is why upper-middle countries are included in the analysis provided here. While we restrict the sample to countries that were in the low- and middle-income brackets around the beginning of the ten-year period, many of these countries "graduated" from one income bracket to another during this period of time.

Figure 29. Assessment Influence (0-5), by Country Income Level



Note: Assessment influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Country income levels are based on the modal classification of each sample country from 2004 to 2013, according to the World Bank. The height of each bar corresponds to the 95% confidence interval around the point estimate for the average country within each income group.

Yet, neither of these explanations easily accounts for why external assessments of government performance seem to exert less influence in lower-middle income countries than in low-income countries. This could be an example of the "seesaw effect," whereby progress in one area often leads to backsliding in others (Acemoglu et al. 2008). That is to say, if the process of reform is indeed non-linear and progress is often followed by setbacks, it is possible that when low income countries pursue a set of "first-generation reforms" the disruptive nature of policy and institutional change engenders high levels of opposition to the future reforms that they might otherwise pursue when they attain lower-middle income status.¹⁵²

5.2. External assessments are more influential in countries with democratically elected leaders and a free press.

We also wondered whether a country's regime type might increase or decrease the influence of external assessments. Liberal regimes may be more likely to consider external policy recommendations than authoritarian regimes because elected leaders have a greater incentive to strengthen policy and institutional quality, provide public goods, and "lock-in" reform commitments (Kapstein and Converse 2008 and Keefer 2007)¹⁵³. However, it is also possible that autocrats face fewer domestic restraints than democrats and can more easily implement politically unpopular reforms in pursuit of long-run benefits (Devarajan et al. 2001).

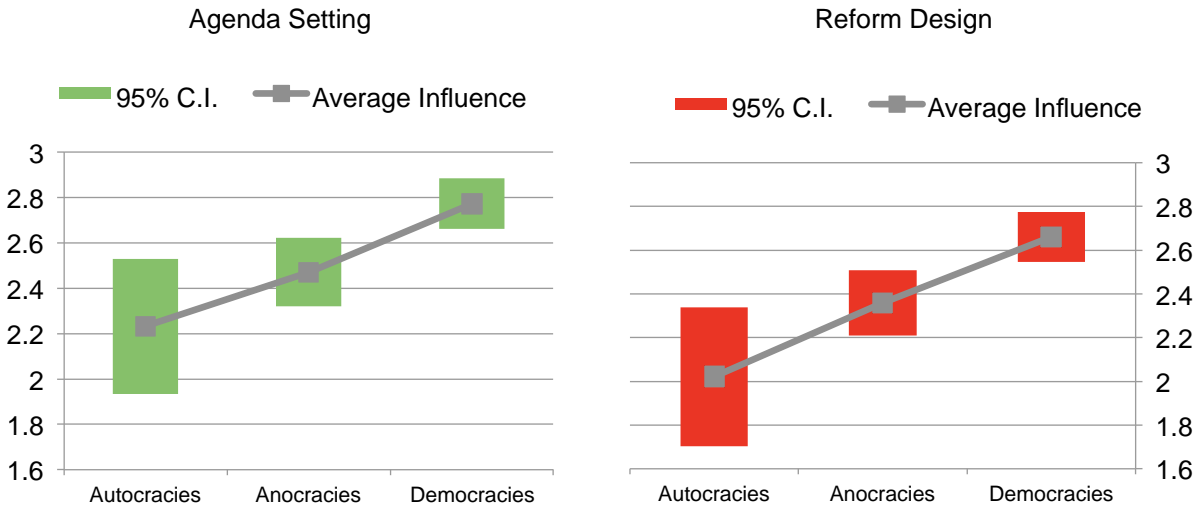
We find that external assessments are most influential in democratic regimes and least influential in autocratic regimes (see Figure 30). Anocracies—regimes that are neither fully democratic nor fully autocratic—fall somewhere in the middle. These findings support the hypothesis that the presence of

¹⁵² The absolute differences in assessment influence across income levels are not statistically significant, however, so one should approach these results with some degree of caution.

¹⁵³ Also see Joyce 2006; Smith and Vreeland 2006; Acemoglu et al. 2008; Amin and Djankov 2009; Bueno de Mesquita and Smith 2009, 2010; and Montinola 2010.

electoral incentives has a significant impact on how governments respond to external pressures for reform (Kapstein and Converse 2008; Parks 2014; and Aklin and Urpelainen 2014).

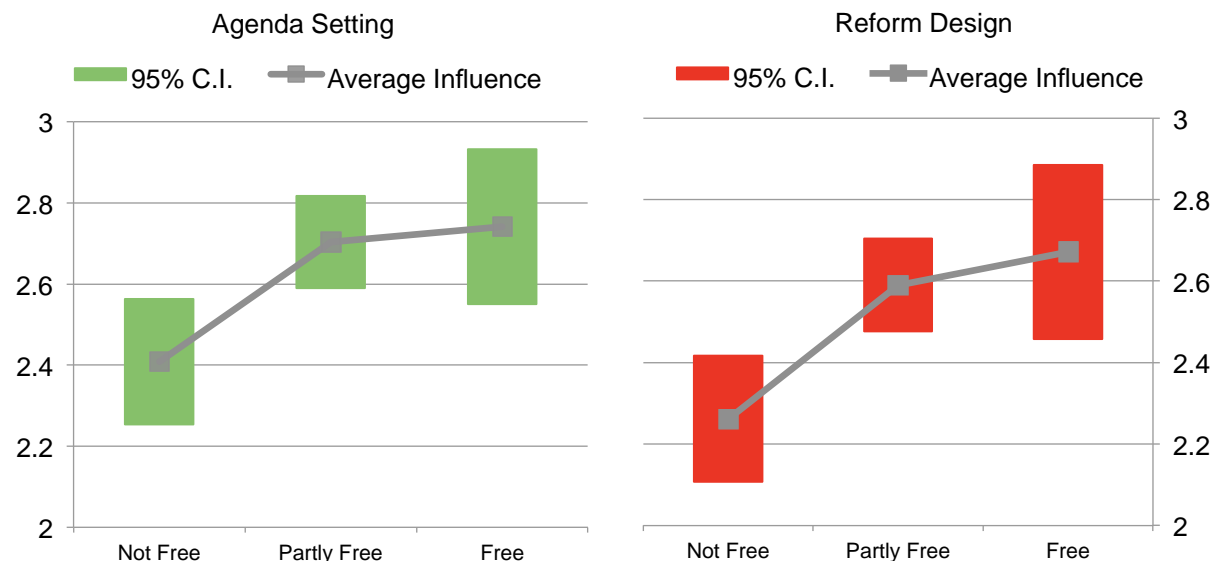
Figure 30. Assessment Influence (0-5), by Regime Type



Note: Assessment influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Regime type designations are based on the modal classification of each sample country from 2004 to 2013 on the Polity IV Index. The height of each bar corresponds to the 95% confidence interval around the point estimate for the average country within each country group.

Similarly, governments characterized by low levels of press freedom (i.e., "not free") are less likely than other governments to be influenced by external assessments (see Figure 31). While a country's regime type and its level of press freedom are closely correlated and difficult to empirically disentangle, there is good reason to believe that the inability of domestic media to pressure the government for reform may contribute to lower levels of assessment influence (Hedger and Agha 2007; Charron 2009; and Diamond 2013).

Figure 31. Assessment Influence (0-5), by Country Press Freedom Status



Note: Assessment influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". Country press freedom designations are based on the modal classification of each sample country from 2004 to 2013 on Freedom House's Freedom in the World Report. The height of each bar corresponds to the 95% confidence interval around the point estimate for the average country within each press freedom group.

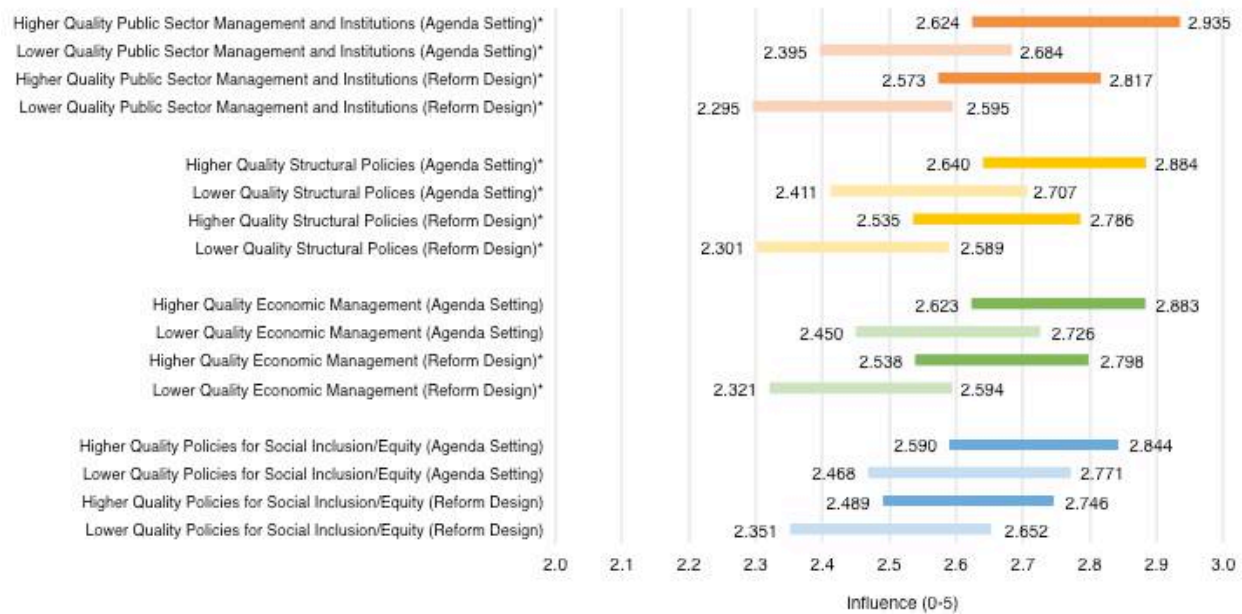
5.3. Countries with effective public sector institutions are more amenable to external analysis and advice.

Drawing upon data from WB's CPIA, we also examined the role public sector policies and institutions play in shaping external assessment influence at the agenda-setting and reform design stages of the policymaking process.¹⁵⁴ As demonstrated in Figure 32, we find that external assessments are more influential in countries with high-quality institutions and policies and less influential in countries with low-quality institutions and policies. This result suggests that well-governed countries with high quality public sector management policies and institutions are more mindful of—and responsive to—external assessments of government performance than other countries.¹⁵⁵ This could be a reflection of the fact that these countries have a larger cadre of competent civil servants and a higher level of “absorptive capacity” for the diagnostic and advisory content contained in external assessments. It could also be the case that competent governments are more willing to consider external analysis and advice in the first place.

¹⁵⁴ Countries are grouped into higher and lower score categories according to the median global scores on the CPIA Public Sector Management and Institutions Cluster Ratings, CPIA Structural Policies Cluster Ratings, CPIA Economic Management Cluster Ratings, and CPIA Policies for Social Inclusion and Equity Ratings.

¹⁵⁵ This finding is consistent with previous research undertaken by Hille and Knill (2006), Pop-Eleches (2009), and Parks (2014).

Figure 32. External Assessment Influence, by Policy and Institutional Performance at the Country Level



Note: * Indicates a significant difference-in-means at $p < 0.05$. Country policy and institutional performance categorizations are based on the modal score for each country from 2004 to 2013 on subcomponent indicators of World Bank's IDA Resource Allocation Index. The World Bank's IDA Resource Allocation Index is measured on a scale from 1 (low) to 6 (high). Influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence." The width of each bar corresponds to the 95% confidence interval surrounding the average assessment influence score for countries with the indicated level of policy and institutional performance. Assessment influence is on a scale from 0-5, where 0 means "No influence at all" and 5 means "Maximum influence." Only countries with a sample size of at least 10 observations were considered for analysis.

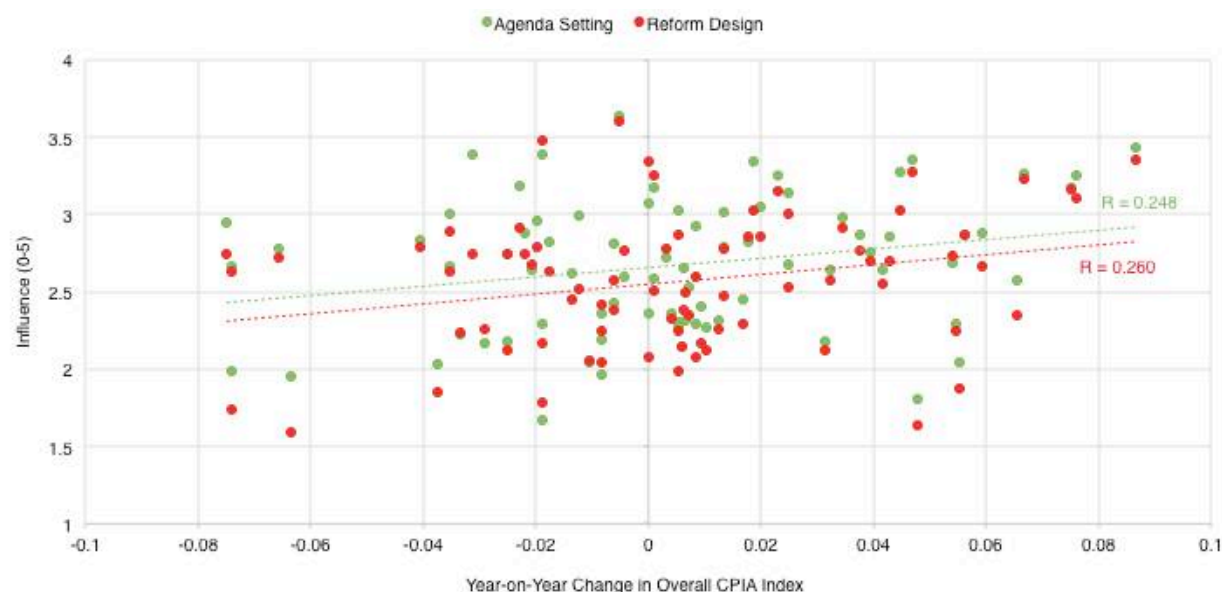
5.4. Reform-minded governments are more likely to draw upon external sources of advice and analysis.

Changes over time in the WB's CPIA scores can also be used to understand the extent and pace of reform implementation in a country and, by extension, the reform credentials of a government.¹⁵⁶ In Figure 33, we calculated a government's reform track record during the 2004-2013 period, with higher scores indicating stronger reform credentials. We find a positive correlation between a government's reform credentials and its propensity to use external assessments at both the agenda-setting and reform design stages of the policymaking process.¹⁵⁷ This positive relationship could indicate either that reform-minded governments are more likely to draw upon external analysis and advice, or that reform is more likely when external assessment influence is high.

¹⁵⁶ This is a good measure of the reform credentials of a government insofar as the successful implementation of reforms in recent history captures a government's *demonstrated* willingness and ability to undertake reform (Parks 2014). Averaging all CPIA cluster averages as a proxy for the overall quality of a government's institutions and policies, we calculated the average year-on-year change in this index score as a measure of successful reform implementation.

¹⁵⁷ Figure 32 excludes one outlier: Serbia. Whereas the average year on year change for the rest of the countries in the sample is 0.005494, Serbia's average year on year change on the CPIA index was -0.608, a number that is based solely on data from the 2005-2006 period—a time at which time Montenegro achieved its independence from Serbia and the country experienced an overhaul in its policies and institutions. Including Serbia in the sample results in a significantly weaker correlation. For an exploration of the construct validity of our CPIA-based measure of success in reform implementation efforts, see the supplemental information and Figures H.3 and H.4 in [Appendix H](#).

Figure 33. The Positive Relationship between CPIA-Based Success in Reform Implementation and Assessment Influence



Note: The "Year-on-Year Change in the Overall CPIA Index" was calculated by taking the year-on-year difference (value at time t minus value at time $t-1$) in the World Bank's IDA Resource Allocation Index for each year 2004-2013 and then finding the average value of this difference. The World Bank's IDA Resource Allocation Index is measured on a scale from 1 (low) to 6 (high). Influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". The R values of 0.248 and 0.260 indicate positive relationships between an average country's reform credentials and the agenda-setting and reform design influence of external assessments in that country

5.5. Aid-dependent countries are more influenced by external assessments.

Another potential determinant of whether governments draw upon external analysis and advice in the policymaking process is their level of bargaining power vis-à-vis external assessment providers. We considered two different measures of bargaining power: net official development assistance as a percentage of gross national income (a proxy for government reliance on revenue from external assessment sponsors)¹⁵⁸ and natural resource rent revenue as a percentage of national income (a proxy for government access to non-aid revenue and insulation from aid-for-policy bargaining).¹⁵⁹

Unsurprisingly, we find that a government's level of aid dependence is positively correlated with its responsiveness to external assessments of government performance (see Figure 34).¹⁶⁰ Access to "outside revenue options" cuts in the opposite direction: governments with significant alternative sources of revenue (i.e., highly fungible natural resource rents) are significantly less likely to rely upon external sources of analysis and advice.¹⁶¹

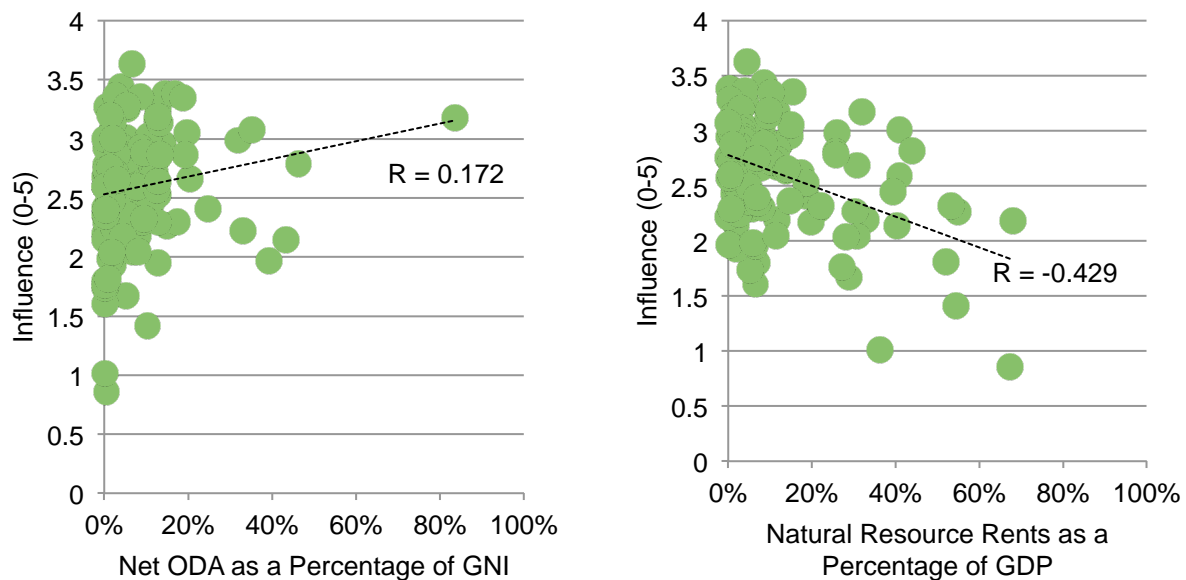
¹⁵⁸ This indicator is imperfect in that nearly all external assessment sponsors and suppliers are also donors. Nevertheless, it does broadly capture the degree to which a government can independently formulate and execute policy without having to take into account the interests of external actors.

¹⁵⁹ See Hamilton and Clemens 1999. This variable sums all rents resulting from 14 fuel and nonfuel mineral resources: oil, gas, hard coal, gold, silver, iron, lead, nickel, phosphate, tin, zinc, lignite, bauxite, and copper.

¹⁶⁰ This relationship is statistically insignificant at $p < 0.05$, but significant at $p < 0.10$. These results are consistent with those reported in de Renzio 2009; Girod and Tobin 2011; Andrews 2011; Parks 2014; Blum 2014; and Buch et al. 2014.

¹⁶¹ This relationship is statistically significant at $p < 0.05$. These results are consistent with those reported in Girod and Tobin 2011 and Parks 2014.

Figure 34. The Importance of Bargaining Power on Agenda-Setting Influence of External Assessments



Note: Influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". For the purposes of this report, AidData employs the standard definition of ODA provided by the OECD-DAC. Natural Resource Rents refer to the sum of rents from 14 fuel and nonfuel mineral resources: oil, gas, hard coal, gold, silver, iron, lead, nickel, phosphate, tin, zinc, lignite, bauxite, and copper. The R-value of 0.172 indicates a positive relationship between an average country's level of aid dependence and the agenda-setting influence of external assessments in that country. The R-value of -0.429, on the other hand, indicates a negative relationship between a country's natural resource rents and the agenda-setting influence of external assessments.

5.6. The influence of external assessments may depend upon the training and work experience of political leaders.

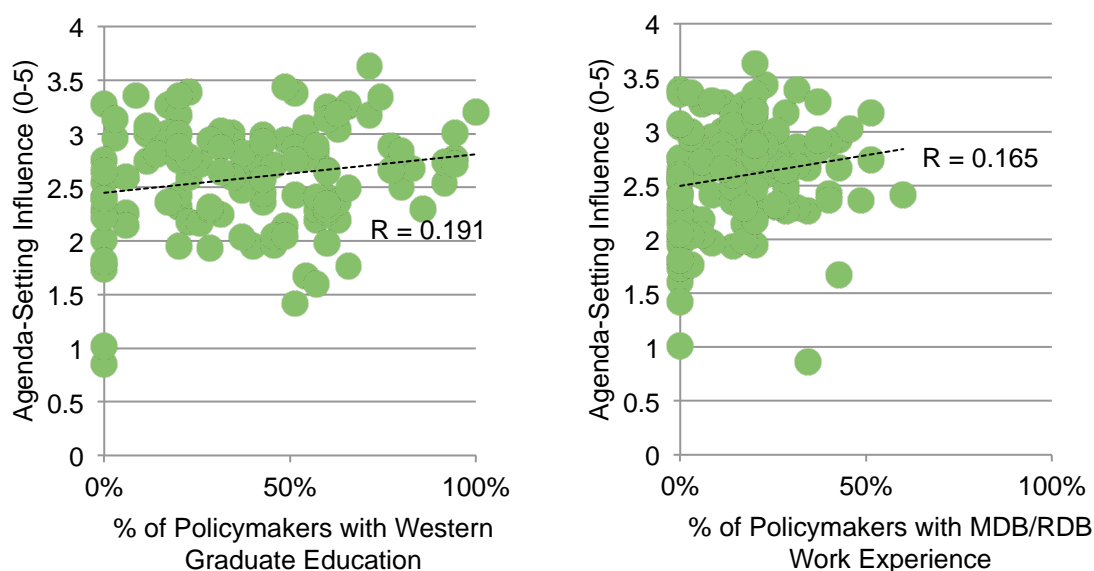
There is a growing literature about the impact that policymakers' normative and causal beliefs may have on their policy priorities and choices. To that end, we examine whether the education and professional socialization of governing senior government decision-makers in developing countries affects the extent to which their countries value and use external sources of analysis and advice.

Recent theoretical work suggests that the policy influence of external actors may hinge on their ability to find and work with developing country officials who share a common set of causal and principled beliefs about development policy (Chwioroth 2007, 2009a; Weymouth and Macpherson 2012; and Parks 2014). In other words, if developing country policymakers share the same professional and educational "DNA" as their foreign counterparts, they may be more amenable to external policy ideas.

In Figure 35, we examine the relationship between external assessment influence at the agenda-setting stage of the policymaking process and the educational training and professional socialization of government decision-makers in developing countries. For the purpose of this analysis, we use two proxy measures for educational training and professional socialization: (1) the percentage of a government's senior policymaking team who received their graduate degrees from OECD countries, and (2) the

percentage of a government's senior policymaking team who have previously worked for bilateral development banks and international financial institutions.¹⁶²

Figure 35. The Positive Relationship between Senior Policymaking Team Socialization and Assessment Influence at the Agenda-Setting Stage



Note: Agenda-setting influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence". We define the senior policymaking team as: Head of State, Minister of Finance, Minister of Planning, Minister of Foreign Affairs, and Central Bank Governor. As "MDB/RDB Work Experience", we counted any prior work experience with the IMF, the World Bank, the AsDB, the IADB, the AfDB, or the EBRD. The R-values of 0.191 and 0.165 indicate a positive, though weak, relationship between the Western socialization of an average country's senior policymaking team and the agenda-setting influence of external assessments in that country.

While we do find a positive correlation between educational training and assessment influence, as well as professional socialization and assessment influence, in both cases the relationship is only weakly significant.¹⁶³

¹⁶² For professional socialization, we counted as "socialization" any prior work experience with the IMF, the WB, the AsDB, the IADB, the AfDB, or the EBRD. These data are drawn from Parks (2014). We define the senior policymaking team as: Head of State, Minister of Finance, Minister of Planning, Minister of Foreign Affairs, and Central Bank Governor.

¹⁶³ These data are drawn from Parks (2014). Whereas the correlation with the measure of education training is significant at $p < 0.05$, the correlation with previous work experience at multilateral development banks and international financial institutions is only significant at $p < 0.10$.

6. Findings: Intended and Unintended Effects

Questions about the upstream policymaking influence of external assessments naturally lead to questions about their downstream effects. Some argue that external assessments have largely positive effects: shoring up support for reform-oriented technocrats, broadening domestic coalitions in favor of reform, weakening domestic political opponents, and facilitating policy coordination and collaboration with the donor community.¹⁶⁴ However, others argue that external assessments have a number of negative, unintended consequences: diverting a government's attention away from higher priority policy issues, limiting domestic policy autonomy, or creating incentives to "game the system".¹⁶⁵

The *2014 Reform Efforts Survey* speaks to this debate. Not only did we ask participants to identify the factors they believe best *explain* the influence of individual external assessments, we also asked them to identify the overall *effects* of those same assessments on the policymaking process. Participants were asked to evaluate the extent to which they believed that each assessment had a series of effects on a scale of 0 to 5, with 5 indicating very strongly and 0 indicating not at all.

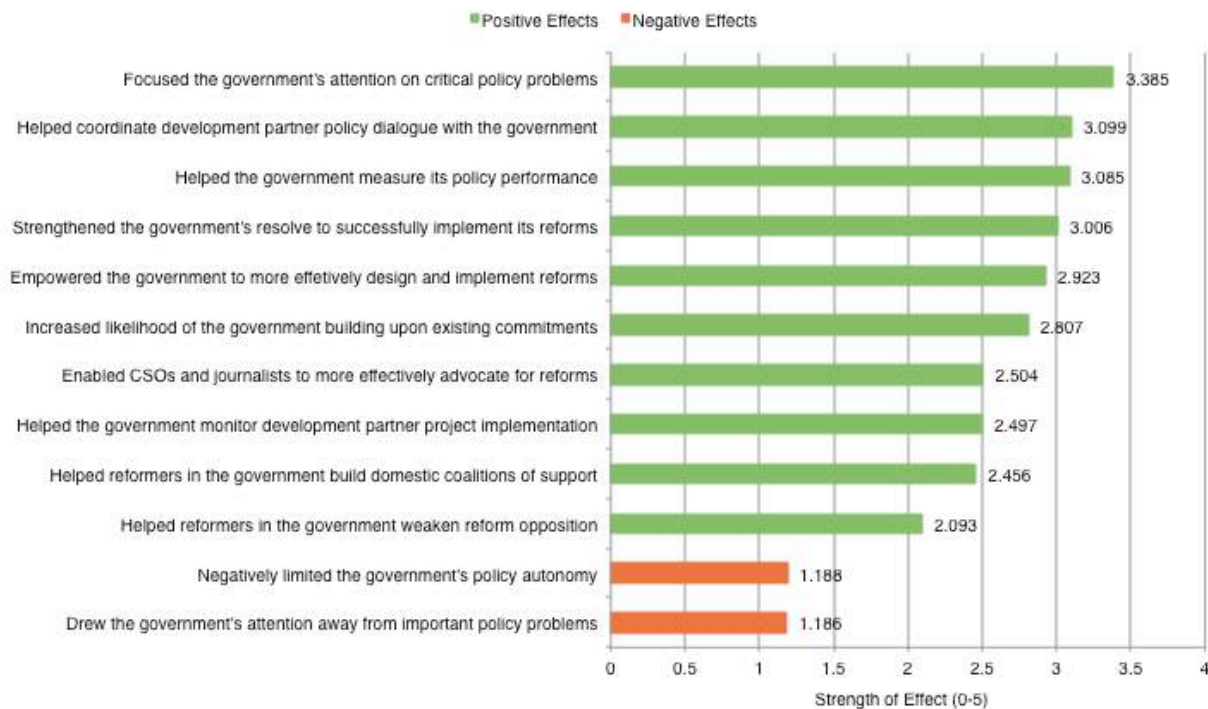
Analyzing the data, we identify twelve findings regarding the effects of external assessments:

1. External sources of analysis and advice help focus the attention of governments on critical policy problems.
2. External assessments help governments monitor their own performance and strengthen reform resolve.
3. The authorities do not appear to be unduly constrained or distracted by external performance assessments.
4. Assessments rarely help to neutralize opposition or build coalitions in support of reform.
5. Governmental and non-governmental stakeholders broadly agree on the effects that assessments have.
6. Assessments help establish a shared diagnosis of policy problems and bridge donor-recipient priorities.
7. In countries where external assessment influence is high, both positive and negative effects are more likely.
8. External assessments have more influence when they are focused on the "right" problems.
9. Countries can successfully implement reforms by going it alone or by working with outsiders.
10. External assessment influence and domestic political support are good predictors of reform success.
11. Assessment suppliers should be mindful that reforms succeed in some policy domains more than others.
12. Some influential assessments correspond closely with reform success, while others do not.

¹⁶⁴ See Kelley 2004; Radelet 2007; Sutherland 2008; Djankov 2009; Kelley and Simmons 2014a; and Ribadu 2010.

¹⁶⁵ See Lim et al. 2008; Chhotray and Hulme 2009; Soederberg 2004; Arruñada 2007; Pham 2009; Goldsmith 2011; Sjöstedt 2013; IBP and ONE 2013; and Hoey 2015.

Figure 36. The Strength of External Assessment Effects



Note: "Strength of Effect" is on a scale of 0-5, where 0 means "No effect at all" and 5 means "A very strong effect."

6.1. Assessments have largely positive and rarely negative effects on the policy process.

Participants reported that external assessments of government performance had mostly positive effects in the low- and lower-middle income countries included in our sample.

6.1.1. External analysis and advice help focus the attention of governments on critical policy problems.

External assessments of government performance, first and foremost, appear to have helped focus the attention of "target" governments on critical policy problems. Of a list of potential assessment effects, this one received the highest weighted average score of 3.385 out of a possible 5.

6.1.2. External assessments help governments monitor their own performance and strengthen reform resolve.

The survey results also provide support for the idea that external assessments helped donors and their host government counterparts to more effectively coordinate their policy dialogue (3.099),¹⁶⁶ measure

¹⁶⁶ On the issue of whether and when donors effectively coordinate their policy dialogue with the host government, see Bourguignon and Platteau forthcoming; Pop-Eleches 2009; and Knack and Rahman 2007.

their own performance (3.085), and shore up internal resolve to successfully implement reforms (3.006).¹⁶⁷

6.1.3. Authorities do not appear to be unduly constrained or distracted by external performance assessments.

Most participants did *not* agree that external assessments drew the attention of developing country governments away from important policy problems (1.186 out of 5) or limited the policy autonomy of governments in a negative manner (1.188 out of 5). These results are consistent with the findings of Meessen et al. (2011) and the findings of the 2012 MCA Stakeholder Survey, reported in Parks and Rice (2013).¹⁶⁸

6.1.4. Assessments rarely help to neutralize opposition or build coalitions in support of reform.

Several other potential effects of assessment influence also registered weighted average scores below the mid-point scale value of 2.5. Two of these relatively weak assessment effects, in particular, merit special attention: helping reformers weaken opposition to reform and helping reformers build coalitions of support for reform. Contrary to claims made by international political economy scholars, our survey evidence suggests that external assessments of government performance tended to have a limited impact on the most basic elements of political economy in developing countries (i.e., who opposes and supports reform). By contrast, external assessments of government performance were able to help solve more technical issues like problem identification, policy coordination, and performance measurement.

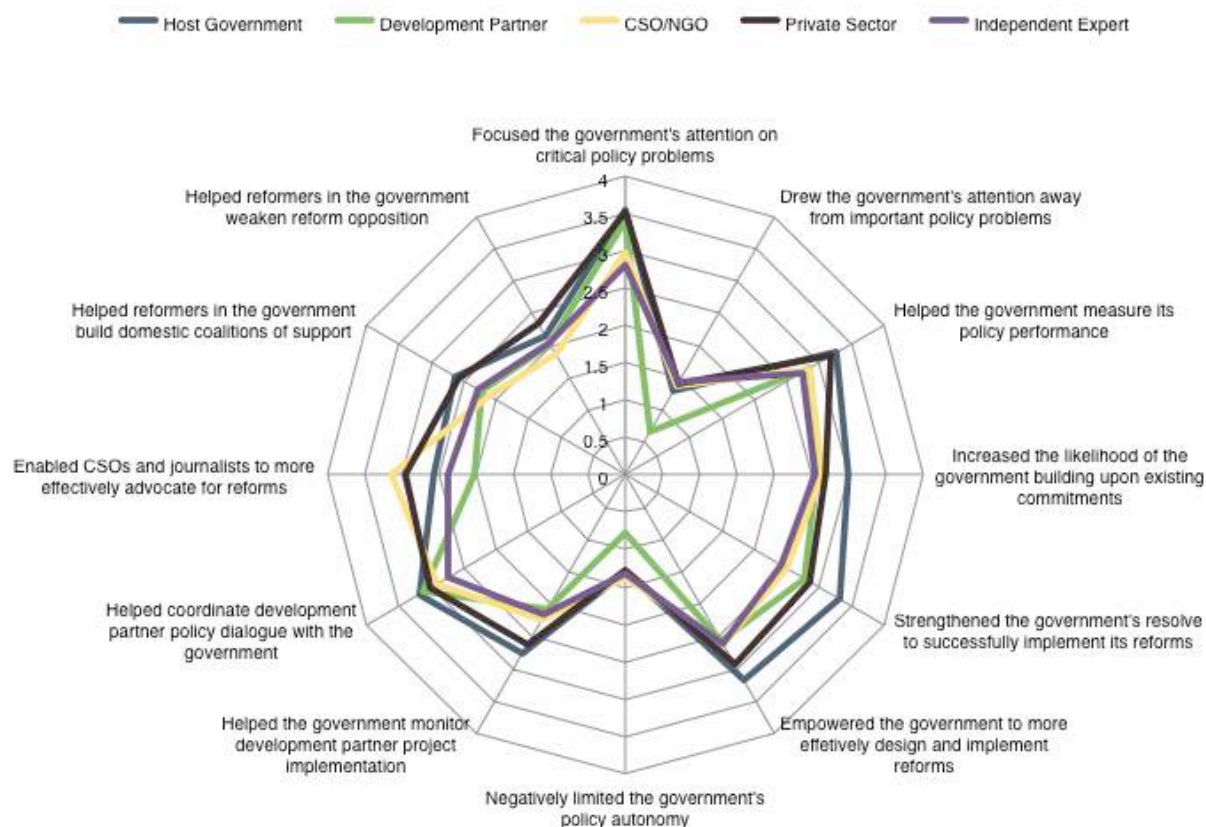
6.2. Governmental and non-governmental stakeholders broadly agree on the effects that assessments have.

Did the host government officials targeted by external assessments have different perceptions of assessment effects than the development partners who produced them? Did the perceived impact of external assessments vary depending on whether were are a policy insider or outsider (e.g., civil society, private sector, and independent experts)? In Figure 37, we examine perceptions of the effects of external assessments on government performance by stakeholder group.

¹⁶⁷ On the issue of whether and when external assessments of government performance are used by the authorities to track their own progress, see de Renzio and Masud 2011; Schueth 2011; Sharman 2011; de Sherbinin et al. 2013; Parks and Rice 2013; McArthur 2013; and Perakis and Savedoff 2015.

¹⁶⁸ This may be because assessments are largely influential only when they align with country priorities (i.e., countries “screen out” otherwise limiting or distracting assessments).

Figure 37. The Strength of External Assessments Effect, by Stakeholder Group



Note: "Strength of Effect" is on a scale from 0-5, where 0 means "No effect at all" and 5 means "A very strong effect." Assessment effects are in no particular order. The spider graph displays every stakeholder group's perception of the strength of each effect. Each effect has its own dedicated radial line emanating from the center of the graph. More strongly felt effects are located further from the center of the graph on each line.

There is substantial agreement among the different stakeholder groups about the effects that external assessments of government performance did (or did not) have. Participants across the stakeholder groups largely agreed that external assessments helped developing country governments measure their own performance and focus their reforms on critical policy problems. There was also broad agreement that external assessments did not negatively limit the policy autonomy of governments or distract governments from confronting important policy problems.

Nevertheless, there were some differences in perception across stakeholder groups. Host government participants were most likely to indicate that external assessments had positive effects on performance measurement, reform resolve, and development partner coordination with the authorities. While assessments may not have had as much influence on the design of reforms as they did on shaping the policy agenda, host government officials still perceived them to be a tool they could use to maximize the likelihood of success in reform design and implementation.¹⁶⁹

¹⁶⁹ The largest margins between host government scores and those of other stakeholder groups pertained to assessment effects on reform resolve and policy and programmatic implementation. Host government scores were substantially higher for the following response options: "strengthened the government's resolve to successfully implement its own reform efforts" (+0.449), "increased the likelihood that the government would build upon previously adopted policy commitments" (+0.306), "empowered the government to

Disaggregating the survey data by stakeholder group also suggests some level of perception bias where participants gave relatively higher scores to those effects to which they were more readily exposed. For example, the assessment effect of “enabling civil society organizations and journalists to more effectively advocate for reforms” received the highest score (3.132) from CSO and NGO participants. The assessment effect of “helping development partners coordinate their policy dialogue with that of the government” received the second highest scores (3.190 and 3.136) from host government and development partner participants. Within the development partner cohort, there seems to be some bias against the idea that external assessments of government performance have had negative effects on host government policy decision-making. Development partner participants provided the lowest scores among all stakeholder groups to “drew the government’s attention away from important policy problems” (0.677) and “limited the policy autonomy of the government in a negative manner” (0.785).

Independent country experts assigned many of the lowest assessment effect scores. This pattern could be attributable to the independence of these country experts and their ability to make arms-length judgments. However, it could also reflect the fact that independent country experts possess less first-hand knowledge of how assessments actually influenced the behavior of governments and non-governmental groups.

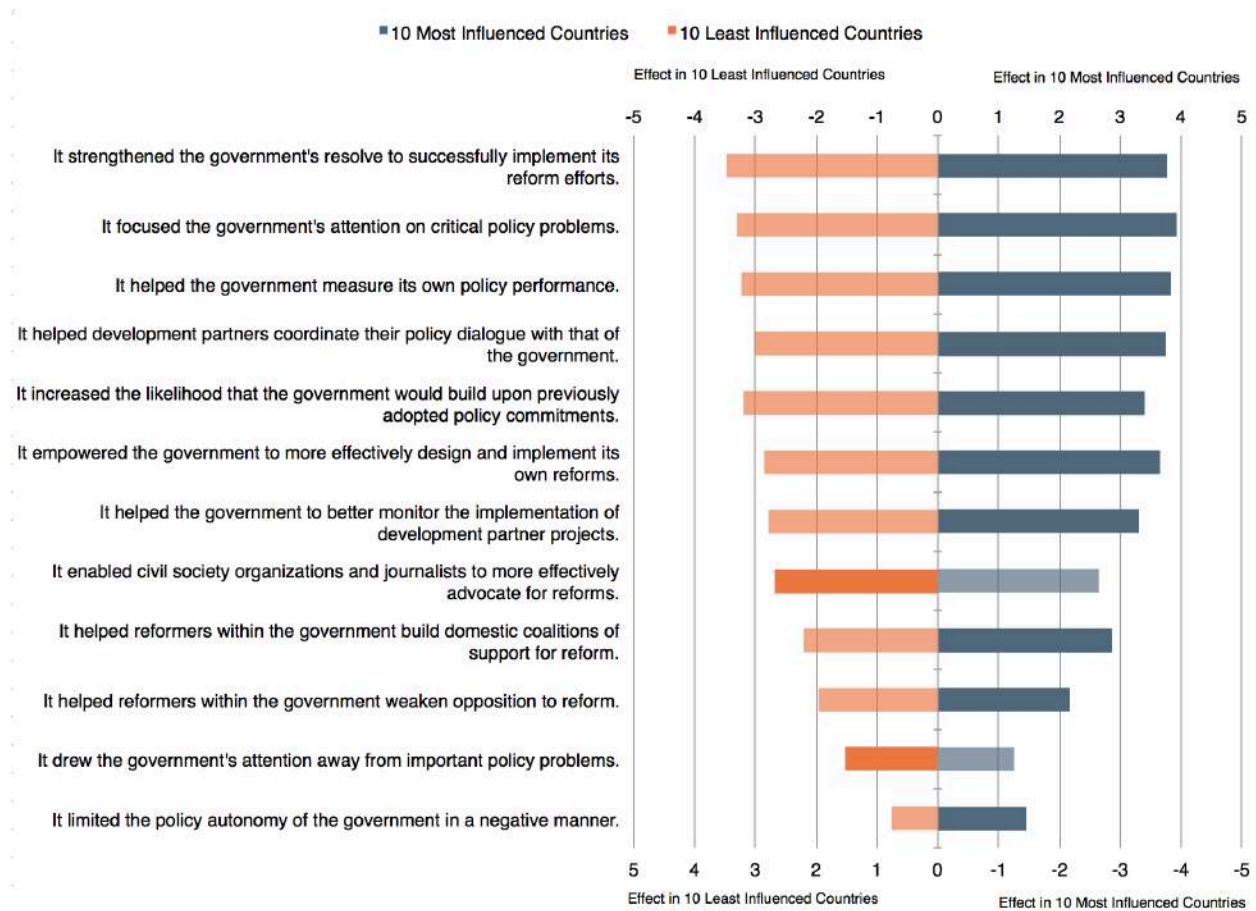
6.3. External Assessment Effects in Highly Influenced and Weakly Influenced Countries

In Figure 38, we report the weighted score for each intended and unintended assessment effect according to host government participants from the 10 countries that were most and least influenced by external assessments.¹⁷⁰

more effectively design and implement its own reforms” (+0.205), and “helped the government to better monitor the implementation of development partner projects” (+0.150).

¹⁷⁰ Figure 16 indicates the top 10 most and least influenced countries at the agenda-setting stage. Somaliland is omitted from this analysis because of its status as a semi-autonomous territory. Fiji is added in its place among the group of 10 least influenced countries. The countries Equatorial Guinea, Turkmenistan, and Cuba are also omitted, as they do not meet the participant threshold of $N \geq 5$ necessary for this analysis.

Figure 38. The Strength of External Assessment Effects in 10 Most and Least Influenced Countries



Note: "Strength of Effect" is on a scale of 0-5, where 0 means "No effect at all" and 5 means "A very strong effect." The 10 most and least influenced countries at the agenda setting stage can be found in Figure 16. Somaliland was excluded from the selection of least influenced countries because of its status as a semi-autonomous region. Fiji was included in its place.

6.3.1. Assessments help establish a shared diagnosis of policy problems and bridge donor-recipient priorities.

In both the most and least influenced countries, the three strongest effects identified by participants were that assessments “strengthened the government’s resolve to successfully implement its reform efforts,” “focused the government’s attention on critical policy problems”, and “helped the government measure its own policy performance.” These results are encouraging because they suggest that—in both highly influenced and weakly influenced countries—**external assessments helped establish a shared diagnosis of policy problems and bridge the gap between the policy priorities of host governments and those of development partners.**¹⁷¹

Participants from both the most and least influenced countries also identified the same three weakest effects of external assessments. For both sets of countries, the weakest overall effects were that assessments “helped reformers within the government weaken opposition to reform,” “drew the government’s attention away from important policy problems,” and “limited the policy autonomy of the government in a negative manner.”

6.3.2. In countries where external assessment influence is high, positive and negative effects are more likely.

Nonetheless, Figure 38 also calls attention to important differences in how the most and least influenced countries experienced external assessments of government performance. **While in general external assessments do not appear to negatively limit the policy autonomy of host governments, assessments are more likely to exert undue influence in highly influenced countries.**¹⁷²

Assessment sponsors and suppliers should be mindful of this problem in countries where their assessments achieve high levels of resonance and uptake.¹⁷³

On the flip side, where external assessments achieved high levels of policy influence, they also tended to bring together domestic and external actors with a shared interest in reform and exert mostly positive effects on government reform efforts. In particular, participants from highly influenced countries more readily indicated that assessments: “empowered the government to more effectively design and implement its own reforms,” “helped reformers within the government build domestic coalitions of support for reform,” and “helped development partners coordinate their policy dialogue with that of the government.”

6.4. Influence To What End: Do External Assessments Focus Reforms on the Right Problems?

A common criticism of external assessments of government performance is that they bring to bear a set of outside policy priorities that may not align well with the policy priorities of the domestic authorities (Soederberg 2004; Arruñada 2007; and IBP and ONE 2013).¹⁷⁴ This disconnect may unintentionally

¹⁷¹ The results reported in Mackie and Caprio (2011), IMF and WB (2009), and IEG (2008) echo this finding. For example, in a broad review of WB-sponsored public sector reform programs between 1999 and 2006, independent evaluators concluded that, “starting with [analytical and advisory activities] has been a successful way for the [World] Bank to develop a trusting relationship with governments to work on sensitive areas of [public sector reform]” (IEG 2008: xvii).

¹⁷² It is also possible that, in the least influenced countries (which elsewhere in this report we show to be closed and autocratic), external assessments tend either not to limit policy autonomy at all or to only limit policy autonomy in a manner that is not seen as negative.

¹⁷³ The fact that participants from the “least influenced” countries were more likely than participants from the “most influenced” countries to say that external assessments “drew the government’s attention away from important policy problems” underscores the importance of producing analytical and advisory content that resonates with the authorities.

¹⁷⁴ Pritchett et al. (2014: 65, emphasis added) speak to the use of PEFA indicators in Mozambique: “PEFA indicators reflecting de jure changes are above average while PEFA dimensions reflecting de facto implementation and functional adjustment are only

incentivize government decision-makers in developing countries to adopt cosmetic, rather than deep and durable reforms in order to reap the financial and reputational rewards provided by external actors (Andrews 2013; Pritchett et al. 2013). Others suggest that a poor fit between domestic and external priorities may encourage “gaming” and other forms of opportunistic behavior (Global Integrity 2008; Pham 2009; and Goldsmith 2011).¹⁷⁵

If external assessments of government performance have substantially *more* influence on reform efforts where the host government *is* focused on solving the right set of problems, one could interpret this as a situation in which assessment sponsors/suppliers are “pushing on an open door” and have better odds of influencing the policy decisions of the governments they have assessed. On the other hand, if external assessments of government performance have *the same or more* influence on reform efforts where the host government *is not* focused on solving the right set of problems, one could interpret this as evidence that the assessments in question are not fit for purpose, that their sponsors/suppliers are swimming against a strong tide, or even that they are driving governments to engage in isomorphic mimicry and ignore key policy problems.¹⁷⁶

To examine this question further, we split our sample according to whether or not participants reported that the reforms pursued by the host government addressed the most critical problems in a particular policy domain. We then compared the agenda-setting influence of individual external assessments of government performance in these two cohorts.

6.4.1. External assessments have more influence when they are focused on the “right” problems.

We find that external assessments of government performance had significantly *more* influence when reform efforts *were* focused on the right set of problems. The average assessment has an influence score of 2.737 at the agenda-setting stage when reforms *were* focused on the most critical policy problems, and an influence score of 2.062 when reforms *were not* focused on the most critical policy problems. On average, therefore, it seems that external assessments of government performance have been used to help governments focus their reform efforts on issues of critical importance. Our finding complements the evidence presented in Figures 24 and 26, which suggested that assessments are influential largely because of their ability to help governments identify and solve critical policy problems.

Nevertheless, not all external assessments of government performance focused the authorities’ attention on critical policy problems all the time. Some did a better job of this than others. Table 4 lists the ten assessments whose influence depended the most on whether reforms focused on the most critical policy problems. In other words, these assessments were more influential than others in helping governments focus their attention on and solve key policy problems.

Table 5. Assessments with Substantially More Influence Over Reforms Focused on Critical Problems

Assessment	Influence when Focused on	Influence when Not Focused on Most	Δ in Influence	Difference from Expected Δ
------------	---------------------------	------------------------------------	----------------	----------------------------

average. When asked about this, officials in line ministries, departments, and agencies note that the new laws and systems *are part of the problem*. They may look impressive, but are often poorly fitted to the needs of those using them, requiring management capacities users do not have and institutionalizing organizational scripts and allocation modalities that reflect international best practice but not political and organizational realities on the ground. These officials note that *they were never asked about the kind of system needed*, and while recognizing the impressive nature of the new [public financial management] system, they lament the missed opportunity to craft a system that *works to solve their specific needs*.”

¹⁷⁵ See Wynia et al. 2000; Jacob 2005; Hood 2006, 2012; Ward and John 2013; and Hoey 2015.

¹⁷⁶ It is possible that these assessments exert influence for reasons that are orthogonal to the policy issues they raise.

	Most Critical Problems	Critical Problems		in Influence ¹⁷⁷
EU Governance Initiative and Incentive Tranche	3.046	1.292	+1.754	+1.064
The IMF's Article IV Consultations	3.761	2.200	+1.561	+0.835
The Egmont Group of Financial Intelligence Units' Membership Requirements	2.853	1.545	+1.308	+0.627
US State Department's "Trafficking in Persons" Report	2.886	1.646	+1.240	+0.558
AsDB's Country Economic Reviews	2.681	1.466	+1.215	+0.543
Financial Action Task Force (FATF) Blacklist	2.526	1.357	+1.169	+0.504
EU Special Incentive Arrangement for Sustainable Development, Good Governance	2.687	1.562	+1.125	+0.453
AfDB's Country Governance Profiles	2.761	1.641	+1.119	+0.443
Variable Tranche of the EU Budget Support Program	2.715	1.633	+1.081	+0.407
UNESCO Education for All Development Index	3.202	2.130	+1.071	+0.373

Conversely, Table 5 lists the top ten external assessments of government performance whose influence was either roughly the same when reform efforts did or did not focus on the most critical policy problems, or were higher when reform efforts *were not* focused on the most critical policy problems. Assessments that fall in this category should give their sponsors and suppliers pause. Notable among these is the GEF's Performance Index and Resource Allocation Framework, which received a higher influence score when reforms *were not* focused on critical policy problems (2.933) than when reforms *were* focused on critical policy problems (2.567). The likely source of this poor level of performance is that this assessment emphasizes global environmental issues that benefit stakeholders in a wide variety of countries, rather than local environmental issues that benefit local constituents and stakeholders (Hicks et al. 2008; Buntaine and Parks 2013).¹⁷⁸

Table 6. Assessments with Roughly the Same Influence Whether or Not Reforms Focused on Critical Problems

Assessment	Influence when Focused on Most Critical Problems	Influence when Not Focused on Most Critical Problems	Δ in Influence	Difference from Expected Δ in Influence
Global Environment Facility's Performance Index and Resource Allocation Framework	2.567	2.933	-0.367	-1.033
EU Association Agreements	3.083	3.022	+0.061	-0.631
World Bank Development Policy Loans Program	2.829	2.761	+0.068	-0.612

¹⁷⁷ *Difference from Expected Δ in Influence* is equal to the residual value from a simple linear regression where *Expected Δ in Influence* = $-0.05 * \text{Influence when Focused on Most Critical Problems} - 0.5384$. This line-of-best fit was calculated using data from all assessments where the number of observations is ≥ 5 both when focused on the most critical problems and when not focused on the most critical problems.

¹⁷⁸ The GEF seems to be aware of this challenge. See Watson 2008 and GEF 2013.

AsDB Country Environmental Analysis	1.994	1.889	+0.105	-0.533
World Bank Rural Access Index	2.539	2.364	+0.175	-0.490
EU Stabilization and Association Agreements	3.753	3.562	+0.191	-0.535
World Bank Health Sector Review	3.292	3.056	+0.237	-0.466
ILO Global Monitoring and Analysis of Conditions of Work and Employment	2.875	2.605	+0.270	-0.412
African Growth and Opportunity Act (AGOA) Eligibility Criteria	2.365	2.088	+0.277	-0.380
European Neighborhood Policy Action Plans and Country Reports	2.822	2.535	+0.287	-0.393

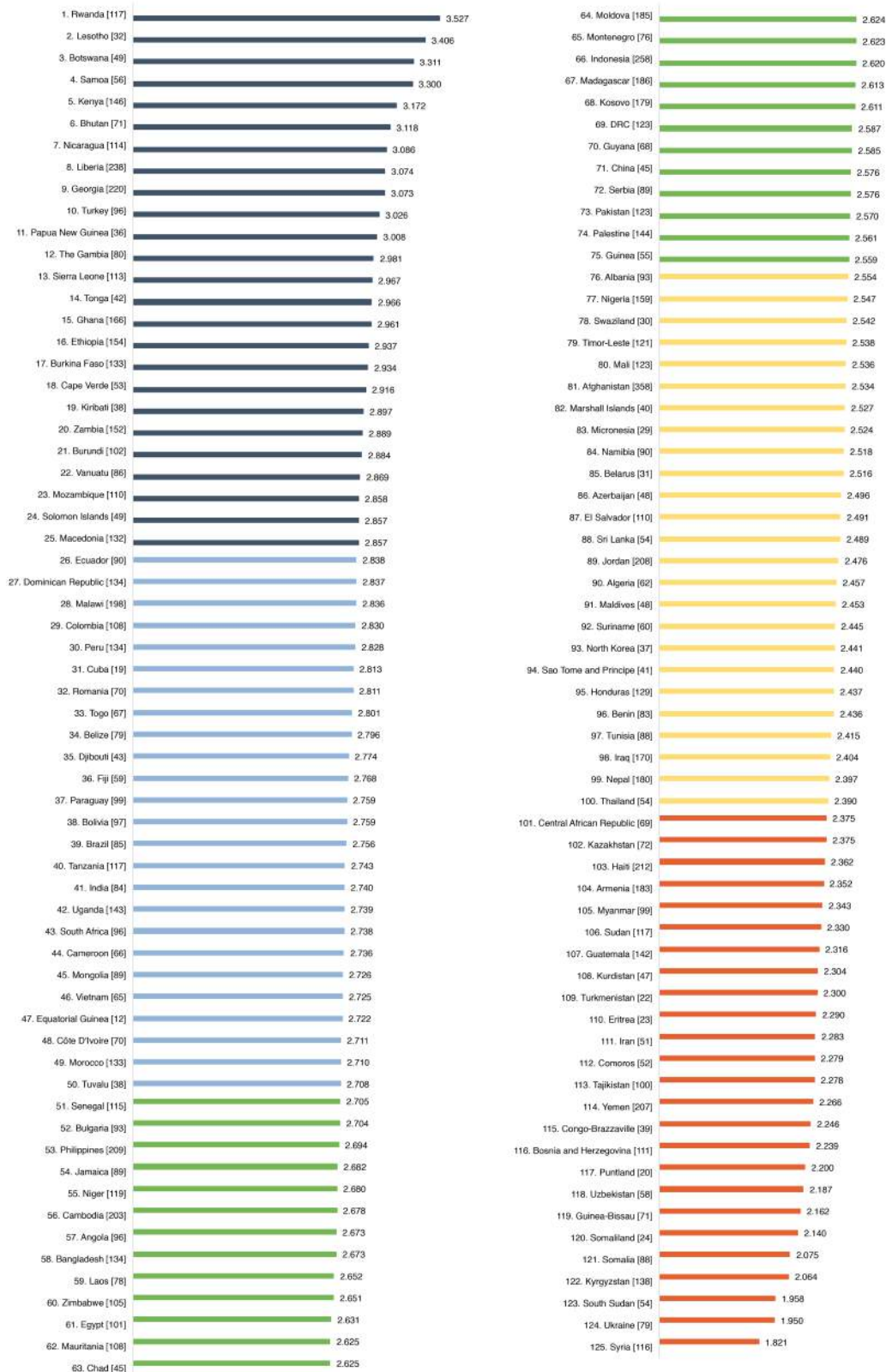
6.5. The Success of Reforms Influenced (or Not) by External Assessments

Another contested question is whether and when external assessments of government performance facilitate the successful implementation of reform efforts in developing countries.¹⁷⁹ Our survey participants identified three problems that reforms tried to solve—bounded by their country, policy domain, and the years in which they held a specific position between 2004 and 2013. Each participant was then asked to indicate on a scale of 1 to 4 the amount of progress that each reform made towards solving those three specific problems—1 indicating no progress at all and 4 indicating a great deal of progress.¹⁸⁰ Figure 39 presents these data on reform success by country.

¹⁷⁹ See Chhotray and Hulme 2009; Goldsmith 2011; Booth 2011; Brown and Tirnauer 2009; Andrews 2010, 2013; IBP and ONE 2013; Global Integrity 2008a, 2008b; and Parks 2014.

¹⁸⁰ Throughout this report, we refer to this amount of progress as reform success, or the success of reform implementation. 1 = “No progress at all”, 2 = “Only a little progress”, 3 = “A moderate amount of progress”, and 4 = “A great deal of progress”.

Figure 39. Success in Reform Implementation, by Country



Note: Participants in each country were asked to evaluate the success of reform efforts, measured as the progress made by each of up to three reform efforts towards solving particular problems, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress". The estimates presented in this chart are presented in rank order by quintile, and the location of a country within a specific quintile indicates the relative progress made by that country towards solving problems vis-a-vis other countries. The number of reform-specific observations pertaining to each country is in brackets. Zanzibar was excluded from analysis because it did not meet a response sample size threshold of at least 10 observations. Due to respondent security concerns, the North Korean assessment influence score only reflects the experiences and perceptions of those development partner staff and officials and independent country experts who were physically located outside of North Korea on the survey launch date.

6.5.1. Countries can achieve reforms whether by going it alone or working with outsiders.

Interestingly, many of the countries most influenced by external assessments of government performance are *not* the same countries that were most successful in implementing reform efforts. However, it is also true that we observe a moderately positive and statistically significant correlation between external assessment influence and reform success.¹⁸¹ We interpret this as evidence that, while assessment influence has a net positive effect on reform success, there are different paths to successful reform implementation. Some reformers, such as Rwanda and Ethiopia, adopt a "go-it-alone" approach and limit the extent to which external pressures are brought to bear on the domestic policy formulation and execution process (see Box 9). Other reformers, such as Liberia and Georgia, rely more heavily and proactively on information, advice, assistance, and incentives from external actors (see Box 10).

¹⁸¹ See Figure H.2 in [Appendix](#).

Box 9. Rwanda and Ethiopia: A “Go it Alone” Approach to Reform

Rwanda and Ethiopia are paradigmatic examples of countries with successful track records of reform implementation, but that remain fiercely independent and wary of external efforts to influence their policy priorities and behavior.¹⁸²

Ethiopia ranks 16th out of 125 countries on our measure of reform success (see Figure 39), yet only 73rd on our measure of the agenda-setting influence of external assessments (see Figure 16). A 2012 *Christian Science Monitor* profile of the late Ethiopian Prime Minister Meles Zenawi speaks directly to this issue of how the political leadership in Addis Ababa has approached external actors and pressures: “Meles positioned the country expertly, allowing it to develop according to his unique [policy] prescriptions, while being supported by diverse allies” (Davison 2012).¹⁸³

Rwanda has pursued a similar reform path. The authorities in Kigali have successfully overseen the implementation of a wide range of economic and social sector reforms. These policy and institutional changes have delivered real results. Between 2001 and 2011, the country experienced average annual GDP growth of over 8 percent and reduced poverty by 14 percentage points (Meessen et al. 2011; Levy 2014; Routley 2014; World Bank 2009, 2013). The WB identified Rwanda as the world’s leading economic reformer in 2009 and second-fastest reforming country in the world in 2013. Rwanda’s performance on the WB and IFC’s “Ease of Doing Business” index has increased meteorically: from 158th place in 2005 to 67th place in 2009 (WB 2009). By 2013, Rwanda had risen to 32nd place on this index (IFC 2013).

Stephen Kinzer, a former *New York Times* bureau chief, points out that “many development specialists [now] consider Rwanda the most promising poor country on earth.” Kinzer (2009) writes that Rwanda “has become not only stable but, according to its many admirers, a model for developing countries.” Time Magazine’s 2009 issue of “The World’s Most Influential People” identifies President Paul Kagame as “the face of emerging African leadership. His reconciliation strategy, management model, empowerment of women in leadership and insistence on self-reliance are transforming a failed state into one with a bright future” (Warren 2009).

Nevertheless, Rwanda jealously guards its policy autonomy (Human Rights Watch 2009; Whitfield 2009; and Swedlund 2013). This pattern is evident in our survey data. Whereas Rwanda tops our global ranking (Figure 39) as the country with single highest level of achievement in terms of successfully designing and implementing reforms during the 2004-2013 period, its relative ranking on a measure of external assessment influence (see Figure 16) is 37 places lower (38th out of 125 countries).

¹⁸² These two countries are not only strong reformers according to our metric reported in Figure 39, but also by the reckoning of many others (Soeters et al. 2006; Hartwig et al. 2008; Hedger and de Renzio 2010; Meessen et al. 2011; WB 2009; Levy 2014; and Routley 2014).

¹⁸³ On Ethiopia’s efforts to preserve its policy autonomy and maneuverability, see Whitfield 2009; Birdsall and Perakis 2012; Del Biondo and Orbie 2014; and Hackensch 2013.

Box 10. Liberia and Georgia: A “Go Together” Approach to Reform

In contrast to Ethiopia and Rwanda, Liberia and Georgia have pursued a more open and collaborative path. Both countries have aggressively pursued and successfully implemented reform efforts, but have done so with substantially more external input, influence, and assistance.

Liberia ranks 15th on our measures of external assessment influence (see Figure 16) and 8th on our measure of reform success (see Figure 39). Shortly after Ellen Johnson-Sirleaf was elected as President in December 2005, her administration sprang into action, assessing whether and how Liberia could benefit from various programs sponsored by foreign governments and intergovernmental organizations. These programs—the HIPC Initiative, the MCC Account, the EITI, and the AGOA—involved the same fundamental arrangement: domestic reforms in exchange for external rewards (Robinson 2008a, 2008b; Cook 2010; and Flores et al. 2013). Additionally, they all involved submission to explicit or implicit forms of performance assessment (Parks 2014).

President Johnson-Sirleaf and her cabinet were candid and unapologetic about their interest in mobilizing these external pressures and incentives in the service of their domestic reform objectives (Parks 2014).¹⁸⁴ The Liberian authorities subsequently succeeded in executing a series of wide-ranging anti-corruption and investment climate reforms: requiring senior government appointees to publicly disclose their income and assets, removing thousands of “ghost workers” from the government’s payroll, putting in place strict controls on discretionary government spending, simplifying business registration and land registration, and lowering barriers to trade (IMF 2006c; EIU 2007; and Konneh and Radelet 2008).

Georgia bears a striking resemblance to Liberia. Georgia ranks 12th on our measures of external assessment influence (see Figure 16) and 9th on our measure of reform success during the 2004-2013 period (see Figure 39). Mikheil Saakashvili, the country’s Western-educated President from 2004 to 2013, filled his administration with other like-minded technocrats educated in the West and introduced sweeping economic, anti-corruption, and civil service reforms.¹⁸⁵ Georgian authorities were also eager to marshal external incentives for reform and external sources of advice and assistance in support of their domestic reform priorities.

Kakha Bendukidze, State Minister of Reform Coordination, was particularly aggressive in his use of the WB and IFC *Doing Business* Report to identify, prioritize, and monitor government reform efforts (WB 2007, 2008). Schueth (2011) reports that Bendukidze and his colleagues viewed the annual *Doing Business* rankings as an opportunity to improve the country’s international reputation (Schueth 2011). The European Stability Initiative (2010) notes that “[r]ising in these rankings became a national policy priority, closely monitored by [President] Saakashvili.” By almost most accounts, the government’s strategy was a success: Georgia catapulted from 137th place in 2004 to 14th place in 2013 on the “Ease of Doing

¹⁸⁴ For example, the President reported to Liberia media, “the country would do all it could to meet the indicators in order to benefit from the Millennium Challenge Account” (Smith 2007). Several months later, the Deputy Chief of Mission at the US Embassy in Monrovia sent a cable to Foggy Bottom, describing a raft of reform measures that the Government of Liberia (GOL) had introduced: “The GOL has announced 21 reforms to improve Liberia’s ranking in the IFC ‘Doing Business’ index by enhancing the business climate. The rapid progress reflects President Sirleaf’s commitment and the productive collaboration between the GOL and the business community to target regulatory changes that do not require legislative action. There was special emphasis on reforms that will also affect Millennium Challenge Account ratings” (Robinson 2008).

¹⁸⁵ After a popular uprising in November 2003 known as the Rose Revolution, the Saakashvili administration launched a series of reforms to overhaul tax and customs administration, business and property registration, the courts, and the civil service. They fired 15,000 members of the notoriously corrupt police force, dramatically increased the salaries of 10,000 public servants to combat petty corruption, reduced the time and cost of property registration by 75%, cut the corporate income tax rate from 20% to 15%, streamlined customs by reducing the number of days it takes to export from 54 to 10, and slashed the number of activities requiring a business license from 909 to 159 (Bagaudinova et al. 2007). These reforms played a major role in reversing the country’s economic fortunes: between 2003 and 2007, Georgia registered average annual economic growth of 10%, FDI inflows tripled, and tax revenues jumped by 300% (Mitchell 2009; IMF 2009). Increased revenue enabled the government to dedicate significantly more resources to health, education, and vulnerable social groups (AsDB 2010).

Business Index" and surged ahead of Canada, Switzerland, Austria, and the Netherlands in the global rankings.¹⁸⁶

6.5.2. External assessment influence and domestic political support are good predictors of reform success.

The cases of Rwanda, Ethiopia, Georgia, and Liberia (described in Boxes 9 and 10) not only call attention to the widely varying levels of external orientation that are observed among fast-reforming countries, but also the need for a more complete explanation of the domestic and external factors that shape countries' odds of success in implementing reforms. To this end, we developed a multivariate regression model to account for country-level variation in reform success with ten explanatory variables: government effectiveness, regime type, population size, aid dependence, access to natural resource rents, checks and balances in the political system, the net level of reform support/opposition, the Western educational training of members of the government's senior leadership team, experience working for the multilateral development banks and international financial institutions within the government's senior leadership team, and external assessment influence (at the agenda-setting stage of the policymaking process).¹⁸⁷

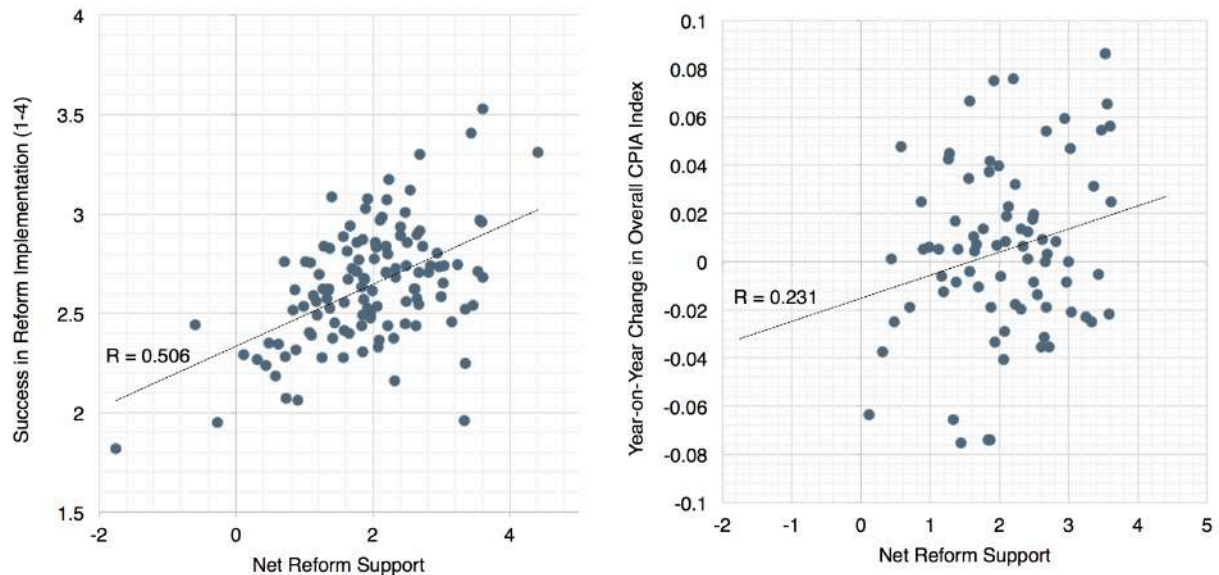
We found evidence that only four of these variables registered statistically significant and positive effects on the government's probability of successfully implementing reforms. The first key finding is that when external assessment influence is high, governments are more likely to experience success in reform implementation. We also found that aid-dependent countries and countries with effective public sector institutions are more likely to succeed when they attempt to implement reforms.

Another key determinant of success is the relative level of support for—and opposition to—reform within a given country (see Figure 40). This finding is consistent with a large body of literature on the domestic political economy of reform, which suggests that reforms often fail when two conditions are met: (1) the costs of reform are largely felt by a few individuals who are prepared to organize themselves against change, and (2) those who stand to benefit face significant collective action problems and demonstrate low levels of political organization and support (Alesina and Drazen 1991; Fernandez and Rodrik 1991; Dewatripont and Roland 1992; Lewis 1996; Acemoglu and Robinson 2000; Rajan 2004; and Kikeri et al. 2006).

¹⁸⁶ The USG, the WB, and other donors enthusiastically embraced the Georgian government and identified it as a "model of reform" (Dinmore 2005) and "[something] we hope to see replicated in many other parts of the world" (Taylor 2004).

¹⁸⁷ The multivariate regression results are available upon request.

Figure 40. The Positive Relationship Between Net Reform Support and Success in Reform Implementation



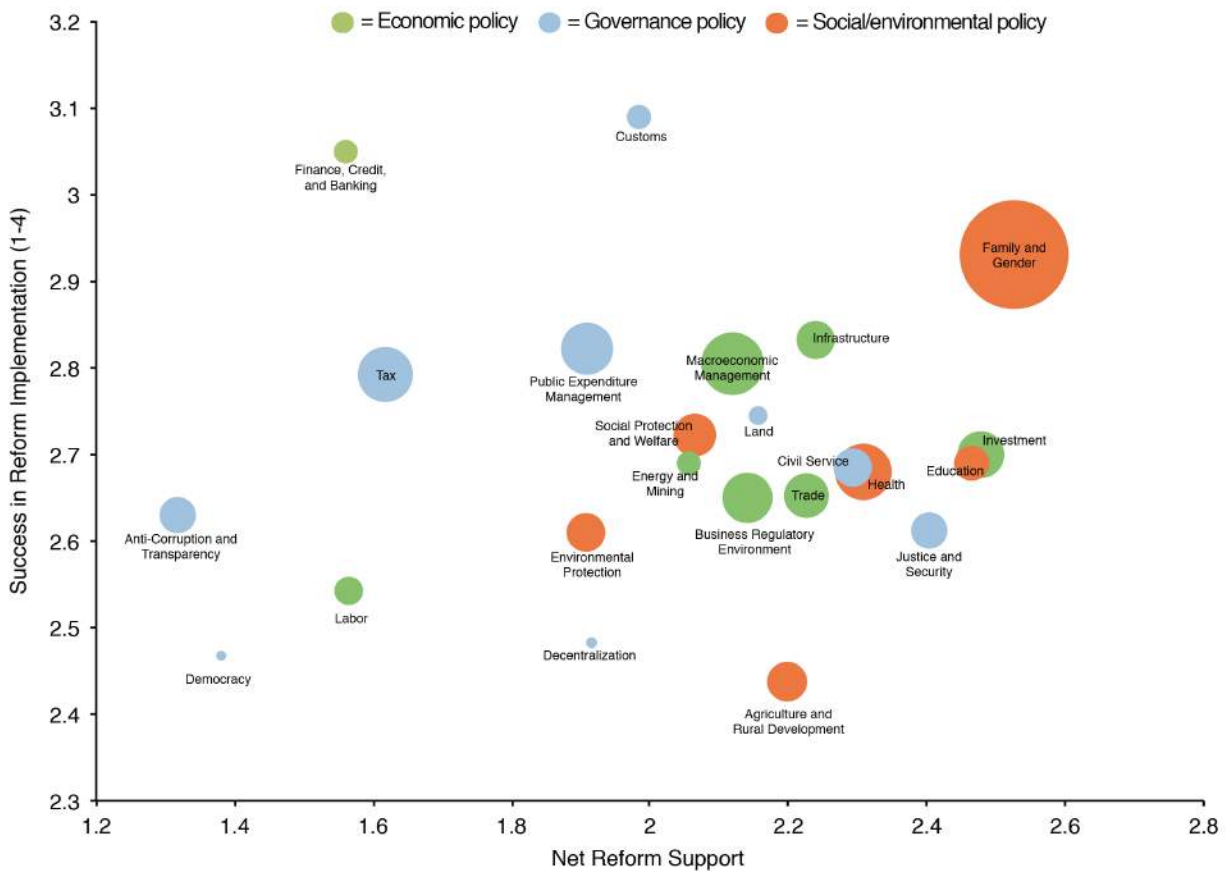
Note: "Net Reform Support" was estimated by subtracting the number of oppositional groups selected in question 38 (up to 13, including "Other") from the number of supportive groups selected in question 37 (up to 13, including "Other") and calculating the weighted average of this difference for each sample country. Because of the method used, "Net Reform Support" is more reflective of the relative breadth of support for reform in a country than it is of the relative intensity of support. The survey-based "Success in Reform Implementation" estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress". The year-on-year change in the overall CPIA index was calculated by taking the year-on-year difference (value at time t minus value at time $t-1$) in the WB's IDA Resource Allocation Index for each year 2004-2013 for a country and then finding the average value of this difference. The WB's IDA Resource Allocation Index is measured on a scale from 1 (low) to 6 (high). The R-values of 0.506 and 0.231 indicate a positive linear relationship between an average country's net breadth of reform support and the success of reform efforts in that country.

6.5.3. Assessment suppliers should be mindful that reforms succeed in some policy domains more than others.

At the country-policy domain level, we find a positive and statistically significant relationship between the relative level of reform support and opposition and the likelihood that reforms were successfully implemented (see Figure 41).¹⁸⁸ However, several "outlier" policy domains had higher or lower than expected levels of reform success given their levels of relative reform support and opposition: customs; finance, credit and banking; decentralization; agriculture and rural development; democracy; labor; and family and gender.

¹⁸⁸ At the country-policy domain level (2,898 observations), we find that $R=0.215$ and $p<0.00001$. Figure 41 represents, by policy domain, the average country's level of net reform support/opposition, reform success, and assessment influence at the reform design stage of the policy process. Each point in the chart represents the policy domain-specific experience of an average sample country, and some policy domain estimates are more precise than others. For example, the customs policy domain has the smallest, country-level standard deviation of 0.982, indicating relatively low variation in influence between countries, while the public expenditure management policy domains has a country-level standard deviation of 1.642, indicating a relatively high variation in influence between countries. Given each point's role as presenting the policy domain-specific values of an average country, it is more accurate to explicitly account for the inter-country variation in policy domain-specific scores.

Figure 41. Net Reform Support, Success in Reform Implementation, and External Assessment Influence



Note: Bubble size corresponds to the relative reform design influence of external assessments within each policy domain. "Net Reform Support" was estimated by subtracting the number of oppositional groups selected in question 38 (up to 13, including "Other") from the number of supportive groups selected in question 37 (up to 13, including "Other") and calculating the weighted average of this difference for each sample country. Because of the method used, "Net Reform Support" is more reflective of the relative breadth of support for reform in a country than it is of the relative intensity of support. The survey-based "Success in Reform Implementation" estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress".

In four policy domains—democracy; labor; decentralization; and agriculture and rural development—the probability of success in reform implementation was lower-than-expected, given levels of support for and opposition to reform. These findings corroborate previously reported results—specifically, that reform efforts in these policy domains engender relatively high levels of resistance and that such reforms are less likely to succeed in implementation. The sizes of the circles in Figure 42 reveal that these are also the policy domains where external actors have relatively low levels of policy influence, which may reflect some level of “pass through” where the insolubility of a domestic problem limits the influence that any external actor can exert.

For three policy domains—finance, credit, and banking; customs; and family and gender—the probability of success in reform implementation was higher-than-expected, given the levels of support for and opposition to reform. Specifically, family and gender seems to represent a policy domain where there were not only few reform opponents vis-à-vis reform champions, but also a significant number of successfully implemented reforms influenced by external analysis and advice. This finding is timely and merits further examination, as a growing number of influential organizations (e.g., the Bill and Melinda

Gates Foundation, the Clinton Foundation, USAID, the WB) are redoubling their efforts to encourage and incentivize gender reforms.

Conversely, while anti-corruption and transparency is the policy domain with both the lowest level of (net) support for reform and the worst track record of reform implementation, external assessments played a non-trivial role in shaping the anti-corruption policy priorities and decisions of government leaders. This finding can be interpreted in several different ways, but it is consistent with previous research about the difficulty of implementing anti-corruption reforms that actually combat and prevent corruption, as opposed to merely sending signals that satisfy external “demandeurs” of better governance.¹⁸⁹

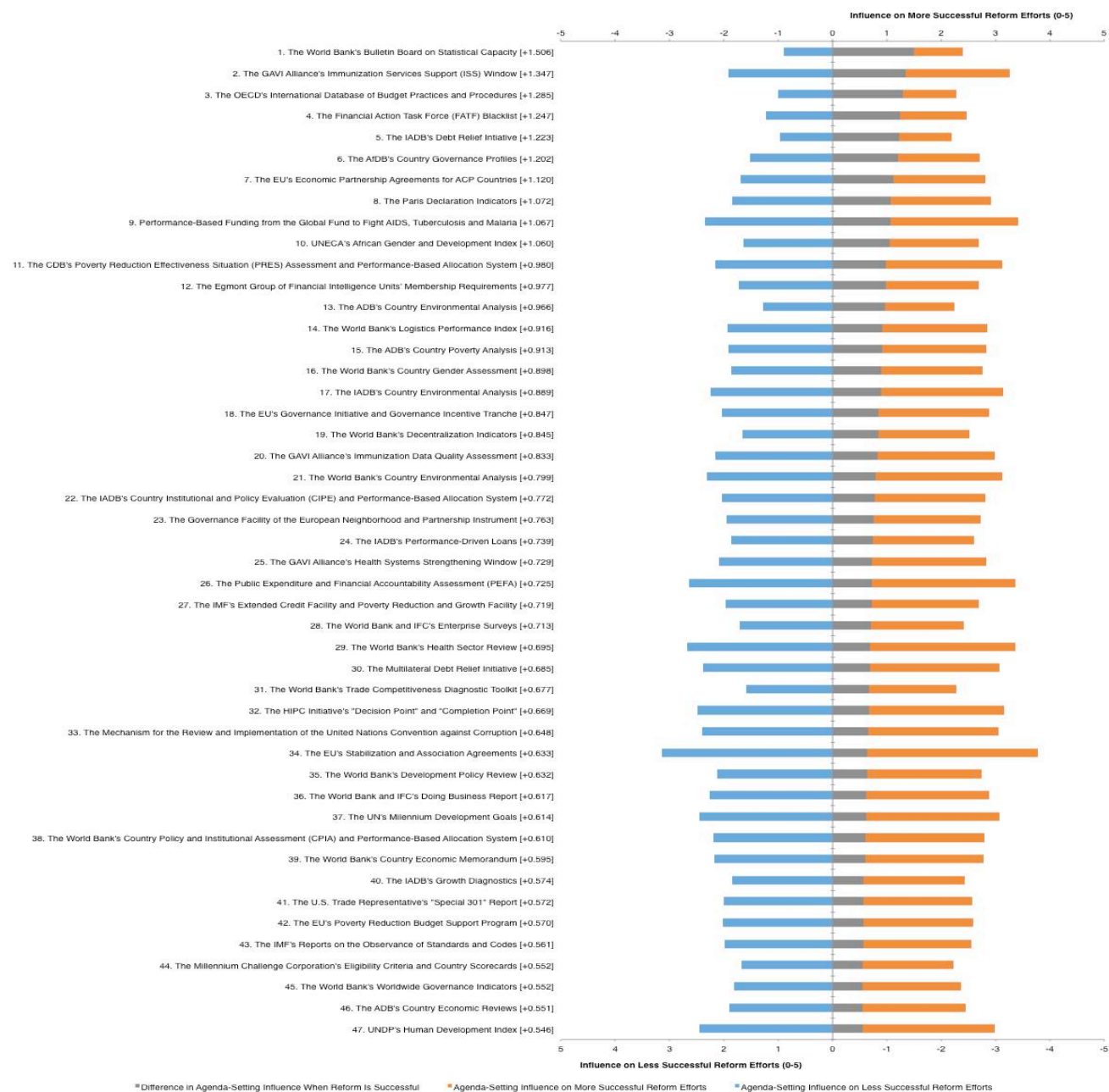
6.5.3. Some influential assessments correspond closely with reform success, while others do not.

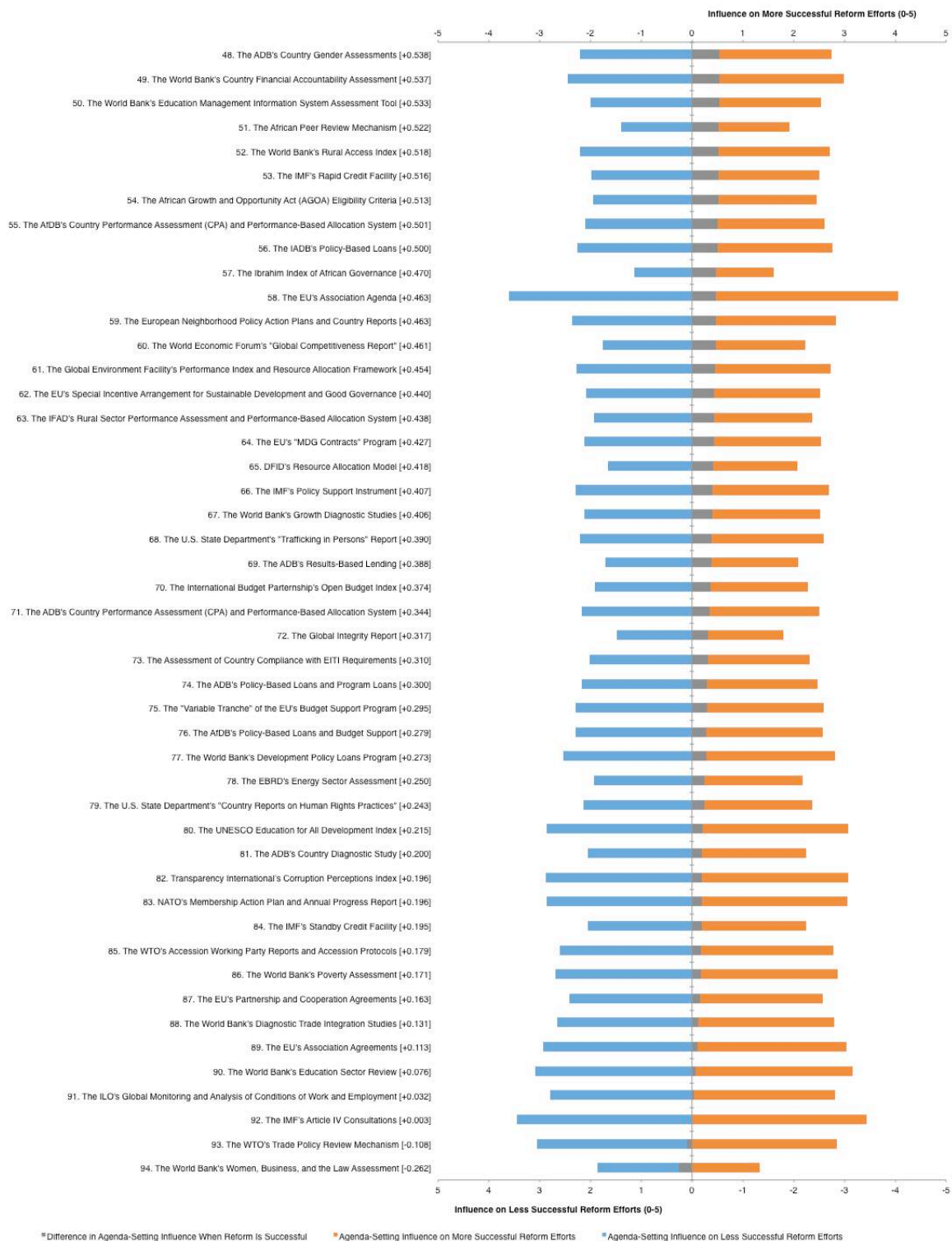
In order to gauge whether the influence of certain external assessments of government performance is more or less strongly associated with successful reform efforts, we also asked our survey participants to estimate the amount of progress each of up to three reforms made towards solving specific policy problems. Then he or she was asked to identify the extent to which individual external assessments of government performance influenced the authorities’ decision to prioritize the reforms focused on those problems. Using the responses provided to the first question, we were able to divide reforms into two groups: less successful reforms with a progress score of no more than two (at most “only a little progress”) and more successful reforms with a progress score exceeding two (at least “a moderate amount of progress”). We then used responses to the second question to estimate the degree to which each external assessment of government performance influenced both more and less successful reform efforts.

We find that all but two external assessments of government performance had more influence on successful reform efforts than unsuccessful reform efforts, which is encouraging. However, the degree to which external assessment influence corresponded with reform success does vary by assessment. Figure 42 presents the degree to which each assessment influenced more and less successful reform efforts.

¹⁸⁹ In a 2006 study, the WB’s Independent Evaluation Group concluded that, “countries have met...requirements on paper, and even agreed to the ‘cosmetic’ changes that the [World] Bank requires, while the underlying incentives and opportunities for corruption remain unaffected” (WB 2006b: 1). On this topic, also see The Economist 2000, 2003; Heilbrunn 2004; Meagher 2005; WB 2006b; Doig and Norris 2012; Innes 2013; Andrews 2013; and Persson et al. 2013.

Figure 42. External Assessment Influence on More and Less Successful Reform Efforts





Note: The change in influence when reform is successful (indicated in brackets) is equal to an assessment's agenda setting influence score when reform efforts are more successful minus its influence score when reform efforts are less successful. Reforms were categorized into more and less successful groups using a threshold score of greater than two for inclusion as a more successful reform effort. Reform success estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress". Assessment influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence".

The WB's Bulletin Board on Statistical Capacity tops the list of external assessments of government performance that tend to have exerted influence in cases where reform efforts were successful. This finding implies that this assessment of government performance encouraged governments to pursue reform efforts that had a high likelihood of success. Given that the current policy dialogue surrounding the Post-2015 SDGs has zeroed in on the critical importance of building strong and sustainable national statistical systems, our findings should provide comfort—and inspiration—to those who advocate for a “Data Revolution” (UN 2013; CGD and APHRC 2014).

Figure 42 also reveals that two performance-based aid schemes in the health sector—the GAVI Alliance's Immunization Services Support Window and performance-based funding of the Global Fund to Fight HIV/AIDS, Tuberculosis, and Malaria—demonstrated a strong correlation with more successful reform efforts. This finding supports the argument that, in the health sector, such pay-for-performance programs endow governments with the policy autonomy and maneuverability that they need to experiment, iteratively adapt, and “crawl the design space” in pursuit of deep and durable reform (Pritchett et al. 2012; also see Meessen et al. 2011; and Perakis and Savedoff 2015).

By contrast, the fact that the WTO's Trade Policy Review Mechanism *increased* in policy influence in cases when reform efforts were less successful suggests that this particular WTO assessment has been steering governments down a path that will likely end in disappointment. This could either be because the Trade Policy Review Mechanism promoted reforms that were not focused on soluble problems, or because it proposed reforms that failed to gain traction or generate the necessary level of support with the host country authorities.

Other assessments that fall in the unenviable category of having exerted roughly equal influence over successful and unsuccessful reform efforts include the IMF's Article IV Consultations, the EU's Association Agreements, the WB's Women, Business and the Law Report, and TI's Corruption Perceptions Index. These findings support previous research suggesting that some of these assessments can motivate countries to pursue cosmetic or otherwise ill-conceived reform efforts.¹⁹⁰

In our analysis of the assessment characteristics that are correlated with more or less successful reform efforts,¹⁹¹ we found that four attributes are associated with significant increases or decreases in reform success.¹⁹² These four characteristics are reported in Figure 43.¹⁹³ Some results immediately emerge from the data.

External assessments of government performance that source their underlying data from the host government were generally associated with more successful reforms than those that do not. While many

¹⁹⁰ See UNDP and Global Integrity 2008a, 2008b; Andersson and Heywood 2009; Stahl 2011; Cobham 2013; Sherr 2013; and Stewart 2013.

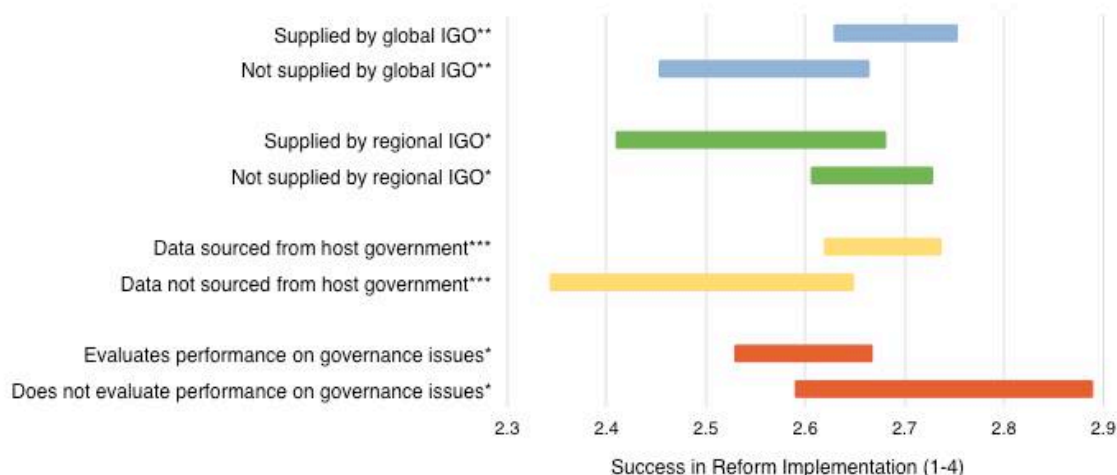
¹⁹¹ Independently of the data collected through the 2014 Reform Efforts Survey, our research team assembled a database on the observable attributes of external assessments (e.g. the identity of the assessment sponsor, where the assessment sources its data, and whether the assessment is country-specific or cross-country in nature). See [Appendix G](#) for the methodology of how this database was built. By coding our sample of assessments based on their observed characteristics, we were able to identify the extent to which assessment attributes were associated with more or less successful reform efforts between 2004 and 2013.

¹⁹² Each participant was asked to identify three key problems that reform efforts in his or her country tried to solve within a particular period of time. They were then asked about the amount of progress the same reforms made towards solving the problems they targeted, on a scale of one to four, with four indicating a great deal of progress. Finally, participants were asked to estimate the amount of influence—on a 0-5 scale, where 5 represents maximum influence—that individual assessments had on both the government's decision to undertake reforms focused on the three identified problems and the design of the resulting reforms. We were then able to isolate the set of reforms on which each individual external assessment of government performance had the most influence of any assessment and measure the amount of progress that was made, on average, on those particular reform efforts.

¹⁹³ The bars depicted in Figure 43 show the 95% confidence intervals around the mean reform success scores associated with each assessment characteristic. Difference-in-means tests were conducted for all variables listed in Appendix G, except for the following: *PolicyDomainCount*, *CoverageThreshold_5*, *PolicyDomainScope*, and policy-domain specific variables. The assessment characteristics listed in Figure 43 correspond to the following codebook variables and values: Supplied by global IGO = *NatureofSupplier* value 4, Supplied by regional IGO = *NatureofSupplier* value 5, Data sourced from host government = *PartyResponsible_2*, and Evaluates performance on broad governance issues = *GovernancePolicyBroad*.

scholars argue that delegating data collection and policy surveillance responsibilities to outside parties with significant expertise and credibility leads to greater assessment influence and thus reform success (Hawkins et al. 2006; Momani 2007; and Hagen 2009), our empirical findings show the opposite to be true: limiting measurement responsibilities to parties outside of the target government seems to result in less successful reform efforts.¹⁹⁴ One way of interpreting this finding is that assessments are more likely to yield positive reform results when they rely on host government data sources *that are already being used by the authorities to make policy decisions*.

Figure 43. Assessment Characteristics Associated with Significantly More or Less Success in Reform Implementation



Note: * Indicates a significant difference-in-means at $p < 0.10$, ** indicates a significant difference-in-means at $p < 0.05$, and *** indicates a significant difference-in-means at $p < 0.01$. The width of each bar corresponds to the 95% confidence interval surrounding the average assessment reform success score for assessments with the indicated characteristic. The survey-based "success in reform implementation" estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress". Only assessments with a sample size of at least 10 observations were considered for analysis.

Second, assessments from international organizations with a global reach (e.g., WB, UN, and IMF) were associated with significantly more successful reform efforts than those supplied by regionally focused international organizations (e.g., AfDB, AsDB, and IADB). This finding suggests that the policy expertise that resides within the largest international organizations is uniquely valuable to reform-oriented policymakers in the developing world. It may also reflect the WB's efforts to position itself as a "knowledge bank" over the last twenty years (Gilbert et al. 1999; Phillips 2009; Ravallion 2011; IEG 2012; and Doemeland and Trevino 2014) and the fact that "providing policy advice to ... member countries is an integral part of ... the [IMF's] mandate" (IMF 2013: v).

External assessments that focus on issues of governance appear to be associated with less successful reform efforts. This is a sobering finding and deserves further exploration. For now, we simply note that at least two explanations are plausible: (1) successfully implementing governance reform is fundamentally more difficult because there are relatively few champions and many opponents, and (2) governance assessments may promote ill-conceived reforms that have a low likelihood of succeeding in implementation. Whereas the former finding would be consistent with research that suggests government officials will resist reforms when they reach into the areas of rent-seeking and corruption that elites want

¹⁹⁴ See Hagen 2009; Fang and Stone 2012; Hawkins et al. 2006; and Momani 2007.

to preserve and protect (Rajan 2009),¹⁹⁵ the latter interpretation would be consistent with recent work by Lant Pritchett, Matt Andrews, and Michael Woolcock on the phenomena of “isomorphic mimicry” and “premature load-bearing” in developing countries (Andrews 2013; Andrews et al. 2013; and Pritchett et al. 2013, 2014).

Finally, it is important to note that some external assessment characteristics are *not* associated with more or less successful reform efforts. We find no statistically significant difference between the success of reforms influenced by external assessments with prescriptive policy guidance and the success of reforms influenced by external assessments with only diagnostic content. There is also no apparent relationship between reform success and the cross-country or country-specific nature of an external assessment.

Nor do we find evidence that external assessments attached to material or reputational incentives (positive or negative) are associated with significantly more or less successful reform efforts.¹⁹⁶ While some governments may choose to design and implement reforms in order to obtain the rewards or avoid the penalties associated with improved or declining assessment performance, rewards and penalties do not by themselves seem to increase or decrease the likelihood that reforms will help governments actually solve policy problems.

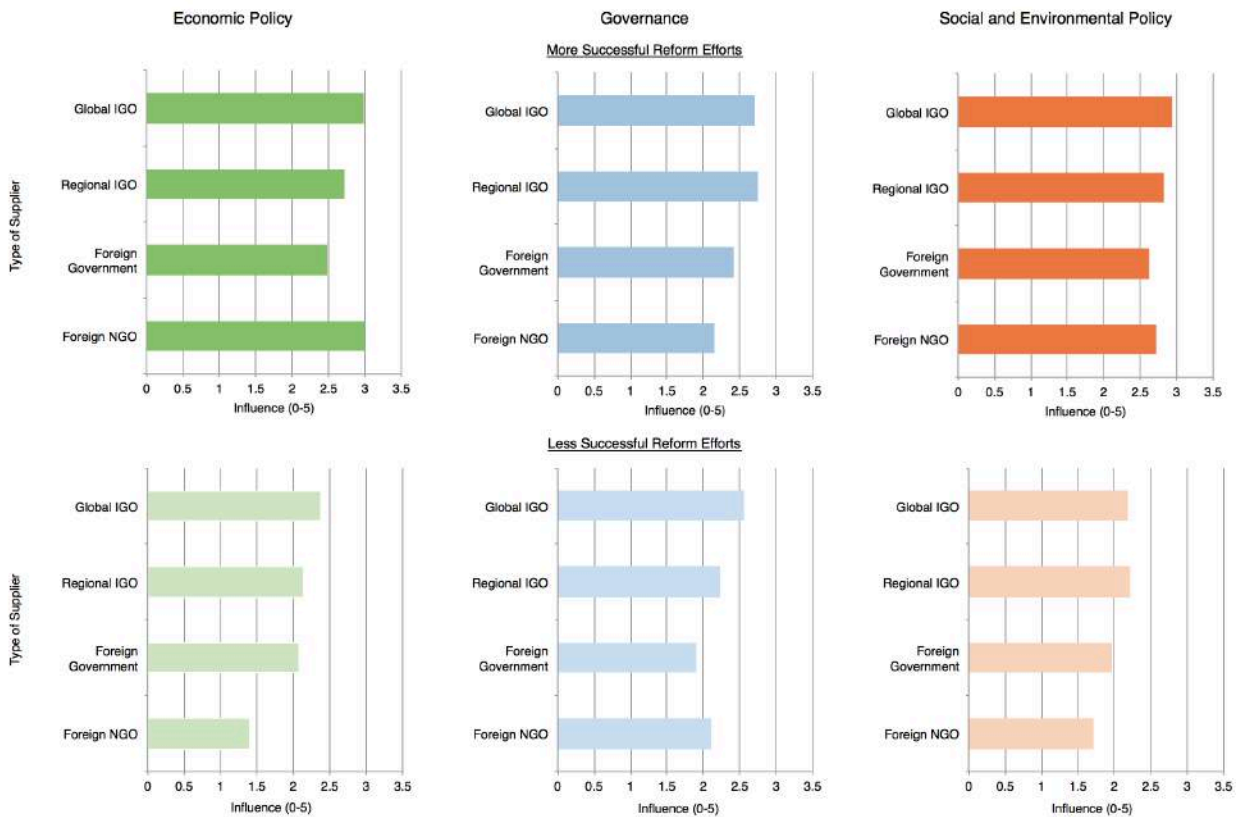
In Figures 44 and 45, we present the influence of external assessments by the type of assessment supplier at the agenda-setting and reform design stages, as reported by participants who perceived reforms as more or less successful.¹⁹⁷ The results reflect our earlier findings that nearly all of the external assessments have greater influence over successful reform efforts than unsuccessful reform efforts.

¹⁹⁵ For example, State Department cables made available by Wikileaks and IMF monitoring of its member states demonstrate that government officials in resource-rich countries often seek to avoid or undermine transparency-related reforms in the energy and mining sector (Yovanovitch 2006a, 2006b; Litzenberger 2007; Efirid 2007a, 2007b; and IMF 2013). There is also a body of literature that demonstrates natural-resource rich countries are significantly less like to respond to external pressures for reform (Girod and Tobin 2011; Mumssen et al. 2013; Parks 2014; and Buch et al. 2014).

¹⁹⁶ There is a rich literature on the extent to which material incentives attached to external assessments of government performance influence the behavior of policymakers in “target” countries (Chee et al. 2007; Orbie and Tortell 2009; Krasner 2011; Kelley 2004; Schimmelfennig and Sedelmeier 2004; Fan et al. 2013; David-Barret and Okamura 2013; and Parks 2014). There is also a body of literature on whether financial and other material penalties deter backsliding and/or encourage reform (Martin 1992; Noland 1997; and Hovi et al. 2005).

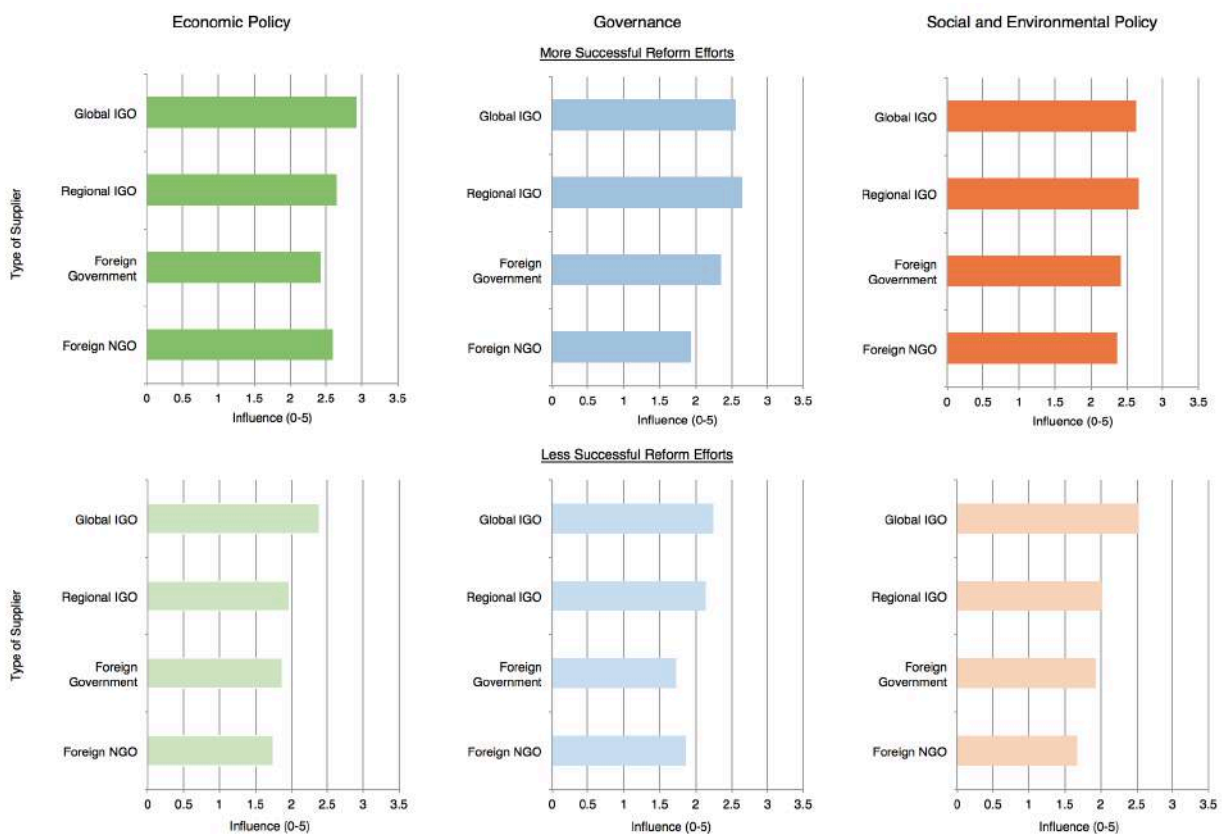
¹⁹⁷ We define reform success as progress towards solving specific, self-identified policy problems.

Figure 44. The Agenda-Setting Influence of Assessment Suppliers, by Level of Success in Reform Implementation



Note: Agenda-setting influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence." Reforms were categorized into more and less successful groups using a threshold score of greater than two for inclusion as a more successful reform effort. Reform success estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all," 2 = "Only a little progress," 3 = "A moderate amount of progress," and 4 = "A great deal of progress."

Figure 45. Reform Design Influence of Assessment Supplier Types, by Level of Success in Reform Implementation



Note: Reform design influence is on a scale of 0-5, where 0 means "No influence at all" and 5 means "Maximum influence." Reforms were categorized into more and less successful groups using a threshold score of greater than two for inclusion as a more successful reform effort. Reform success estimates progress made toward solving specific problems and is on a scale of 1-4, where 1 = "No progress at all", 2 = "Only a little progress", 3 = "A moderate amount of progress", and 4 = "A great deal of progress".

There are at least two ways of interpreting these results. Reform-oriented governments may be more likely to marshal external assessments in the service of their own domestic policy objectives. Similarly, external evaluations of government performance may be less important to governments that face the strong domestic impediments to reform. It could, however, also be the case that reforms are more likely to succeed *because* of the influence that external assessments exert on government policies and programs.

Irrespective of whether participants perceived reforms as being more or less successful, both global and regional IGO suppliers of assessments wielded more influence than other types of assessment suppliers at the agenda-setting and reform design phases of the policymaking process. The outsized influence of foreign NGOs at the agenda-setting stage of successful economic reform efforts serves as the only exception to this overall trend.

In governance policy, the absolute difference in the overall levels of policy influence between "more successful reform" contexts and "less successful reform" contexts, was largest for foreign government assessment suppliers. This finding suggests that the success of governance reforms may be associated with the target country's openness to assessment by foreign governments (e.g., US, UK, and Germany). Economic and social and environmental policy reform efforts run counter to this finding.

The largest decline in assessment influence between more and less successful economic and social and environmental policy reform efforts is for foreign NGOs. Notably, the narrowest decline in assessment influence between more and less successful economic reform efforts is for foreign governments. The influence of foreign government-supplied assessments over economic reform is least strongly associated with reform success.

Our results pertaining to foreign non-governmental (NGO) assessment suppliers merit special attention. With respect to more successful economic policy and governance reform efforts, we observe the greatest declines in influence between the two stages of the policymaking process.¹⁹⁸ This pattern suggests that, for non-governmental organizations, there may be a “policy influence penalty” at the reform design stage associated with the absence of a strong in-country presence.

It is important to note that many of the other assessment supplier types represented in Figure 45 have a reasonably strong “ground game” in our sample of low- and middle-income countries (Parks and Rice 2014). Their in-country presence likely makes it possible to work in closer collaboration with the host government authorities on the design of new policy, institutional, legal, regulatory, and administrative reforms (once the authorities have decided that they want to pursue a particular set of reforms in the first place).¹⁹⁹

¹⁹⁸ Declines range from average assessment influence scores of 3.001 (economic policy) and 2.145 (governance) during the agenda-setting stage to average assessments influence scores of 2.597 (economic policy) and 1.935 (governance) during the reform design stage.

¹⁹⁹ It is interesting, however, that this pattern does not hold in social and environmental policy—an area in which some specialized health and education NGOs may have more influence in reform design—or in less successful reform contexts.

7. Conclusion

We now conclude where we began—the reform-minded Minister of Finance in an aid-dependent country who must decide where to invest her time, money, and political capital. Facing a patchwork quilt of external performance assessments, she must now decide which sources of analysis and advice to use and how to use them. Thanks to the *2014 Reform Efforts Survey*, the Minister’s most trusted adviser is now able to draw upon the experiences and insights of 6,750 other development policymakers and practitioners who have faced similar dilemmas.

Having internalized the key lessons from this report, the adviser would likely tell the Minister that while external assessments of government performance can serve many purposes, past experience suggests that such assessments are most useful for diagnosing policy problems and identifying potential solutions. He would likely warn the Minister that while some external performance assessments are tied to attractive financial or reputational benefits, she should be careful not to let external incentives distract her from the country’s most critical policy problems; such incentives are not associated with significantly more or less successful reform efforts.

The Minister’s adviser would probably tell her that, of the wide variety of assessments at her disposal, the policy analysis and advice contained in the assessments of large, global international organizations have proven to be particularly valuable to other reformers. He might also warn the Minister not to rely on the analysis and advice provided by assessment suppliers who lack an in-country presence. These suppliers may not be able to provide the contextual insights needed for the effective design of specific reform features. Additionally, after explaining that external performance assessments are generally associated with more successful reform efforts when they source their data from the governments they assess, the adviser would probably suggest that the Minister pay particularly close attention to assessments that draw upon the data that the government is already producing.

The Minister and her adviser would eventually confront the issue of whether to use country-specific performance assessments or cross-country benchmarking assessments. The adviser would likely explain that each type of assessment has its own advantages and disadvantages. Country-specific assessments contain in-depth diagnostic and advisory content, which may resonate with local stakeholders because they are more attuned to local needs and realities. These assessments tend to be more popular with policymakers across the developing world; however, they can also leave host government officials with a false sense of confidence in the analysis and advice they provide, thus bypassing critical internal processes of introspection, deliberation, and iterative problem solving. Cross-country benchmarking assessments, on the other hand, are less attractive at first blush because they lack nuance and context specificity. However, when decision-makers confront complex issues of corruption, informality, and institutional dysfunction, they can provide greater policy flexibility and maneuverability than country-specific assessments, enabling the government to experiment with different reform strategies and iteratively adapt in pursuit of better, de facto outcomes.

Noting the Minister’s interest in Greek mythology, the adviser might remind her of the many perils Odysseus faced on his journey home from Troy and counsel the Minister to avoid steering her country into the “capability trap” that has ensnared other well-intentioned, would-be reformers. He would encourage her to focus on simple, solvable problems before undertaking more complex reforms. To underscore his point, he might even slip a copy of a short article by Lant Pritchett, Michael Woolcock, and Matt Andrews into the Minister’s read file, which describes the capability trap as “a dynamic that enables [policymakers] to document instances of apparent reform and thus assure a continued flow of development resources to their country or sector, despite the fact that the reforms themselves may be generating few actual improvements in performance.”²⁰⁰

²⁰⁰ Pritchett et al. 2014: 63.

Finally, the adviser would likely counsel the Minister that officials within her government will view and value external performance assessments in very different ways, depending on the positions that they hold and the incentives they face. The President will emphasize his government's sovereign authority to make final policy decisions, while his chief of staff and various advisers will separately convey the importance of securing external validation and rewards from donors and international organizations. Senior civil servants, for their part, will emphasize complementarity with existing reform efforts. The Minister's adviser would therefore urge her to be smart and strategic in her interactions with these government counterparts and to not lose sight of the end goal: pursuing reforms that solve real problems and are grounded in local needs and realities.

How do money, ideas and, reforms come together to produce better development outcomes? This report begins to shed light on some facets of this question, but more data and research is needed. With survey data anchored in the experiences of real-world decision-makers, we believe scholars and policymakers can make major advances in our understanding of the causes and consequences of reform in developing countries. To this end, AidData will release a series of publications in 2015 and 2016 that continue to make use of the rich data contained in the *2014 Reform Efforts Survey*. In subsequent reports, we will seek to further understand how Western and non-Western development partner organizations are attempting to help their host government counterparts implement reforms and improve development outcomes – and to what effect.

8. References

Acemoglu, Daron and James A. Robinson. 2000. Political Losers as a Barrier to Economic Development. *American Economic Review* 90: 126-130.

Acemoglu, Daron, Simon Johnson, Pablo Querubin, and James A. Robinson. 2008. When Does Policy Reform Work: The Case of Central Bank Independence. *Brookings Papers on Economic Activity* (1): 351-418.

Alesina, Alberto and Allan Drazen. 1991. Why Are Stabilizations Delayed? *American Economic Review* 81(5): 1170-1188.

Aklin, Michaël and Johannes Urpelainen. 2014. The Global Spread of Environmental Ministries: Domestic–International Interactions. *International Studies Quarterly* 58: 764–780.

Amin, Mohammad, and Simeon Djankov. 2009. Democracy and Reforms. World Bank. Policy Research Working Paper 4835. Washington D.C.: World Bank.

Andersson, Staffan, and Paul Heywood. 2009. The politics of perception: Use and abuse of Transparency International's approach to measuring corruption. *Political Studies* 57 (4): 746–767.

Andrews, Matt. 2009. Isomorphism and the Limits to African Public Financial Management Reform. Harvard Kennedy School Faculty Research Working Paper Series, RWP09-012. Cambridge, MA: Harvard University.

Andrews, Matt. 2010. How Far Have Public Financial Management Reforms Come in Africa? Harvard Kennedy School Faculty Research Working Paper Series, FRW 10-018. Cambridge, MA: Harvard University.

Andrews, Matt. 2011. Which Organizational Attributes Are Amenable to External Reform? An Empirical Study of African Public Financial Management. *International Public Management Journal* 14(2): 131–56.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge, UK: Cambridge University Press.

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2013. Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA). *World Development* 51(C): 234-244.

Arndt, C. 2008. The politics of governance ratings. *International Public Management Journal* 11 (3): 1-23

Arruñada, Benito, 2007. Pitfalls to avoid when measuring institutions: Is Doing Business damaging business? *Journal of Comparative Economics* 35(4): 729-747.

Asian Development Bank (ADB). 2005. Multilateral Development Bank Technical Meeting on Performance Based Allocation Methods. Meeting at AsDB Headquarters on 24-25 January 2005. Manila: ADB.

Asian Development Bank (ADB). 2009. *ADB Perceptions Survey: Multinational Survey of Stakeholders 2009*. Manila: ADB.

Asian Development Bank (ADB). 2010. Georgia: A Country at an Ancient Crossroads Looks to the Future. Development Effectiveness Brief. Manila: ADB.

- Asian Development Bank (ADB). 2013. *ADB Perceptions Survey: Multinational Survey of Stakeholders 2012*. Manila: ADB.
- Atkinson, Carol. 2010. Does Soft Power Matter? A Comparative Analysis of Student Exchange Programs 1980-2006. *Foreign Policy Analysis* 6 (1): 1-22.
- Avey, Paul C. and Michael C. Desch. 2014. What Do Policymakers Want From Us? Results of a Survey of Current and Former Senior National Security Decision Makers. *International Studies Quarterly* 58 (2): 227-246.
- Bagaudinova, Svetlana, Dana Omran and Umar Shavurov. 2007. "Georgia: Licensing 159 activities—not 909." In *Celebrating Reform 2007*. Washington DC: World Bank.
- Ban, C. and W. Vandenabeele. W. 2009. Motivation and Values of European Commission Staff. Paper presented at the European Union Studies Association meeting, Marina del Rey, CA.
- Barder, Owen. 2010. Can aid create incentives for politicians in developing countries? 27 April 2010. Retrieved from <http://www.owen.org/blog/3275>.
- Betley, Mary. 2008. Assessing Impact of the PEFA Framework. Manuscript produced for PEFA Steering Committee.
- Brown, Keith and Jill Tirnauer. 2009. Trends in U.S. Foreign Assistance over the Past Decade. Washington D.C.: USAID/Management Systems International.
- Birdsall, Nancy and Rita Perakis. 2012. Cash on Delivery Aid: Implementation of a Pilot in Ethiopia. April 27, 2012. Washington DC: Center for Global Development.
- Biron, Carey L. 2013. World Bank Formally Urged to Overhaul 'Doing Business' Report. *Inter Press Service*. 24 June 2013. Retrieved from <http://www.ipsnews.net/2013/06/world-bank-formally-urged-to-overhaul-doing-business-report/>
- Bishin, Benjamin G., Robert R. Barr, and Matthew J. Lebo. 2006. The Impact of Economic Versus Institutional Factors in Elite Evaluations of Presidential Progress Toward Democracy in Latin America. *Comparative Political Studies* 39 (10): 1-26.
- Bjuremalm, H. 2006. Power Analysis: Experiences and Challenges. Concept Note. Department for Democracy and Social Development. Stockholm: SIDA.
- Blum, Jürgen René. 2014. What Factors Predict How Public Sector Projects Perform? A Review of the World Bank's Public Sector Management Portfolio. Policy Research Working Paper 6798. Washington DC: World Bank.
- Böhmelt, Tobias and Tina Freyburg. 2013. The temporal dimension of the credibility of EU conditionality and candidate states' compliance with the *acquis communautaire*, 1998-2009. *European Union Politics* 14 (2): 1-23.
- Böhmelt, Tobias and Tina Freyburg. 2015. Diffusion of Compliance in the 'Race towards Brussels?' A Spatial Approach to EU Accession Conditionality. *West European Politics* 38(3): 601-626.
- Booth, David. 2011. Aid, Institutions and Governance: What Have We Learned? *Development Policy Review* 29 (s1): s5-s26.

Bor, J. 2007. The Political Economy of AIDS Leadership in Developing Countries: An Exploratory Analysis. *Social Science and Medicine* 64 (8): 1585-1599.

Borrman, Axel and Matthias Busse. 2007. The Institutional Challenge of ACP/EU Economic Partnership Agreements. *Development Policy Review* 25 (4): 403-16.

Bourguignon, François and Jean-Philippe Platteau. Forthcoming. The Hard Challenge of Aid Coordination. *World Development*.

Buch, Benjamin P., Mark T. Buntaine, and Bradley C. Parks. 2014. Aiming at the Wrong Targets: The Difficulty of Improving Domestic Institutions with International Aid. AidData Working Paper #4. Williamsburg, VA: AidData.

Bueno de Mesquita, Bruce and Alastair Smith. 2009. Political Survival and Endogenous Institutional Change. *Comparative Political Studies* 42(2): 167-197.

Bueno de Mesquita, Bruce, and Alastair Smith. 2010. Leader Survival, Revolutions and the Nature of Government Finance. *American Journal of Political Science* 54: 936-950.

Bueno de Mesquita, Bruce. 2011. *The Dictator's Handbook: Why Bad Behavior Is Almost Always Good Politics*. New York, NY: PublicAffairs.

Buntaine, Mark T., and Bradley C. Parks. 2013. When Do Environmentally-Focused Assistance Projects Achieve their Objectives? Evidence from World Bank Post-Project Evaluations. *Global Environmental Politics* 13(2): 65-88.

Busia, Kojo. 2014. "The Search for Domestic Accountability in Development Aid: What Role for the APRM in Reshaping Governance in Africa?" *Problems, Promises, and Paradoxes of Aid: Africa's Experience*, edited by Muna Ndulo and Nicolas van de Walle. Pp. 171-210. Newcastle upon Tyne, UK: Cambridge Scholars Publishing/Institute for African Development.

Butts, Cassandra. 2011. Deepening MCC's Commitment to Gender Equality. MCC's Poverty Reduction Blog. 18 November 2011. Retrieved from <https://www.mcc.gov/pages/povertyreductionblog/entry/deepening-mccs-commitment-to-gender-equality>.

Cadier, David. 2013. "Is the European Neighbourhood Policy a Substitute for Enlargement?" in *The Crisis of EU Enlargement*, LSE IDEAS Special Report. London, UK: London School of Economics and Political Science.

Caribbean Development Bank (CDB). 2008. Progress Report on Implementation of the Caribbean Development Bank's Poverty Reduction Strategy. Barbados: CDB. Retrieved from http://www.caribank.org/uploads/projects-programmes/special-development-fund/sdf-meetings/negotiation-meetings/CDB%20PRS_Progress%20_Report.pdf

Center for Global Development (CGD) and the African Population and Health Research Center (APHRC). 2014. Delivering on the Data Revolution in Sub-Saharan Africa: A Final Report of the Data for African Development Working Group. Washington DC: CGD. Retrieved from <http://www.cgdev.org/sites/default/files/CGD14-01%20complete%20for%20web%200710.pdf>

Cerny, Philip G. 1997. Paradoxes of the "Competition State": The Dynamics of Political Globalisation. *Government and Opposition* 32 (2): 251-274.

Charron, Nicholas. 2009. The Impact of Socio-Political Integration and Press Freedom on Corruption. *Journal of Development Studies* 45 (9): 1472-1493.

Chee, Grace, Natasha Hsi, Kenneth Carlson, Slavea Chankova and Patricia Taylor. 2007. Evaluation of the First Five Years' of GAVI Immunization Services Support Funding. Bethesda, MD: Abt Associates Inc.

Chhotray, Vasudha and David Hulme. 2009. Contrasting Visions for Aid and Governance in the 21st Century: The White House Millennium Challenge Account and DFID's Drivers of Change. *World Development* 37(1): 36-49.

Chwieroth, Jeffrey M. 2007. Neoliberal Economists and Capital Account Liberalization in Emerging Markets. *International Organization* 61 (2): 443-463.

Chwieroth, Jeffrey M. 2008. Cheerleading for Policy Goals: The International Monetary Fund and Capital Account Liberalization in Emerging Markets. Paper presented at the Conference on the Political Economy of International Organizations, Centro Stefano Franscini, Monte Verità, Switzerland, 3 – 8 February 2008.

Chwieroth, Jeffrey M. 2009a. *Capital Ideas: The IMF and the Rise of Financial Liberalization*. Princeton: Princeton University Press.

Chwieroth, Jeffrey. 2009b. 'The Silent Revolution': Professional Characteristics, Sympathetic Interlocutors and IMF Lending. LSE Working Paper.

Chwieroth, Jeffrey M. 2010. How Do Crises Lead to Change? Liberalizing Capital Controls in the Early Years of New Order Indonesia. *World Politics* 62 (3): 496-527.

Cho, Seo-Young, Axel Dreher and Eric Neumayer. 2014. The Determinants of Anti-trafficking Policies: Evidence from a New Index. *Scandinavian Journal of Economics* 116 (2): 429– 454.

Clist, Paul, Alessia Isopi, and Oliver Morrissey. 2012. Selectivity on Aid Modality: Determinants of Budget Support from Multilateral Donors. *The Review of International Organizations* 7 (3): 267-284.

Cobham, Alex. 2013. Corrupting Perceptions: Why Transparency International's flagship corruption index falls short. *Foreign Policy*. 22 July 2013. Retrieved from <http://foreignpolicy.com/2013/07/22/corrupting-perceptions/>

Cook, Nicolas. 2010. Liberia's Post-War Development: Key Issues and U.S. Assistance. Washington D.C.: Congressional Research Service.

Corrales, Javier. 2006. "Political Obstacles to Expanding and Improving Schooling in Developing Countries." In *Global Educational Expansion: Historical Legacies and Political Obstacles*, edited by Aaron Benavot, Julia Resnik and Javier Corrales. Cambridge, MA: Academy of Arts and Sciences. Pp. 91-148.

Couper, Mick P. 2000. Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly* 64: 464-494.

Couper, Mick P. 2008. *Designing Effective Web Surveys*. Cambridge: Cambridge University Press

Dahl-Østergaard, Tom, Sue Unsworth, Mark Robinson, and Rikke Ingrid Jensen. 2005. Lessons Learned on the Use of Power and Drivers of Change Analyses in Development Cooperation. Review commissioned by the OECD DAC Network on Governance.

David-Barrett, Elizabeth, and Ken Okamura. 2013. The Transparency Paradox: Why Corrupt Countries Join the Extractive Industries Transparency Initiative. Paper presented at 2013 APSA Annual Conference, Chicago, Illinois.

Davis, Kevin, Benedict Kingsbury, and Sally Engle Merry. 2012. Indicators as a Technology of Global Governance. *Law & Society Review* 46 (1): 71-104.

Davison, William. 2012. As Ethiopia looks beyond strongman Meles, fears of instability (+video). *The Christian Science Monitor*. 22 August 2012. Retrieved from <http://www.csmonitor.com/World/Africa/2012/0822/As-Ethiopia-looks-beyond-strongman-Meles-fears-of-instability-video>

de Almeida, Paulo Roberto. 2013. Sovereignty and Regional Integration in Latin America: a political conundrum? *Contexto Internacional* 35 (2): 471-495.

De Bruyn, Tom. 2015. "New Friends", "Easier Partners" and "Bigger Brothers": The Influence of the Rising Powers on Agriculture and Food Security in Malawi. Paper presented at the Chinese Overseas Finance Conference 2015, Johns Hopkins University's School of Advanced International Studies (SAIS), Washington, DC, April, 10, 2015.

de Renzio, Paolo. 2009. Taking Stock: What do PEFA Assessments tell us about PFM systems across countries? ODI Working Paper 302. London, UK: Overseas Development Institute.

de Renzio, Paolo and Harika Masud. 2011. Measuring and Promoting Budget Transparency: The Open Budget Index as a Research and Advocacy Tool. *Governance: An International Journal of Policy, Administration, and Institutions* 24 (3): 607-616.

de Renzio, Paolo and Ngaire Woods. 2008. The Trouble with Cash on Delivery Aid: A note on its potential effects on recipient country institutions. Note prepared for the CGD Initiative on 'Cash on delivery Aid'. Mimeo. Oxford: University of Oxford. Accessed at <http://www.cgdev.org/doc/Cash%20on%20Delivery%20AID/DeRenzio%20Woods.pdf>

de Sherbinin, Alex, Aaron Reuben, Macy Levy, and Laura Johnson. 2013. Indicators in Practice: How Indicators are Being Used in Policy and Management Contexts. Palisades, NY and New Haven, CT: CIESIN, Columbia University and YCELP.

Del Biondo, Karen. forthcoming. Donor Interests or Developmental Performance? Explaining Sanctions in EU Democracy Promotion in sub-Saharan Africa. *World Development*.

Del Biondo, Karen and Jan Orbie. 2014. The European Commission's implementation of budget support and the Governance Incentive Tranche in Ethiopia: democracy promoter or developmental donor? *Third World Quarterly* 35 (3): 411-427.

Dervis, Kemal, Homi Kharas, and Noam Unger. 2010. Aiding Development: Assistance Reform for the 21st century. Brookings Blum Roundtable Report. Washington DC: Brookings Institution.

Devarajan, S., D. Dollar, and T. Holmgren. 2001. *Aid and Reform in Africa: Lessons from Ten Case Studies*. Washington, DC: World Bank.

Dewatripont, Mathias and Gerard Roland. 1992. Economic Reform and Dynamic Political Constraints. *Review of Economic Studies* 59 (4): 703-730.

Dharmasaputra, Metta and Abe Wahyudi. 2014. The Impact of Payment-for-Performance Finance on the Political Economy of Deforestation of Indonesia. *Center for Global Development*. Climate and Forest Paper Series #9: 1-55. <http://www.cgdev.org/sites/default/files/CGD-Climate-Forest-Series-9-Dharmasaputra-Wahyudi-Indonesia.pdf>

Diamond, Larry, 2008. The Democratic Rollback: The Resurgence of the Predatory State. *Foreign Affairs* 87 (2): 36-48.

Diamond, Jack. 2013. Background Paper 1: Sequencing PFM Reforms. January 2013. Washington DC: PEFA Secretariat.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (3rd Ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Dinmore, Guy. 2005. Georgia calls on Europe for coherent policy. *The Financial Times*. September 25, 2005.

Diofasi, Anna and Nancy Birdsall. 2015. Three Comments on the Proposed SDGs. CGD Blog Post. 26 January 2015. Accessed at <http://www.cgdev.org/blog/three-comments-proposed-sdgs>

Director General for External Policies of the European Union. 2011. The Future of EU Budget Support in Developing Countries. Brussels: European Parliament. Retrieved from [http://www.europarl.europa.eu/RegData/etudes/note/join/2011/433769/EXPO-DEVE_NT\(2011\)433769_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/note/join/2011/433769/EXPO-DEVE_NT(2011)433769_EN.pdf)

Djankov, Simeon. 2009. The Regulation of Entry: A Survey. *World Bank Research Observer* 24 (2): 183–203.

Doemeland, Doerte and James Trevino. 2014. Which World Bank Reports Are Widely Read? World Bank Policy Research Working Paper 6851. Washington DC: World Bank.

Doig, Alan and David Norris. 2012. Improving anti-corruption agencies as organisations. *Journal of Financial Crime* 19 (3): 255 – 273.

Dong, Lili and David W. Chapman. 2008. The Chinese Government Scholarship Program: An Effective Form of Foreign Assistance? *International Review of Education* 54 (2): 155-173.

Dreher, Axel, Silvia Marchesi, and James Vreeland. 2008. The political economy of IMF forecasts. *Public Choice* 137(1): 145–171

Easterly, William. 2014. *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*. New York NY: Basic Books.

Eaton, Kent, Kai Kaiser, and Paul Smoke. 2010. *The Political Economy of Decentralization Reforms: Implications for Aid Effectiveness*. Washington DC: World Bank.

Economist Intelligence Unit (EIU). 2007. Country Report: Liberia. September 2007. London, UK: Economist Intelligence Unit.

Edwards, Martin S. 2011. Under the Microscope: Some Findings from the 2011 Triennial Surveillance Review. Written for the New Rules for Global Finance Coalition, December 1, 2011.

Edwards Jr. D. Brent. 2012. Small states and big institutions: USAID and education policy formation in El Salvador, *Current Issues in Comparative Education* 15 (1): 82-98.

Efird, Cythnia. 2007a. Angola Decides Not to Seek IMF Agreement. 14 March 2007. Retrieved at http://www.wikileaks.org/plusd/cables/07LUANDA250_a.html

Efird, Cythnia. 2007b. IMF Article IV Team Gives Angola High Marks. 7 June 2007. Retrieved at http://www.wikileaks.org/plusd/cables/07LUANDA564_a.html

Ellinas, Antonis A. and Ezra Suleiman. 2012. *The European Commission and Bureaucratic Autonomy: Europe's Custodians*. Cambridge, UK: Cambridge University Press.

European Commission (EC). 2008. The "MDG Contract": An Approach for Longer-Term and More Predictable General Budget Support. Brussels. European Commission. Retrieved from http://ec.europa.eu/development/icenter/repository/MDG-Contract-Paper-0608_en.pdf
http://ec.europa.eu/development/icenter/repository/MDG-Contract-Paper-0608_en.pdf

European Stability Initiative (ESI). 2010. Reinventing Georgia: The Story of a Libertarian Revolution. April 2010. Berlin: European Stability Initiative. Retrieved at www.esiweb.org/index.php?lang=en&id=322&debate_ID=3

Fan, Victoria, Denizhan Duran, Rachel Silverman, and Amanda Glassman. 2013. Performance-based financing at the Global Fund to Fight AIDS, Tuberculosis and Malaria: an analysis of grant ratings and funding, 2003-2012. *Lancet Global Health* 1(3): e161-68.

Fang, Songying, and Randall Stone. 2012. International Organizations as Policy Advisors. *International Organization* 66(4): 537-71.

Fernandez, Raquel and Dani Rodrik. 1991. Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty. *American Economic Review* 81(5): 1146-1155.

Flores, Thomas Edward, Gabriella Lloyd, and Irfan Nooruddin. 2013. The Technocratic Advantage? Leadership and IMF Programs. Paper presented at the 2013 Annual Convention of the International Studies Association, San Francisco, CA.

Fox, William, Sayeed Bayat, and Naas Ferreira. 2007. *A Guide to Managing Public Policy*. Cape Town, South Africa: Juta Academic.

Gabre-Madhin, Eleni Z., and Steven Haggblade, S. 2001. *Successes in African Agriculture: Results of an Expert Survey*. Washington DC: International Food Policy Research Institute. Retrieved from http://www.ifpri.org/sites/default/files/publications/syn04_survey.pdf

Gamurari, Ludmila and Cristian Ghinea. 2014. It has only just begun: EU and anticorruption institutions in Moldova. 1 August 2014. Brussels, Belgium: European Policy Centre.

Ganster, D. C., H. W. Hennessey, and F. Luthans. 1982. The role of social desirability in organizational behavior research. *Academy of Management Proceedings* 1: 141-145.

Gavrillis, George. 2014. Forging a National Strategy Through EU Accession: Serbia, 2007–2012. *Innovations for Successful Societies Case Study*. Princeton, NJ: Princeton University.

Geertson, Michael. 2010. Achieving Quick Impact in Millennium Challenge Corporation Threshold Projects. *Yale Journal of International Affairs* Winter 2010: 34-48.

Gilbert, C., A. Powell and D. Vines. 1999. Positioning The World Bank. *The Economic Journal* 109(459): 598-633.

Girod, Desha M. and Jennifer Tobin. 2011. Take the Money and Run: The Determinants of Compliance with Aid Agreements. Paper presented at the American Political Science Association Annual Meeting, Seattle, Washington.

Girod, Desha and Meir Walters. 2012. Elite-Led Democratisation in Aid-Dependent States: The Case of Mauritania. *Journal of North African Studies* 17 (2): 181-193.

Global Environment Facility (GEF). 2013. Mid-Term Evaluation of the System of Transparent Allocation of Resources. Prepared by the GEF Evaluation Office. Washington DC: GEF. [http://www.thegef.org/gef/sites/thegef.org/files/documents/GEF.ME_C.45.04%20MTE%20of%20STAR%20\(EO\).pdf](http://www.thegef.org/gef/sites/thegef.org/files/documents/GEF.ME_C.45.04%20MTE%20of%20STAR%20(EO).pdf)

Global Integrity. 2008a. Cameroon's War on Corruption: Reality or Good PR? 03 July 2008. Retrieved on 20 January 2012 at <http://www.globalintegrity.org/node/226>

Global Integrity. 2008b. Cameroon's War on Corruption: Probably Good PR. Retrieved on 20 January 2012 at <http://www.globalintegrity.org/node/224>

Goldsmith, Arthur A. 2011. No Country Left Behind? Performance Standards and Accountability in US Foreign Assistance. *Development Policy Review* 29: s157–s176.

Grabbe, H. 2002. European Union conditionality and the Acquis Communautaire. *International Political Science Review* 23 (3): 249-268.

Gray, Julia and Jonathan Slapin. 2012. How Effective Are Preferential Trade Agreements? Ask the Experts. *Review of International Organizations* 7(3): 309-333.

Grek, Sotiria. 2009. Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy* 24 (1): 23-37.

Grindle, Merilee. 2004. *Despite the Odds: The Contentious Politics of Education Reform*. Princeton, NJ: Princeton University Press.

Grossman, Herschel I. 1980. "Rational Expectations, Business Cycles, and Government Behavior." In *Rational Expectations and Economic Policy*. Edited by Stanley Fischer. Chicago: University of Chicago Press.

Gruber, Lloyd. 2000. *Ruling the World: Power Politics and the Rise of Supranational Institutions*. Princeton, NJ: Princeton University Press.

Hackenesch, Christine. 2013. Aid Donor Meets Strategic Partner? The European Union's and China's Relations with Ethiopia. *Journal of Current Chinese Affairs* 42 (1): 7–36.

Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor, and James H. Fowler. 2014. Decision Maker Preferences for International Legal Cooperation. *International Organization* 68: 845–876.

Hagen, Rune Jansen. 2009. Basic Analytics of Multilateral Lending and Surveillance. *Journal of International Economics* 79 (2009):126-13

Hallward-Driemeier, Mary, Gita Khun-Jush, and Lant Pritchett. 2010. Deals versus Rules: Policy Implementation Uncertainty and Why Firms Hate It. NBER Working Paper 16001. Cambridge, MA: National Bureau of Economic Research.

Hallward-Driemeier, Mary and Lant Pritchett. 2011. How Business is Done and the ‘Doing Business’ Indicators: The Investment Climate when Firms have Climate Control. World Bank Policy Research Working Paper 5563. Washington D.C.: World Bank.

Hamilton, Kirk and Michael Clemens. 1999. Genuine savings rates in developing countries. *World Bank Economic Review* 13:333–356.

Hartwig, K., J. Pashman, E. Cherlin, M. Dale, and M. Callaway. 2008. Hospital management in the context of health sector reform: a planning model in Ethiopia. *International Journal of Health Planning and Management* 23: 203–218.

Hawkins, Darren, David Lake, Daniel Nielson, and Michael Tierney. 2006. *Delegation and Agency in International Organizations*. Cambridge, UK: Cambridge University Press.

Hedger, Edward and Zainab Kizilbash Agha. 2007. Reforming Public Financial Management When the Politics Aren’t Right: A Proposal. ODI Opinion Paper No. 89. London, UK: Overseas Development Institute.

Hedger, Edward and Paulo de Renzio. 2010. What do Public Financial Management assessments tell us about PFM reform? ODI Background Note. July 2010. London, UK: Overseas Development Institute.

Heilbrunn, John R. 2004. Anti-corruption commissions: panacea or real medicine to fight corruption? Working Paper Series 37234. Washington DC: World Bank Institute.

Helleiner, Eric and Bessma Momani. 2008. “Slipping Into Obscurity: Crisis and Institutional Reform as the IMF,” in *Can the World Be Governed? Possibilities for Effective Multilateralism*, edited by Alan Alexandroff. Wilfred Laurier University Press. pp. 353-388.

Herbert, R., and S. Gruzd. 2008. The African Peer Review Mechanism: Lessons from the Pioneers. Johannesburg: SAIIA.

Hicks, Robert, Bradley Parks, Timmons Roberts and Michael Tierney. 2008. *Greening Aid? Understanding the Environmental Impact of Development Assistance*. Oxford: Oxford University Press.

Hille, Peter, and Christoph Knill. 2006. It’s the Bureaucracy, Stupid: The Implementation of the Acquis Communautaire in EU Candidate Countries, 1999-2003. *European Union Politics* 7 (4): 531-52.

Hoey, Lesli. 2015. “Show me the Numbers”: Examining the Dynamics Between Evaluation and Government Performance in Developing Countries. *World Development* 70: 1–12.

Holder, William E. 2003. The International Monetary Fund’s Involvement in Combating Money Laundering and the Financing of Terrorism. *Journal of Money Laundering Control* 6 (4): 383–87.

Hood, Christopher. 2006. Gaming in Targetworld: The Targets Approach to Managing British Public Services. *Public Administration Review* 66 (4): 515-521.

Hood, Christopher. 2012. Public management by numbers as a performance-enhancing drug: two hypotheses. *Public Administration Review* 72(1): S85–S92.

Honig, Dan. 2014. Navigation by Judgment: Organizational Autonomy in the Delivery of Foreign Aid. Paper presented at the 2014 Northeast Universities Development Consortium Conference in Boston, Massachusetts.

Hovi, Jon, Robert Huseby, and Detlef F. Sprinz. 2005. When Do (Imposed) Economic Sanctions Work? *World Politics* 57(4): 479-499.

Hout, Wil. 2007. *The Politics of Aid Selectivity Good Governance Criteria in World Bank, U.S. and Dutch Development Assistance*. London, UK: Routledge.

Hout, Wil. 2012. The anti-politics of development: Donor agencies and the political economy of governance. *Third World Quarterly* 33(3): 405–422.

Hyden, Goran. 2008. After the Paris declaration: Taking on the issue of power. *Development Policy Review* 26 (3): 259–74.

Hoffmann-Lange, Ursula. 2007. "Methods of Elite Research." In *Oxford Handbook of Political Behavior*, edited by Russell J. Dalton, Hans-Dieter Klingemann. Pp. 910-927. Oxford, UK: Oxford University Press.

Høyland, Bjørn, Karl Moene, and Fredrik Willumsen. 2012. The tyranny of international index rankings. *Journal of Development Economics* 97:1–14.

Human Rights Watch. 2009. *World Report 2009*. New York: Human Rights Watch.

Ibrahim, Hadeel. 2012. Fertilizing the Fields for Honest Government: Reflections on leaders, leadership and Africa's challenges: Interview with Hadeel Ibrahim. *Development* 55 (4): 453-457.

Independent Evaluation Group of The World Bank (IEG). 2008. *Doing Business: An Independent Evaluation. Taking the Measure of The World Bank-IFC Doing Business Indicators*. Washington D.C.: World Bank. Retrieved from http://siteresources.worldbank.org/EXTDOIBUS/Resources/db_evaluation

Independent Evaluation Group of The World Bank (IEG). 2008. *Public Sector Reform: What Works and Why? An IEG Evaluation of World Bank Support*. Washington D.C.: World Bank.

Independent Evaluation Group (IEG). 2012. *The Matrix System at Work: An Evaluation of The World Bank's Organizational Effectiveness*. Washington. DC: World Bank.

Innes, Abby. 2013. The Political Economy of State Capture in Central Europe. *Journal Of Common Market Studies* Volume 52 (1): 88-104.

Inter-American Development Bank (IADB). 2010. Evaluation of the Fund for Special Operations during the Eighth Replenishment (1994-2010) -- Part I. Office of Evaluation and Oversight. Washington DC: IADB. Retrieved from <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=38027222>

International Aid Transparency Initiative (IATI). 2013. Country Survey - IATI Data and Aid Information Management Systems (AIMS). Presented at a 2014 Meeting of IATI Steering Committee Members and Observers. Copenhagen, Denmark. Retrieved from <http://www.aidtransparency.net/wp-content/uploads/2013/01/Paper-4a-Country-Survey-of-AIMS.pdf>

International Budget Partnership (IBP) and ONE. 2013. How Can Donors Promote Budget Transparency and Accountability? The Use of Benchmarks and Incentives for Budget Transparency in Foreign Aid Programs. Retrieved from http://internationalbudget.org/wp-content/uploads/Donor-benchmarks-paper_Sept2013.pdf on 25 February 2015.

International Finance Corporation (IFC). 2013. Leading the pack: Rwanda is Africa's best reformer in Doing Business 2014. Retrieved on 26 February 2015 at http://www.ifc.org/wps/wcm/connect/region_ext_content/regions/sub-saharan+africa/news/za_ifc_as_rwanda_doing_business_2014

International Labour Office (ILO). 2007. World Bank Doing Business report: The employing workers indicator. November 2007. Geneva: International Labour Office. Retrieved from http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/documents/meetingdocument/wcms_085125.pdf

International Monetary Fund (IMF) and The World Bank (WB). 2009. The Financial Sector Assessment Program After Ten Years: Experience and Reforms for the Next Decade. Washington DC: IMF and World Bank. Retrieved from <http://www.imf.org/external/np/pp/eng/2009/082809B.pdf>
<http://www.imf.org/external/np/pp/eng/2009/082809B.pdf>

International Monetary Fund (IMF). 2006a. Evaluation Report: Financial Sector Assessment Program. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieo-imf.org/ieo/files/completedevaluations/FSAP_Main_Report.pdf

International Monetary Fund (IMF). 2006b. Evaluation Report: Multilateral Surveillance. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from <http://www.ieo-imf.org/ieo/files/completedevaluations/09012006report.pdf>

International Monetary Fund (IMF). 2006c. Liberia: First Review of Performance Under the Staff-Monitored Program—Staff Report. IMF Country Report No. 06/412. Washington DC: IMF.

International Monetary Fund (IMF). 2009. *An IEO Evaluation of IMF Interactions with its Member Countries*. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from <http://www.ieo-imf.org/ieo/files/completedevaluations/A.%20%20Full%20Text%20of%20Main%20Report.pdf>

International Monetary Fund (IMF). 2007. Evaluation Report: IMF Exchange Rate Policy. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieo-imf.org/ieo/files/completedevaluations/05172007exrate_full.pdf

International Monetary Fund (IMF). 2013. Evaluation Report: The Role of the IMF as Trusted Advisor. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieo-imf.org/ieo/files/completedevaluations/RITA_-_Main_Report.pdf

International Monetary Fund (IMF). 2014. Review of the Financial Sector Assessment Program—Further Adaptation to the Post-Crisis Era. SM/14/249. Washington DC: International Monetary Fund. Retrieved from <http://www.imf.org/external/np/pp/eng/2014/081814.pdf>

IREX. 2014. Media Sustainability Index 2014: The Development of Sustainable Independent Media in Europe and Eurasia. Washington DC: IREX. Retrieved from http://www.irex.org/sites/default/files/u105/EE_MSI_2014_Turkmenistan.pdf

Jacob, Brian A. 2005. "Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics* 89 (5): 761-796.

Jacoby, Wade. 2006. Inspiration, Coalition, and Substitution: External Influences on Postcommunist Transformations. *World Politics* 58 (4): 623-651.

Jennings Jr., Edward T. and Jeremy L. Hall. 2011. Evidence-Based Practice and the Use of Information in State Agency Decision Making. *Journal of Public Administration Research and Theory* 22: 2245-266.

Johnson, Juliet. 2008. "Two-Track Diffusion and Central Bank Embeddedness: The Politics of Euro Adoption in Hungary and the Czech Republic," in *Transnational Actors in Central and Eastern European Transitions*, edited by Mitchell A. Orenstein, Stephen Bloom, and Nicole Lindstrom. Pittsburgh, PA: University of Pittsburgh Press. Pg. 77-97.

Jones, Nicola, Harry Jones, and Cora Walsh. 2008. Political science? Strengthening science–policy dialogue in developing countries. ODI Working Paper 294. London, UK: Overseas Development Institute.

Joyce, Joseph P. 2006. Promises made, promises broken: A model of IMF program implementation. *Economics and Politics* 18 (3): 339–365.

Kahler, Miles. 1992. "External Influence, Conditionality, and the Politics of Adjustment." in *The Politics of Economic Adjustment*, edited by Stephan Haggard and Robert R. Kaufman. Princeton, NJ: Princeton University Press. Pp. 89-133.

Kapstein, Ethan B. and Nathan Converse. 2008. *The Fate of Young Democracies*. Cambridge, UK: Cambridge University Press.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2005. Governance Matters IV: Governance Indicators for 1996-2004. World Bank Policy Research Working Paper No. 3630. Washington, D.C.

Keefer, Philip. 2007. Clientelism, Credibility, and the Policy Choices of Young Democracies. *American Journal of Political Science* 51(4): 804-821.

Kelley, Judith and Beth Simmons. 2014a. Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science*.

Kelley, Judith G. and Beth A. Simmons. 2014b. The Power of Performance Indicators: Rankings, Ratings and Reactivity in International Relations. Paper prepared for the annual meeting of the American Political Science Association, August 27-September 1, 2014, Washington DC.

Kelley, Judith. 2004. International Actors on the Domestic Scene: Membership Conditionality and Socialization by International Institutions. *International Organization* 58(3): 425-45.

Kelley, Judith. 2006. New Wine in Old Wineskins: Policy Learning and Adaption in The new European Neighborhood Policy. *Journal of Common Market Studies* 44(1): 29-55.

Kelsall, Tim. 2011. Going with the Grain in African Development? *Development Policy Review*.29 (S1): s223–s251.

Kelsall, Tim. 2012. Neo-Patrimonialism, Rent-Seeking and Development: Going with the Grain? *New Political Economy* 17 (5): 677-682.

Kikeri, Sunita, Thomas Kenyon, and Vincent Palmade. 2006 *Reforming the Investment Climate: Lessons for Practitioners*. Washington DC: International Finance Corporation.

Kinzer, Stephen. 2009. His country's saved, but he can't go back. *The Washington Post*. November 22, 2009.

- Kleeman, Kristof. 2010. European Neighbourhood Policy: A Reality Check. How effective is the European Neighbourhood Policy in promoting good governance? Berlin, Germany: Hertie School of Governance.
- Konneh, Amara and Steve Radelet. 2008. Liberia Shows Dramatic Improvement in Controlling Corruption. June 2008. Monrovia, Liberia: Executive Mansion. http://www.emansion.gov.lr/doc/Liberia_Corruption.pdf
- Knack, Stephen. 2007. Measuring Corruption: A Critique of Indicators in Eastern Europe and Central Asia. *Journal of Public Policy* 27 (3): 255-291.
- Knack, Stephen and Aminur Rahman. 2007. Donor fragmentation and bureaucratic quality in aid recipients. *Journal of Development Economics* 83(1): 176-197.
- Knack, Stephen. 2014. Building or Bypassing Recipient Country Systems: Are Donors Defying the Paris Declaration? *The Journal of Development Studies* 50 (6): 839-854.
- Krampe, Eva, and Christian Henning. 2011, May. Implementing the Common Africa Agricultural Development Programme in Malawi: Conflicts and Communication among Stakeholders. Paper prepared for the 4th European Conference on African Studies, Uppsala, Sweden. Retrieved from <http://www.nai.uu.se/ecas-4/panels/121-140/panel-132/Eva-Krampe-full-paper.pdf>
- Krasner, Stephen. 2011. Foreign Aid: Competing Paradigms. *Journal of Intervention and Statebuilding* 5 (2): 123-149.
- Krasner, Stephen D. and Thomas Risse. 2014. External Actors, State-Building, and Service Provision in Areas of Limited Statehood: Introduction. *Governance* 27 (4): 545–567.
- Krever, Tor. 2013. Quantifying Law: legal indicator projects and the reproduction of neoliberal common sense. *Third World Quarterly* 34(1): 131–150.
- Khwaja, Asim. 2008. Can Good Projects Succeed in Bad Communities? *Journal of Public Economics* 93(7): 899-916.
- Lake, David, and Christopher J. Farris. 2014. Why International Trusteeship Fails: The Politics of External Authority in Areas of Limited Statehood. *Governance*. 27 (4): 569–587.
- Laurance W.F. 2007. A new initiative to use carbon trading for tropical forest conservation. *Biotropica* 39: 20–24.
- Lawson, Andrew, Richard Gerster, and David Hoole. 2005. Learning from Experience with Performance Assessment Frameworks for General Budget Support. November 2005. Synthesis Report. Swiss State Secretariat for Economic Affairs. Retrieved from http://www.gersterconsulting.ch/docs/synthesis_paf_report.pdf
- Levy, Brian. 2014. *Working with the Grain: Integrating Governance and Growth in Development Strategies*. Oxford, UK: Oxford University Press.
- Lewis, Peter. 1996. Economic Reform and Political Transition in Africa: The Quest for a Politics of Development. *World Politics* 49(1): 92-129.
- Liefferink, Duncan, Bas Arts, Jelmer Kamstra and Jeroen Ooijevaar. 2009. Leaders and laggards in environmental policy: A quantitative analysis of domestic policy outputs. *Journal of European Public Policy* 16(5): 677-700.

Lim, S.S., D.B. Stein, A. Charrow, and C.J.L. Murray. 2008. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *Lancet* 372 (9655): 2013-2046.

Lin, Min-Wei and Chilik Yu. 2014. Can Corruption Be Measured? Comparing Global Versus Local Perceptions of Corruption in East and Southeast Asia. *Journal of Comparative Policy Analysis: Research and Practice* 16 (2): 140-157.

Litzenberger, Lee. 2007. IMF/WB “Not Optimistic” As Kyrgyz HIPC Decision Nears. 08 February 2007. Retrieved at <http://wikileaks.org/cable/2007/02/07BISHKEK143.html>

Loayza, Norman, Jamele Rigolini, and Gonzalo Llorente. 2012. Do Middle Classes Bring Institutional Reforms? World Bank Policy Research Working Paper 6015. Washington DC: World Bank.

Lombardi, Domenico and Ngaire Woods. 2008. The Politics of Influence: An Analysis of IMF Surveillance. *Review of International Political Economy* 15 (5): 709–737.

Lu, Chunling, Catherine M Michaud, Emmanuela Gakidou, Kashif Khan, Christopher J. L. Murray. 2006. Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: An independent assessment. *Lancet* 368: 1088–1095.

Mackie, Andrew and Giovanni Caprio. 2011. Assessing the Impact of the PEFA Framework A Study for the PEFA Steering Committee. Volume 1: Main Report. PEFA. Retrieved from <http://www.pefa.org/sites/pefa.org/files/attachments/Assessing%20the%20Impact%20of%20PEFA%20Framework%20Vol%201%20Final%20JUNE%202011%2020611.pdf>

Martin, Lisa. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton: Princeton University Press.

McArthur, John W. 2013. Own the goals: What the Millennium Development Goals have established. *Foreign Affairs* 92 (2): 152-162.

Meagher, Patrick. 2005. Anti-corruption agencies: Rhetoric versus reality. *The Journal of Policy Reform* 8 (1): 69-103.

Meessen, Bruno, Agnès Soucat, and Claude Sekabaraga. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bulletin of the World Health Organization* 89: 153-156.

Millennium Challenge Corporation (MCC). 2013a. The MCC Effect. Issue Brief. Washington D.C.: MCC. <http://www.mcc.gov/documents/reports/issuebrief-2013002131301-mcc-effect.pdf>

Millennium Challenge Corporation (MCC). 2013b. The MCC Effect: The Prospect of Compact Eligibility is Incentivizing Policy Reform. February 2013 Fact Sheet. Washington D.C.: MCC. Retrieved from https://assets.mcc.gov/press/factsheet-2012002115602-mcc-effect-compact-eligibility_1.pdf

Milner, Helen and Dustin Tingley 2013. The Choice for Multilateralism: Foreign Aid and American Foreign Policy. *Review of International Organizations* 8(3): 313-341.

Ministry of Foreign Affairs of Denmark. 2013. Guidelines for development contracts. June 2013. Retrieved from <http://amg.um.dk/en/~media/amg/Documents/Technical%20Guidelines/Guidelines%20for%20Development%20Contracts/Guidelines%20for%20development%20contracts%20final%20June%202013.pdf>

[mg.um.dk/en/~media/amg/Documents/Technical Guidelines/Guidelines for Development Contracts/Guidelines for development contracts final June 2013.pdf](http://mg.um.dk/en/~media/amg/Documents/Technical_Guidelines/Guidelines_for_Development_Contracts/Guidelines_for_development_contracts_final_June_2013.pdf)

Mitchell, Lincoln. 2009. Compromising democracy: state building in Saakashvili's Georgia. *Central Asian Survey* 28(2): 171-183.

Molenaers, Nadia. 2012. The Great Divide? Donor perceptions of budget support, eligibility and policy dialogue. *Third World Quarterly* 33 (5): 791-806.

Molenaers, Nadia and Leen Nijs. 2011. Why the European Commission Fails to Adhere to the Principles of Good Donorship: The Case of the Governance Incentive Tranche. *European Journal of Development Research* 23: 409-425.

Molenaers, Nadia, Anna Gagiano, Lode Smets, and Sebastian Dellepiane. Forthcoming. What determines the suspension of budget support? *World Development*.

Molina-Gallart, Nuria. 2008. Outcome-based conditionality: Too good to be true? Brussels: EUODAD. http://www.euodad.org/uploadedfiles/whats_new/reports/outcome_based_conditionality_final_august_2010.pdf

Momani, Bessma. 2005. Recruiting and Diversifying IMF Technocrats. *Global Society* 19(2): 167- 187.

Momani, Bessma. 2007. IMF Surveillance and America's Turkish Delight. *Central European Review of International Affairs* 27: 5–24.

Montinola, Gabriella R. 2010. When Does Aid Conditionality Work? *Studies in Comparative International Development* 45: 358–382.

Mumssen, Christian, Yasemin Bal Gündüz, Christian Ebeke, and Linda Kaltani. 2013. IMF-Supported Programs in Low Income Countries: Economic Impact over the Short and Longer Term. IMF Working Paper WP/13/273. Washington DC: IMF.

Muth, John F. 1961. Rational Expectations and the Theory of Price Movements. *Econometrica* 29 (3): 315-35.

Naimoli, J. and P. Vergeer, eds. 2010. Verification at a Glance: A Series of Snapshots of Experiences Verifying Performance Linked to Financial Incentives for Results-Based Financing (RBF) Programs from Selected Countries. Washington DC: The World Bank.

Natsios, Andrew. 2010. The clash of the counter-bureaucracy and development. Washington DC: Center for Global Development.

Noland, Marcus. 1997. Chasing Phantoms: The Political Economy of USTR. *International Organization* 51: 365-87.

Norad. 2008. Norway's provision of budget support to developing countries -- guidelines. Oslo, Norway: NORAD.

North, Douglass C., John Joseph Wallis, Steven B. Webb, and Barry R. Weingast. 2007. Limited Access Orders in the Developing World: A New Approach to the Problems of Development. World Bank Policy Research Working Paper 4359. Washington D.C.: World Bank.

Nsouli, Saleh M. 2006. The Euro-Mediterranean Partnership Ten Years On: Reassessing Readiness and Prospects. Statement by Director of the IMF Offices in Europe at Crans-Montana Forum, Monaco. June 23, 2006. Retrieved from <http://www.imf.org/external/np/speeches/2>

Odell, John S. and Dustin H. Tingley. 2013. "Negotiating Agreements in International Relations." In *Negotiating Agreement in Politics*, edited by Jane Mansbridge and Cathie Jo Martin. Washington DC: American Political Science Association.

OECD. 2015. Making Development Co-operation Fit for the Future: A Survey of Partner Countries. OECD Development Co-operation Working Papers, No. 20. Paris: OECD. Retrieved from <http://dx.doi.org/10.1787/5js6b25hzv7h-en>

OECD and German Development Institute (DIE). 2014. Technical workshop on results-based funding. 19-20 May 2014. Paris, France: OECD. Retrieved from http://www.die-gdi.de/uploads/tx_veranstaltung/20140519_Results_based_Funding_Workshop_Report.pdf

Öhler, Hannes, Axel Dreher, and Peter Nunnenkamp. 2012. Does Conditionality Work? A Test for an Innovative US Aid Scheme. *European Economic Review* 56: 138-153.

Orbie, Jan and Lisa Tortell. 2009. The New GSP + Beneficiaries: Ticking the Box or really Consistent with ILO Findings? *European Foreign Affairs Review* 14 (5): 663–681.

Parks, Bradley C. 2014. Brokering Development Policy Change: The Parallel Pursuit of Millennium Challenge Account Resources and Reform. Ph.D. Dissertation. London, UK: London School of Economics and Political Science.

Parks, Bradley C. and Michael Tierney. 2004. Cooperation or Collusion: Explaining Bilateral and Multilateral Environmental Aid to Developing Countries. Paper presented at the 2004 American Political Science Association Meeting, Chicago, IL, September 1-5, 2004. http://www.wm.edu/offices/itpir/documents/aiddata/cooperation_collusion_2004.pdf

Parks, Bradley C., Taylor Wilson, Diana Winter, Michael Testa, and Torey Beth Jackson. 2014. Expanding the Policy Influence of the State Department's Investment Climate Statements. A Diplomacy Lab Report Prepared by the Institute for the Theory and Practice of International Relations. Williamsburg, VA: The College of William and Mary.

Parks, Bradley C. and Zachary J. Rice. 2013. Measuring the Policy Influence of the Millennium Challenge Corporation: A Survey-Based Approach. Working Paper. Williamsburg, VA: The College of William and Mary.

Parks, Bradley C. and Zachary J. Rice. 2014. Buying Influence in a Crowded Market: Aid Policy in the Post- DAC World. Presentation delivered at the "South-South Development Cooperation: Chances and Challenges for the International Aid Architecture" Workshop at Heidelberg University, September 26-27, 2014.

Perakis, Rita and William Savedoff. 2015. Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies. CGD Policy Paper 053. Washington DC: Center for Global Development.

Perrin, Burt. 2002. Implementing the vision: Addressing challenges to results-focused management and budgeting. Paris: OECD. <http://www.seachangeop.org/sites/default/files/documents/2002%20Results-focused%20managing%20and%20budgeting.pdf>

Persson, Anna, Bo Rothstein, and Jan Teorell. 2013. Why Anticorruption Reforms Fail—Systemic Corruption as a Collective Action Problem. *Governance* 26 (3): 449–471.

Pew Research Center. 2012. *U.S. Public, Experts Differ on China Policies*. Washington DC: Pew Research Center. Retrieved from <http://www.pewglobal.org/files/2012/09/US-Public-and-Elite-Report-FINAL-FOR-PRINT-September-18-2012.pdf>

Pham, J. Peter. 2009. Behind the Curve: Corrupt governments cash in on the Millennium Challenge Corporation’s outdated metrics. *Stanford Social Innovation Review* Fall 2009.

Phillips, David A. 2009. *Reforming The World Bank: Twenty Years of Trial —and Error*. New York: Cambridge University Press.

Phillips, Michael. 2007. Bush's Aid Policy Prods Countries: Yemen and Lesotho Embrace Overhauls; The Gambia Balks. *The Wall Street Journal*. March 14, 2007; Page A6.

Piot, Peter, Sarah Russell, and Heidi Larson. 2007. Good Politics, Bad Politics: The Experience of AIDS. *American Journal of Public Health* 97 (11): 1934–1936.

Pitlik, Hans, Björn Frank, and Mathias Firchow. 2010. The demand for transparency: An empirical note. *Review of International Organizations* 5: 177-195.

Pop-Eleches, Grigore. 2009. *From Economic Crisis to Reform: IMF Programs in Latin American and Eastern Europe*. Princeton, NJ; Princeton University Press.

Pritchett, Lant, and Michael Woolcock. 2004. Solutions When the Solution is the Problem: Arraying the Disarray in Development. *World Development* 32 (2): 191-212.

Pritchett, Lant, Michael Woolcock, and Matt Andrews. 2013. Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation. *Journal of Development Studies* 49 (1): 1-18.

Pritchett, Lant, Salimah Samji and Jeffrey Hammer. 2012. “It’s All About MeE: Using Structured Experiential Learning (‘e’) to Crawl the Design Space. UNU-WIDER Working Paper 012/104. Helsinki: UNU-WIDER.

Pritchett, Lant. 2013. Folk and Formula: Fact and Fiction in Development. Annual UNU Wider Lecture #16, Helsinki, Finland: UNU-WIDER.

Pritchett, Lant, Michael Woolcock, and Matt Andrews. 2014. Capability Traps in Development: How Initiatives to Improve Administrative Systems Succeed at Failing. *Prism* 3 (3): 63-74.

Radelet, Steve. 2003. *Challenging Foreign Aid: A Policymaker's Guide to the Millennium Challenge Account*. Washington D.C.: Center for Global Development.

Radelet, Steve. 2006. “Pulling not Pushing Reforms: Delivering Aid through Challenge Grants.” in *The New Public Finance: Responding to Global Challenges*, edited by Inge Kaul and Pedro Conceicao. New York and Oxford: Oxford University Press.

Radelet, Steve. 2007. The Millennium Challenge Account in Africa: Testimony Before the House Committee On Foreign Affairs Subcommittee on Africa and Global Health. June 28, 2007.

Rajan, Raghuram, 2004. Why Are Structural Reforms So Difficult? *Finance and Development* 41 (2): 56–57.

- Rajan, Raghuram G. 2009. Rent Preservation and the Persistence of Underdevelopment. *American Economic Journal: Macroeconomics* 1(1): 178-218.
- Ravallion, Martin. 2011. Knowledgeable Bankers? The Demand for Research in World Bank Operations. World Bank Policy Research Paper 5892. Washington DC: World Bank.
- Reitmaier, Angela. 2014. Funding African Governance Institutions and Processes: Insights and Lessons. SAIIA Occasional Paper #117. Johannesburg, South Africa: SAIIA.
- Remmer, Karen L. 1986. The Politics of Economic Stabilization: IMF Standby Programs in Latin America, 1954-1984. *Comparative Politics* 19: 1-24.
- Rheault, Magali and Bob Tortora. 2011. Confidence in Institutions: Africans Speak on the Meaning of Being Well Governed. *Harvard International Review* 32 (4): 72-77.
- Ribadu, Nuhu. 2010. *Show Me the Money: Leveraging Anti-Money Laundering Tools to Fight Corruption in Nigeria*. Washington D.C.: Center for Global Development.
- Rieff, David. 2005. Che's Second Coming? *New York Times* 20 November 2005. Retrieved from http://www.nytimes.com/2005/11/20/magazine/20bolivia.html?pagewanted=print&_r=0
- Robinson, Brooks. 2008a. Liberia: 21 Reforms to Improve the Business Climate May Boost MCC Score. 5 May 2008. Retrieved from <http://wikileaks.org/cable/2008/05/08MONROVIA336.html>
- Robinson, Brooks. 2008b. Liberia: President Turns To Young Technocrats For New Economic Team. 7 August 2008. Retrieved from <http://www.cablegatesearch.net/cable.php?id=08MONROVIA612>
- Rodrik, Dani. 1996. "Why Is there Multilateral Lending?" In Annual World Bank Conference on Development Economics 1995, edited by M. Bruno, and B. Pleskovic. Washington DC: the World Bank.
- Rodrik, Dani. 2007. *One Economics, Many Recipes: Globalization, Institutions and Economic Growth*. Princeton, NJ: Princeton University Press.
- Rogelberg, S. G. and J.M. Stanton. 2007. Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods* 10: 195-209.
- Rosenberg, Tina. 2006. To Fight Corruption, One African Offers Presidents Cash. *The New York Times*. November 24, 2006. Retrieved from http://www.nytimes.com/2006/11/24/opinion/24fri4.html?_r=0
- Rosenzweig, Mark R. and K.I. Wolpin. 1982. Governmental Interventions and Household Behavior in a Developing Country: Anticipated and Unanticipated Consequences of Social Programs. *Journal of Development Economics* 10: 209-225.
- Routley, Laura. 2014. Developmental States in Africa? A Review of Ongoing Debates and Buzzwords. *Development Policy Review* 32 (2): 159-177.
- Rotberg, Robert. 2009. Governance and Leadership in Africa: Measures, Methods, and Results. *Journal of International Affairs* 62 (2): 113-126.

Samuel, Boris. 2013. Economic Calculations, Instability, and (In)formalization of the State in Mauritania, 2003-2011, Conference on African Economic Development: Measuring Success And Failure. April 18 – 20, 2013, Simon Fraser University.

Santiso, Carlos. 2003. The Gordian Knot of Brazilian Foreign Policy: Promoting democracy while respecting sovereignty. *Cambridge Review of International Affairs* 16 (2): 343-358.

Sasse, Gwendolyn. 2008. The European Neighbourhood Policy: Conditionality Revisited for the EU's Eastern Neighbours. *Europe-Asia Studies* 60 (2): 295-316.

Savedoff, William. 2011. Incentive Proliferation? Making Sense of a New Wave of Development Programs. Center for Global Development Essay. Washington D.C.: Center for Global Development.

Schäferhoff, Marco. 2014. External Actors and the Provision of Public Health Services in Somalia. *Governance* 27 (4): 675–695.

Schimmelfennig, Frank and Ulrich Sedelmeier. 2004. Governance by conditionality: EU rule transfer to the candidate countries of Central and Eastern Europe. *Journal of European Public Policy* 11(4): 669–687.

Schueth, Samuel J. 2011. Assembling International Competitiveness: The Republic of Georgia, USAID, and the Doing Business Project. *Economic Geography* 37(1): 51-77.

Shafer, D. Michael. 1994. *Winners and Losers: How Sectors Shape the Developmental Prospects of States*. Ithaca NY: Cornell University Press.

Sharman, Jason. 2008. Power, Discourse and Policy Diffusion: Anti-Money Laundering in Developing States. *International Studies Quarterly* 52: 635-656.

Sharman, Jason. 2009. The Bark is the Bite: International Organizations and Blacklisting. *Review of International Political Economy* 16: 573-596.

Sharman, Jason. 2011. Small is Beautiful: The Perils of Gigantism and the Pay-offs of Studying Small International Organisations. Presented at the AGORA Workshop on Institutional Diversity in Governance of the Global Economy, Brisbane, Australia, January 11-12, 2011.

Sheehan, Kim B. 2006. E-mail Survey Response Rates: A Review. *Journal of Computer-Mediated Communication* 6 (2).

Sherr, James. 2013. Ukraine and Europe: Final Decision? London, UK: Chatham House.

Sherry, James, Sangeeta Mookherji, and Leo Ryan. 2009. The Five-Year Evaluation of the Global Fund to Fight AIDS, Tuberculosis, and Malaria: Synthesis of Study Areas 1, 2 and 3. Macro International Inc.

Shih, Tse-Hua, and Xitao Fan. 2008. Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods* 20: 249-271.

Simmons, Beth A. 2000. The Legalization of International Monetary Affairs. *International Organization* 54: 573-602.

Sjöstedt, M. 2013. Aid effectiveness and the Paris Declaration: A mismatch between ownership and results-based management? *Public Administration and Development* 33: 143–155.

Smets, Lodewijk. 2014. World Bank Policy Lending for Public Sector Reform. Paper presented at CESifo Reforming the Public Sector Workshop. 25-26 July 2014. Bay of Venice, Italy.

Smith, Wellington Geevon. 2007. Prez Johnson-Sirleaf, MCC hold discussion. STAR Radio. February 12, 2007. <http://www.starradio.org.lr/latest-news/prez-Johnson-Sirleaf-mcc-hold-discussion.html>

Smith, Alastair and James Raymond Vreeland. 2006. "The Survival of Political Leaders and IMF Programs." in *Globalization and the Nation State: The Impact of the IMF and The World Bank*, edited by G. Ranis, J. R. Vreeland and S. Kosack. New York: Routledge. Pp. 263- 289.

Soederberg, Susanne. 2004. American empire and 'excluded states': the millennium challenge account and the shift to pre-emptive development. *Third World Quarterly* 25 (2): 279-302.

Soeters R, Habineza C, Peerenboom PB. 2006. Performance-based financing and changing the district health system: experience from Rwanda. *Bulletin of the World Health Organization* 84: 884-889.

Spitzmüller, C., D. Glenn, M. Sutton, C. Barr, and S. Rogelberg. 2007. Survey nonrespondents as bad soldiers: Examining the relationship between organizational citizenship and survey response behavior. *International Journal of Selection & Assessment* 15(4): 449-59.

Srivastava, Vivek and Marco Larizza. 2013. Working with the Grain for Reforming the Public Service: A Live Example from Sierra Leone. *International Review of Administrative Sciences* 76 (3): 458–85.

Stahl, Bernhard. 2011. Perverted Conditionality: The Stabilisation and Association Agreement between the European Union and Serbia. *European Foreign Affairs Review* 16 (4): 465–487.

Stecklov, Guy, and Alexander Weinreb. 2010. Improving the Quality of Data and Impact Evaluation Studies in Developing Countries. Impact Evaluation Guidelines No. 1: Strategy Development Division. Washington DC: Inter-American Development Bank.

Steen, Anton. 2003. *Political Elites and the New Russia: The Power Basis of Yeltsin's and Putin's Regimes*. London, UK: Routledge.

Stewart, Susan. 2013. Public Procurement Reform in Ukraine: The Implications of Neopatrimonialism for External Actors. *Demokratizatsiya* 21 (2): 197-214.

Stone, Randall W. 2004. The Political Economy of IMF Lending in Africa. *American Political Science Review* 98 (4): 577-591.

Stone, Randall. 2002. *Lending Credibility: The International Monetary Fund and the Post- Communist Transition*. Princeton: Princeton University Press.

Sutherland, Peter D. 2008. Transforming Nations: How the WTO Boosts Economies and Opens Societies. *Foreign Affairs* 87 (2): 125-136.

Swedlund, Haley. 2013. The domestication of governance assessments: Evidence from the Rwandan Joint Governance Assessment. *Conflict, Security & Development* 13 (4): 449–470.

Swift-Morgan, Jennifer. 2014. Confronting the Informal in Good Governance: The Case of Decentralised Education-System Reform in Guinea. *Development Policy Review* 32 (4): 399–425.

Tang, Man-Keung and Shang-Jin Wei. 2008. Does WTO Accession Raise Income? When External Commitments Create Value. NBER Working Paper.

Tavakoli, Heidi, Ismaila Cessay, and Winston Cole. 2014. Success when stars align: Public financial management reforms in Sierra Leone. WIDER Working Paper 2014/081. Helsinki: UNU-WIDER.

Taylor, Jeannette. 2011. Factors Influencing the Use of Performance Information for Decision Making in Australian State Agencies. *Public Administration* 89(4): 1316-1334.

Taylor, John B. 2004. Economic Freedom and Georgia's Rose Revolution. Remarks at the Caucasus Business School, Tbilisi, Georgia. November 22, 2004.

The Economist. 2000. Dancing in Kenya to the donors' tune. 3 August 2000. Retrieved from <http://www.economist.com/node/28457>

The Economist. 2003. Dirt out, cash in. 27 November 2003. Retrieved from <http://www.economist.com/node/2249436>

The Economist. 2005. Unblocking business. 17 September 2005.

The Economist. 2014. The final frontier: Fixing dreadful sanitation in India requires not just building lavatories but also changing habits. 19 July 2014. Retrieved from <http://www.economist.com/news/asia/21607837-fixing-dreadful-sanitation-india-requires-not-just-building-lavatories-also-changing>

Transparency International. 2015. Corruption Perceptions Index: Overview. Retrieved from <http://www.transparency.org/research/cpi/overview>

Trapnell, Stephanie E. 2011. Actionable Governance Indicators: Turning Measurement into Reform. *Hague Journal on the Rule of Law* 3 (22): 317-348.

UNDP and Global Integrity 2008. A User's Guide to Measuring Corruption. Oslo, Norway: UNDP Oslo Governance Centre. Retrieved from http://www.undp.org/content/dam/aplaws/publication/en/publications/democratic-governance/dg-publications-for-website/a-users-guide-to-measuring-corruption/users_guide_measuring_corruption.pdf

United Nations (UN). 2013. A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development. New York, NY: UN. Retrieved from <http://www.post2015hlp.org/wp-content/uploads/2013/05/UN-Report.pdf>

Unsworth, Sue. 2009. What's Politics Got to Do With It? Why Donors Find it So Hard to Come to Terms with Politics, and Why this Matters. *Journal of International Development* 21(6): 883-894.

Vachudova, Milada Anna. 2013. EU Leverage and National Interests in the Balkans: The Puzzles of Enlargement Ten Years On. *Journal of Common Market Studies* 52 (1): 1-17.

Van De Walle, Nick. 2001. *African Economies and the Politics of Permanent Crisis*. New York: Cambridge University Press.

Volcker, Paul. 2009. The Changing Nature of Government Service. Princeton, NJ: Woodrow Wilson School.

Vreeland, James Raymond. 2003. *The IMF and Economic Development*. Cambridge, UK: Cambridge University Press.

Wallander, Celeste. 2002. NATO's Price: Shape up or Ship out. *Foreign Affairs* 81 (6): 2-8.

Ward, Hugh, and Peter John. 2013. Competitive Learning in Yardstick Competition: Testing Models of Policy Diffusion With Performance Data. *Political Science Research and Methods* 1 (1): 3-25.

Warren, Rick. 2009. Paul Kagame. *Time Magazine*. April 30, 2009. http://www.time.com/time/specials/packages/article/0,28804,1894410_1893847_1893843,00.html. Retrieved on March 7, 2010.

Watson, Kenneth. 2008. Mid-Term Review of the GEF Resource Allocation Framework: Comparison of the GEF RAF with other Performance-Based Allocation Systems. Technical Paper #8. Global Environment Facility (GEF) Evaluation Office. Washington DC: GEF. Retrieved from https://www.thegef.org/gef/sites/thegef.org/files/documents/RAF_MTR-TECH_Paper_8.pdf

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago, IL: The University of Chicago Press.

Weyland, Kurt. 2007. *Bounded Rationality and Policy Diffusion: Social Sector Reform in Latin America*. Princeton, NJ: Princeton University Press.

Weymouth, Stephen and J. Muir Macpherson. 2012. The Social Construction of Policy Reform: Economists and Trade Liberalization around the World. *International Interactions* 38 (5): 670-702.

Whitfield, Lindsay (ed.). 2009. *The Politics of Aid: African Strategies for Dealing with Donors*. Oxford, UK: Oxford University Press.

Wolff, Hendrik, Howard Chong, and Maximilian Auffhammer. 2011. Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index. *Economic Journal* 121: 843–870.

World Bank (WB). 2006a. *Global Monitoring Report 2006*. Washington DC: World Bank.

World Bank (WB). 2006b. Governance and Anti-Corruption: Ways to Enhance the World Bank's Impact. Washington DC: Independent Evaluation Group (IEG) of the World Bank.

World Bank (WB). 2007. *Celebrating Reform 2007: Doing Business Case Studies*. Washington D.C.: World Bank.

World Bank (WB). 2009. *Doing Business 2010: Reforming Through Difficult Times*. New York: Palgrave MacMillan.

World Bank (WB). 2010. World Bank Economic Reports on Growth Diagnostics in Four African Countries: Ghana, Mauritius, Nigeria, and Uganda. Performance Assessment Review. Report No.: 55404. Washington DC: World Bank.

World Bank (WB). 2011. *World Development Report 2012: Gender Equality and Development*. Washington DC: World Bank.

World Bank (WB). 2013. Rwanda Economic Update. Fourth Edition. Maintaining the Momentum – With a Special Focus on Rwanda's Pathway out of Poverty. Washington DC: World Bank.

World Bank (WB). 2014. Client Feedback Survey of FY13 Analytical and Advisory Activities (AAA): Summary Results. Washington DC: World Bank. Retrieved at

http://www.worldbank.org/content/dam/Worldbank/document/ProductsandServices/CFI%20FY13%20Summary%2009_04_2014.pdf

World Bank (WB) and the International Finance Corporation (IFC). 2013. *Women, Business and the Law 2014: Removing Restrictions to Enhance Gender Equality*. London, UK: Bloomsbury Publishing Plc.

Wynia, Matthew K., Deborah S. Cummins, Jonathan B. VanGeest, and Ira B. Wilson. 2000. Physician manipulation of reimbursement rules for patients: between a rock and a hard place. *Journal of the American Medical Association* 283 (14): 1858–1865.

Yovanovitch, Marie. 2006a. Kyrgyz HIPC Debate Becomes Political Hot Potato. 30 November 2006. Retrieved at http://www.wikileaks.org/plusd/cables/06BISHKEK1692_a.html

Yovanovitch, Marie. 2006b. Kyrgyz Parliament Confirms New Prime Minister. 29 January 2007. Retrieved at <http://wikileaks.org/cable/2007/01/07BISHKEK87.html>

Zimelis, Andris. 2011. Conditionality and the EU–ACP Partnership: A Misguided Approach to Development? *Australian Journal of Political Science* 46 (3): 389–406.

Zwane, Alix Peterson and Kremer, Michael. 2007. What works in fighting diarrheal diseases in developing countries? A critical review. *The World Bank Research Observer* 22 (1): 1–24.